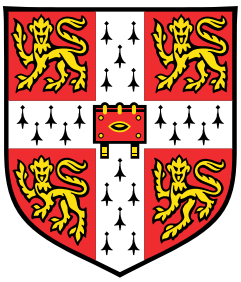
Understanding tropical forest dynamics through remote sensing and deep learning

James G. C. Ball

Understanding tropical forest dynamics through remote sensing and deep learning



James George Clifford Ball

Department of Plant Sciences
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Magdalene College November 2023

I cannot see what flowers are at my feet,
    Nor what soft incense hangs upon the boughs,
But, in embalmed darkness, guess each sweet
    Wherewith the seasonable month endows
The grass, the thicket, and the fruit-tree wild;
    White hawthorn, and the pastoral eglantine;
        Fast fading violets cover'd up in leaves;
            And mid-May's eldest child,
    The coming musk-rose, full of dewy wine,
        The murmurous haunt of flies on summer eves.

– John Keats, *Ode to a Nightingale*

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

James George Clifford Ball
November 2023

Acknowledgements

Completing this PhD has been a journey filled with ups and downs, yet it stands as one of the most rewarding, exhilarating, and enjoyable experiences of my life. I extend my deepest gratitude to all those who made this possible.

I am profoundly grateful to Professor David Coomes, my supervisor, for his invaluable contributions to my academic and personal growth. His remarkable depth of knowledge, open-minded approach, and patience with my explorations in AI have been pivotal in shaping my research. I thank him for being there in my times of need including when he drove down to Magdalene College with snorkeling equipment so that I could rescue my phone from the river.

Dr. Gregoire Vincent, my secondary supervisor, deserves special thanks for introducing me to fieldwork with his gentle but precise guidance. His warm hospitality when hosting me in Montpellier was greatly appreciated - the cheese and wine was exquisite. His sharp attention to detail and his connection to the amazing Paracou site in French Guiana have greatly enriched my research experience. I also owe a debt of gratitude to Dr. Nicolas Barbier, the PhenObs project leader, for welcoming me to the project, showing me his drones, and for giving me the thrilling opportunity to climb rainforest trees, a memory I will always cherish. My heartfelt thanks to the broader teams at the AMAP and ECOFOG labs for their supportive environment. Their collective wisdom and assistance have been invaluable.

The unwavering support and patience of my Mum, Dad and sister has allowed me to pursue my academic dreams. I promise them that this is my last degree. A special thank you to Gerry for always allowing me to ask questions and demonstrating that trying to answer them is a noble cause. To the KRM, thank you for keeping me grounded and uplifted. Gus, your intellectual prowess goes without saying but your willingness to listen and respond to my ideas has helped me maintain momentum even when things have been tough.

Professor Seth Flaxman's introduction to machine/deep learning helped to set my course. Professor Mark Burgman's life spark has been instrumental in my growth. I am grateful to Jean Louis Smock for his help in the field, Julien Angel, for being a botanical wizard, and Jansen for his spicy fish curry (except for the occasional hidden chilli). Ilona Clocher, your energy and enthusiasm were always uplifting. I extend my gratitude to Yi for all kinds of practical support, and Toby, for being a pal in the lab, including me in projects, and for intellectual guidance and

support. Special thanks to 4C, Anil, and his crypto millions for providing Sherwood which has been crucial for my research. Seb, your AI inspiration was huge and thanks for the MCC guest pass. Leah, your moral support, snacks, energy drinks, and proofreading have been an invaluable help. Beth, your careful proofreading at crunch time was indispensable.

I am grateful to NERC for funding my research and organising the DTP programme. Being part of the C-CLEAR cohort was energising and enriching.

Finally, to the many others who have supported me in various ways, please accept my apologies if I have inadvertently left you out of this acknowledgment. Your contributions have been deeply appreciated.

Abstract

Protection of tropical forests is key to achieving global climate and biodiversity conservation goals. They play an essential role in carbon sequestration, water cycling, and nutrient exchanges, thereby regulating atmospheric composition and global climate patterns. However, they are under interlinked threats from deforestation and climate change, exacerbating biodiversity loss and potentially pushing these systems toward ecological tipping points. Computer vision techniques based on deep learning have emerged as potent tools for monitoring, conservation, and prediction efforts within these expansive and intricate ecosystems. This thesis uses these emerging technologies to understand forest dynamics at scales ranging from the phenology of individual tree crowns to large-scale deforestation. **Chapter 1** introduces tropical forests, explores their ecological and societal value, and discusses the technological challenges and opportunities of studying them, with a focus on deep learning as applied to remote sensing data. **Chapter 2** develops a tool to predict deforestation patterns based on convolutional neural networks (CNNs), working with freely accessible data to successfully forecast spatiotemporal patterns in the Southern Peruvian Amazon. Predicting the location of deforestation is difficult as it results from complex interactions within human-ecological systems but doing so may enable effective, adaptable prevention measures and conservation planning. The models, through their ability to discern deforestation drivers such as new access routes from remote sensing data, highlight the potentially transformational role of deep learning in conservation. In **Chapter 3**, I develop a new approach named *detectree2*, building on the Mask R-CNN architecture, which is capable of accurately detecting and delineating individual tree crowns from airborne RGB imagery taken over dense tropical forests. The foundation for any remote-sensing study of individual tree dynamics is accurate tree delineation. Trialled in diverse geographies, including Malaysian Borneo and French Guiana, I show this tool holds promise for large-scale forest studies. The performance of the detection and delineation, especially for tall trees, enables tracking of tree growth and mortality for the study of carbon dynamics from cheap, widely accessible photographic data. **Chapter 4** develops a pipeline for identifying and mapping tropical tree species, building on the *detectree2* approach. This pipeline combines aerial photographic images taken every three weeks using a UAV with hyperspectral survey. Training and testing on a carefully crafted ground truth dataset, the two-step approach applies

detectree2 to multitemporal UAV-RGB data in order to automatically segment trees and then applies Linear Discriminant Analysis (LDA) to hyperspectral data to assign species. This new approach identified over sixty tree species with high confidence, achieving accurate species level mapping over 70% of the total crown area of the landscape. Key to the improved mapping was the temporal stacking of imagery to delineate tree crowns accurately and a large, rigorously validated dataset of labelled tree crowns to train on. In **Chapter 5**, I use the data and techniques developed in the previous two Chapters to address ecological questions related to the phenology of tropical forests. Seasonal variation in canopy greenness has been observed from space, but the extent to which all species in diverse forests follow a similar pattern of leaf pigment changes, leaf flushing and loss remains unknown. I begin to address that knowledge gap by tracking phenology through drone-mounted sensors, providing a dataset that tracked individual trees in French Guiana at 3-weekly intervals over 34 months. 3,000 tree crowns were mapped and tracked using UAV LiDAR, revealing significant spatiotemporal variability in Plant Area Density (PAD) and distinct species-specific phenological patterns. By juxtaposing PAD with spectral metrics, I start to decipher variation in "leaf amount" and "leaf quality", offering some insights into how individual tree changes might impact forest productivity. Concluding **Chapter 6** discusses ways in which integration of deep learning technologies and remote sensing into ecology research is helping to broaden understanding and conservation capabilities for tropical forests, by providing precise, scalable solutions spanning deforestation prediction, tree level monitoring, species identification, and phenological studies.

Publications and collaborations

Collaborations

I have had the tremendous good fortune to be involved with enriching collaborative projects throughout the course of my PhD which have supported my work described in this thesis. These Chapters have been written as scientific papers, two of which are published and two of which are in an advanced stage of preparation. Since multiple authors contributed to the "chapters", I use the collective pronoun "we" instead of "I" where appropriate. Chapters 4-5 were the result of my work as part of the *PhenObs* project, a Labex CEBA funded proposal led by Dr Nicolas Barbier to establish a phenology observatory in French Guiana to study climate-vegetation feedbacks and the diversity of plant strategies. Owing to the ambitious breadth of this project, I have received support from an extensive network of researchers and technicians. As a result, the simple accreditation of contribution through co-authorship falls short. I aim to comprehensively describe the range of contributions below knowing that I too can only fall short in showing my full appreciation to all those individuals and organisations that have supported my work. Professor David Coomes gave supervisory support throughout work presented in this thesis.

In **Chapter 2**, I develop and implement a deep learning system to forecast deforestation. Katerina Petrova (Imperial College London) assisted in designing the neural network architectures and Professor Seth Flaxman (then Imperial College London, now University of Oxford) helped oversee the technical implementation of the system. Professor Mark Burgman (Imperial College London), while not directly involved in the work, had the initial inspiration to make the connections and form the collaboration. The decision to develop this work was in part due to delays to fieldwork (described below) resulting from the COVID-19 pandemic.

In **Chapter 3**, I develop and apply a convolutional neural network system for detecting individual trees in tropical from aerial photographs and apply it to assessing tree growth and mortality at three tropical forest sites. This project involved Sebastian Hickman who was working towards his MRes degree and was also supported by Dr Tobias Jackson and Professor David Coomes in the Forest Ecology and Conservation group. I worked with Sebastian Hickman on the initial coding and I continued to develop the code base towards the form it took for the final publication. Xian Jing Koay and James Hirst contributed to the code and

helped to tune the model training. The 2014 Sabah, Malaysia, data acquisitions were funded by NERC and William Jay (NEODAAS) orthorectified this imagery from Sabah, Malaysia. Matthew Archer gave technical support for implementing the system as an open source Python package readily available for researchers around the world to install. Mélaine Aubry-Kientz, Grégoire Vincent and I performed the manual tree crown delineation and field validation of the Paracou, French Guiana data. CNES funded the 2016 hyperspectral, RGB and lidar data over Paracou and Labex CEBA contributed financial resource for the field validation of manual crown segmentations (described in more detail below). Tobias Jackson and David Coomes supervised the project.

For **Chapters 4-5**, I am grateful to have been part of the *PhenObs* team and to have received funding for this project through a Labex CEBA (Centre d'étude de la biodiversité amazonienne) grant of € 200,000. The team included Nicolas Barbier (Project coordinator; IRD), Grégoire Vincent (IRD), Isabelle Maréchaux (INRA), Patrick Heuret (INRA), Claire Fortunel (IRD), Jean-Louis Smock (IRD), Chantal Geniez (IRD), Julien Engel (IRD), Raphaël Pelissier (IRD), Géraldine Derroire (Director of Paracou Research Station; CIRAD),Ilona Clocher (CNRS) and Benjamin Lagrange.

Chapters 4-5 (and to a lesser extent Chapter 3) rely heavily on fieldwork that was conducted at Parcou Research Station, French Guiana, that would not have been possible without the support of:

- Grégoire Vincent - supervision and assistance with manual tree crown mapping
- Julien Engel - assistance with manual tree crown mapping and identification of tree species
- Ilona Clocher, Mélaine Aubry-Kientz, Ilona Clocher, Jean-Louis Smock, Chantal Geniez, Tom Hattermann (CNRS), Isabelle Marechaux - assistance with manual tree crown mapping
- Patrick Heuret - botanical guidance in the field
- Bárbara Thaís de Alencar, Uelison Mateus Ribeiro, Thomas Cedat, Romain Tchakamian - Module FTH AgroParisTech field course students I supervised who helped with manual tree crown mapping at Paracou
- Géraldine Derroire and Laetitia Proux - facilitated access to Paracou Research Station

These Chapters also relied on UAV imagery that was collected every three weeks over three years at the site. The UAV-lidar, -RGB and -multispectral data required processing to analysis ready datasets. This was supported by:

- Iloner Clocher, Jean-Louis Smock and Benoit Burban - regular drone surveys
- Grégoire Vincent - processing of lidar data into PAI/PAD data with his software package AMAPvox
- Nicolas Barbier - overall supervision of drone surveys and data processing chain including orthomosaicking and data co-registration
- Phillippe Verly (IRD) - technical support on the processing of all data types to analysis ready products

In **Chapter 4**, Anthony Laybros, Colin Prieur and Grégoire Vincent gave important advice on how to analyse the hyperspectral data. Sadiq Jaffer gave advice on machine learning methods and implemented routines for model training and hyperparameter optimisation. Nicolas Barbier gave supervision and guidance on the use of multitemporal RGB data and the general framing of the paper. Nicolas Barbier and Philippe Verley processed the multitemporal RGB data.

In **Chapter 5**, Géraldine Derroire and Patrick Heuret gave guidance on how to analyse the phenological signals of the trees observed in the remote sensing data. Isabelle Marechaux and Nicolas Barbier gave broad advice on the interpretation and analysis of data. Nicolas Barbier, Gregoire Vincent and Philippe Verley processed the multitemporal RGB data and lidar data to an analysis ready state.

Publications

These chapters have been published:

- **Chapter 2**: Ball, J. G. C., Petrova, K., Coomes, D. A., & Flaxman, S. (2022). Using deep convolutional neural networks to forecast spatial patterns of Amazonian deforestation. *Methods in Ecology and Evolution*, *13*(11), 2622–2634
- **Chapter 3**: Ball, J. G. C., Hickman, S. H. M., Jackson, T. D., Koay, X. J., Hirst, J., Jay, W., Archer, M., Aubry-Kientz, M., Vincent, G., & Coomes, D. A. (2023). Accurate delineation of individual tree crowns in tropical forests from aerial RGB imagery using Mask R-CNN. *Remote Sensing in Ecology and Conservation*, *9*(5), 641–655

I have also contributed to other projects during the course of my PhD that have been published:

- Aubry-Kientz, M., Laybros, A., Weinstein, B., Ball, J. G. C., Jackson, T., Coomes, D., & Vincent, G. (2021). Multisensor data fusion for improved segmentation of individual tree crowns in dense tropical forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 3927–3936

- Cao, Y., Ball, J. G. C., Coomes, D. A., Steinmeier, L., Knapp, N., Wilkes, P., Disney, M., Calders, K., Burt, A., Lin, Y., & Jackson, T. D. (2023). Benchmarking airborne laser scanning tree segmentation algorithms in broadleaf forests shows high accuracy only for canopy trees. *Int. J. Appl. Earth Obs. Geoinf.*, *123*, 103490

- Vincent, G., Verley, P., Brede, B., Delaitre, G., Maurent, E., Ball, J. G. C., Clocher, I., & Barbier, N. (2023). Multi-sensor airborne lidar requires intercalibration for consistent estimation of light attenuation and plant area density. *Remote Sensing of Environment*, *286*, 113442

- Brodie, J. F., Mohd-Azlan, J., Chen, C., Wearn, O. R., Deith, M. C. M., Ball, J. G. C., Slade, E. M., Burslem, D. F. R. P., Teoh, S. W., Williams, P. J., Nguyen, A., Moore, J. H., Goetz, S. J., Burns, P., Jantz, P., Hakkenberg, C. R., Kaszta, Z. M., Cushman, S., Coomes, D., … Luskin, M. S. (2023). Landscape-scale benefits of protected areas for tropical biodiversity. *Nature*, *620*(7975), 807–812

One project that I contributed to remains unpublished but is available as a pre-print:

- Bandyopadhyay, D., Mukherjee, S., Ball, J. G. C., Vincent, G., Coomes, D. A., & Schönlieb, C.-B. (2022). Tree species classification from hyperspectral data using graph-regularized neural networks arXiv 2208.08675

The *detectree2* system is available on GitHub to install as an open source Python package:

- Ball, J. G. C. (2022). Detectree2: Python package for automatic tree crown delineation based on the detectron2 implementation of mask R-CNN

An interview discussing my fieldwork was published as a feature in *Nature*:

- Woolston, C., & Ball, J. G. C. (2021). Unpicking the rhythms of the Amazon rainforest: (Where I work). *Nature*, *591*, 494–494

Table of contents

List of figures

List of tables

Chapter 1

General Introduction

1.1 Tropical forests: stability and change

Today's angiosperm-dominated tropical forests emerged from complex evolutionary processes over the last 100 million years, in the highly energetic yet relatively stable conditions proximate to the equator (Morley, 2011; Morley et al., 2000). These processes gave rise to more biodiversity (Dirzo and Raven, 2003; Plotkin et al., 2000) and terrestrial biomass (Pan et al., 2011; Saatchi et al., 2011) than is attributable to any other biome on the planet. Through the Holocene (the 11,700 years since the last glacial epoch), these ecosystems contributed to the rise of human agricultural societies by exerting an integral stabilising influence on global climate through their interactions with solar radiation, hydrology and biogeochemical cycles (Artaxo et al., 2022a,b), including their capacity to sequester ($\sim$2.4 Pg C year^{-1}; Pan et al., 2011) and store (200-300 Pg; Avitabile et al., 2016; Baccini et al., 2017; Pan et al., 2011; Saatchi et al., 2011) vast quantities of atmospheric carbon. However, since colonial and industrial periods, acceleration of human encroachment and extractive activities (Lewis et al., 2015; Williams, 2003), increased temperatures, altered precipitation patterns and increased frequency of extreme weather events have imposed novel pressures on tropical forests (IPCC, 2022), potentially switching them from a net sink to a net source of carbon (Baccini et al., 2017; Brienen et al., 2015; Gatti et al., 2021; Hubau et al., 2020; Mitchard, 2018). The resilience and response of tropical forests to these compounding changes (Doughty et al., 2023; França et al., 2020) and positive climatic feedbacks remains highly uncertain, calling for the need for an improved mechanistic understanding of tropical forest function that scales from individual trees to ecosystem (Chave, 2014; Cox et al., 2000; IPCC, 2022). In this thesis, I will explore how emerging technologies can be used to better understand the dynamics of current deforestation, tree growth and mortality, and the leaf cycling of tropical forest trees.

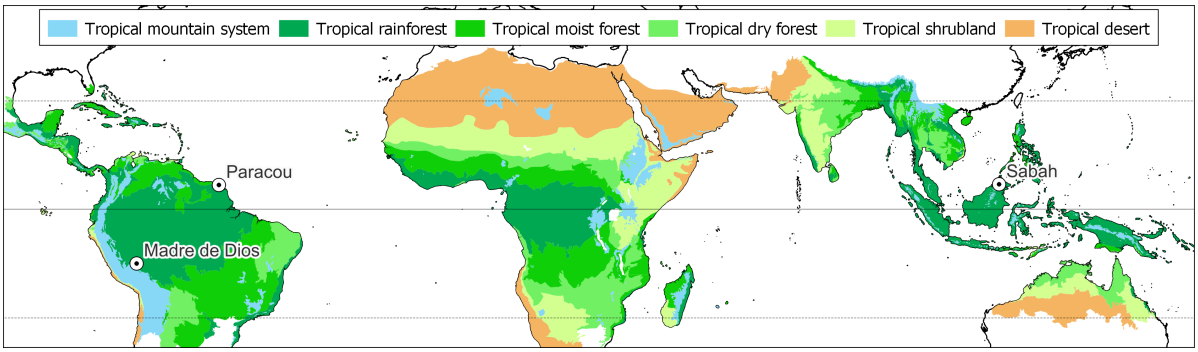


Fig. 1.1 Tropical ecozones (as defined by FAO, 2012) with the location of the sites discussed in this thesis. Note that Paracou has been the site of all primary research.

1.1.1 Biodiversity

Tropical forests are among the most biodiverse systems on the planet, both in terms of absolute numbers and density (i.e. per unit area) of species (Dirzo and Raven, 2003; Gatti et al., 2022; Plotkin et al., 2000). Approximately half of all terrestrial biodiversity is found in tropical forests (Jenkins et al., 2013; Kier et al., 2009) and species richness declines as one moves from the equator to the poles (Hillebrand, 2004). In the case of trees, 67-88% of species are tropical (Beech et al., 2017; Ter Steege et al., 2016), which equates to at least 40,000, and possibly more than 53,000 species (Slik et al., 2015). Part of the challenge in giving precise estimates is the difficulty of quantifying the number of species that have to be described (Gatti et al., 2021; Gatti et al., 2022; Ter Steege et al., 2020). Some regions of the tropics support several hundred species of tree per hectare (Valencia et al., 1994), but their abundance is highly skewed (following an approximately log-normal distribution) with a few common species dominating and most species existing in comparative rarity (Fauset et al., 2015; ter Steege et al., 2013), making them inherently hard to discover. For example, in the Amazon, approximately 1.4-2.3% of species account for half of all trees (Cooper et al., 2024; ter Steege et al., 2013) while 36% of species are estimated to have population size of fewer than 1000 individuals (ter Steege et al., 2013). Climate and its seasonality are thought to play a key role in shaping diversity patterns, with wetter and aseasonal regions of the tropics harbouring greater tree diversity than drier and more seasonal regions (Givnish, 1999; Leigh et al., 2004; Ter Steege et al., 2003). High productivity and metabolic rates in the wetter regions may give rise to enhanced rates of speciation from faster mutagenesis (Rohde, 1992) and relatively stable environmental conditions may support low rates of extinction (Jablonski et al., 2006). Within a particular region, edaphic, topographic and climatic heterogeneity creates niches that select for plants that are adapted to fill them and enables stable co-existence with different species exploiting different resource pools (Zuppinger-Dingley et al., 2014), but this alone cannot account for the

magnitude of diversity observed (Grubb, 1977; Wright, 2002). It has long been recognised that negative density dependence (i.e. the fitness of a species decreases as it becomes more abundant) is critical for the coexistence of tropical diversity, and that pests and disease pressure generate the negative density dependence (Givnish, 1999; Leigh et al., 2004; Ter Steege et al., 2003). More recently, it was realised that stochasticity in the colonisation, extinction and disturbance processes that determine forest dynamics is effective at slowing down extinction rates (Hubbell, 2001; Maire et al., 2012; Rand et al., 2006), and promotes stable coexistence if combined with negative density dependence (Purves and Turnbull, 2010).

Species diversity in tropical forests is thought to be an important determinant of ecosystem functioning (van der Plas, 2019) and ecosystem services (Gamfeldt et al., 2013; Thompson et al., 2014), particularly through its influence on stability (Jactel et al., 2017) and resilience (Schmitt et al., 2022), although the theoretical and empirical underpinnings are debated (Brockerhoff et al., 2017; McCann, 2000; Naeem et al., 2009). Various mechanisms have been suggested to explain the links including niche complementarity, and complementarity of functional effect traits and functional response traits (Isbell et al., 2011). For instance, some plant species facilitate the growth of other plants through co-existence (e.g. nitrogen fixers in nitrogen limited sites), improving the growth of certain tree combinations (Forrester and Bauhus, 2016; Thompson et al., 2014). On the other hand, the "selection effect" posits that having a wider array of species raises the chance of having a particular species with superior growth rates in the prevailing conditions or necessary resistance to disturbance, thus enhancing ecosystem function and service relative to less diverse communities (Lefcheck et al., 2015; Loreau and Hector, 2001; Wardle, 2001). Diversity can increase resistance to disturbances by diluting resources that pests would otherwise exploit and through multi-trophic interactions including increased activity of predators (Jactel et al., 2017). Furthermore, when a community has several species performing the same ecological function (functional redundancy) it is likely to be more resilient to disturbance as the loss of any one species has limited overall impact (Biggs et al., 2020). Broad genetic diversity provides a pool from which traits can be selected (or evolve in the longer term) to adapt to changing conditions such as climate change (Hoban et al., 2021). Ensuring the maintenance of tropical forest biodiversity and the contingent ecosystem services is critical to the wellbeing and livelihoods of 800 million people who live in to next to tropical forests (Keenan et al., 2013), the development of new pharmaceuticals (Calderon et al., 2009), global food security, and meeting global climate goals (IPCC, 2022).

Deforestation is a major threat to tropical forest biodiversity, with many taxa at risk of extinction (Alroy, 2017; Ceballos et al., 2017; Giam, 2017). Climate change is forecast to accelerate this decline (Esquivel-Muelbert et al., 2019) but the anticipated rate of loss is very uncertain, particularly because we are unsure whether sudden, non-linear shifts in

population sizes may occur (Chave, 2014; Laurance, 2007; Wright and Muller-Landau, 2006). Understanding the scale and rate of the shifts requires systems that can monitor compositional changes over large forest extents. Forest plot networks have laid the foundation but they are limited in scale. Further refining our understanding of the role of biodiversity in ecosystem function will allow more efficient allocation of resources and development of policy to mitigate and adapt to these impacts.

1.1.2 Carbon cycle

Gross primary productivity (GPP) of trees is the rate at which they fix CO_2 by photosynthesis. Some of the organic molecules produced by photosynthesis are energy sources used to power a multitude of processes within cells through mitochondrial respiration, releasing CO_2 back to the atmosphere, while other organic molecules become the building blocks for biomass growth. Net primary productivity (NPP) is the rate of accumulation of carbon that has not been respired to fuel metabolic processes. This carbon can stay with the tree briefly (e.g. leaves) or stay locked up in the woody material until a tree dies. From dead matter, carbon may be released back into the atmosphere via (respiratory) decomposition, or stored for a further period in the soil (Malhi, 2012). The balance of these fluxes at the macro-level plays an important role in the global carbon cycle. In particular, the carbon uptake by the world's forests (and oceans) ensured that less than 45% of total anthropogenic emissions from 1960-2015 remained in the atmosphere although the contribution of forests remains poorly quantified (Friedlingstein et al., 2022; Mitchard, 2018).

Accounting for all carbon sources and sinks associated with tropical forests is challenging; the different methods used to calculate them often give diverging estimates which need to be carefully triangulated (Mitchard, 2018; Mitchard et al., 2014). There is thought to be 200-300 Pg C stored in living tropical trees (Baccini et al., 2017; Brienen et al., 2015; Gatti et al., 2021; Hubau et al., 2020; Mitchard, 2018) which capture around 72 Pg C yr^{-1} through photosynthesis (Beer et al., 2010), with slightly less released via respiration across all tropical trophic levels (plants, animals, fungi etc.) and burning (Gaubert et al., 2019; Sitch et al., 2015). For illustration, we can compare deforestation, forest regrowth and intact forests. Deforestation is a clear source of carbon but the magnitude of emissions associated with degradation is less easy to assess; estimates of these emissions are 0.5 - 3.5 Pg C yr^{-1} (Achard et al., 2014; Baccini et al., 2017; Harris et al., 2012; Houghton et al., 2012; Keenan, 2015; Tyukavina et al., 2015; Zarin et al., 2016). In contrast, recovering and regrowing forests are are carbon sinks, but there is limited available data on the magnitude of this sink and a broad range of estimates. The average recovery rate over the first 20 years of regrowth has been estimated to be 3 Mg C ha^{-1} yr^{-1} (for context, the carbon density of an old growth forest was 70-125 Mg C ha^{-1}, Poorter

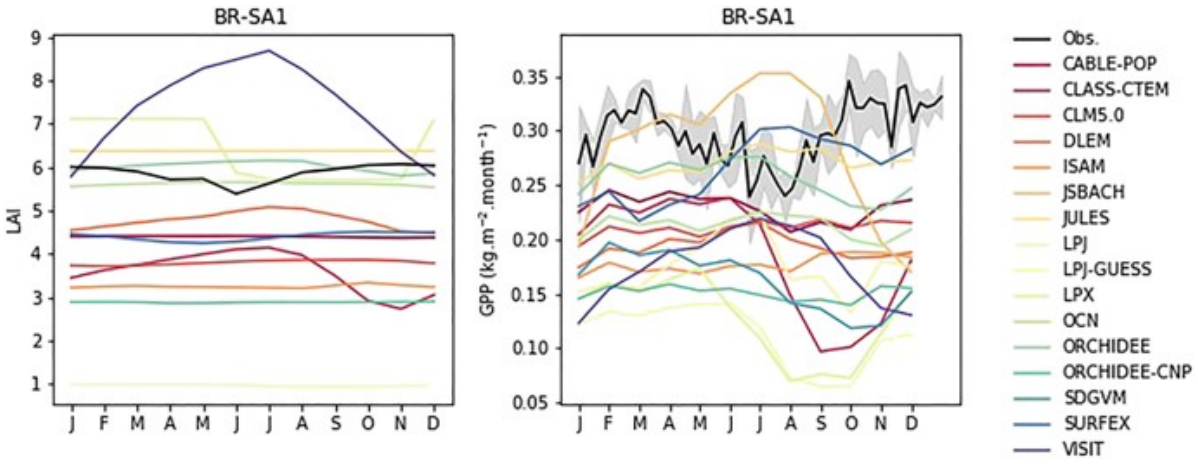


Fig. 1.2 Seasonality of simulated and observed leaf area index (LAI) and Gross Primary Productivity (GPP) at the Amazonian BR-Sa1 site (Tapajós National Forest, Brazil) using 16 land surface models. Simulated LAI shows little seasonal variability throughout the year or drops in the dry season which is in contrast to the camera-based observations of Wu et al. (2016). There is little consistency between simulated and observed GPP values. Source: Chen et al. (2020).

et al., 2016) but it is difficult to determine the pan-tropical extents of disturbed and regrowing forests. Undisturbed tropical forest is more finely balanced. It is likely that they have been an average net sink of $\sim$1 Pg C yr^{-1} (with a high degree of inter-annual variability; Friedlingstein et al., 2022) since at least the late 1970s, supported by enhanced growth rates from atmospheric CO_2 fertilisation (Pan et al., 2011; Phillips et al., 1998; Sitch et al., 2015). However, there is evidence from forest plots that this sink is saturating (most obviously in the Amazon) due to increasing tree mortality (Brienen et al., 2015; Hubau et al., 2020; Qie et al., 2017), and risks flipping to a net source in the near future (Mitchard, 2018). However, these fluxes are all estimated with high uncertainty.

Attempts to better understand and predict the fluxes associated with forests have been made by representing their dynamics mechanistically in Dynamic Global Vegetation Models (DGVMs). DGVMs simulate how vegetation grows, competes, and responds to environmental factors including climate and disturbance, and can be included as components of Earth System Models (ESMs) which additionally represent interactions between atmosphere, oceans and land surface. Such models have suggested that, as a result of increased vapour pressure deficit due to climate change, the Amazon could reach a tipping point beyond which it would begin an irreversible transition to a drier, savanna-like ecosystem (the Amazon dieback scenario), with a substantial further injection of carbon into the atmosphere (Huntingford et al., 2008; Malhi et al., 2009). However, it has since been shown that DGVMs are unable to replicate observed seasonality of carbon fluxes in the Amazon (Restrepo-Coupe et al., 2017) and patterns of leaf area index across tropical rainforests (Zou et al., 2023), failing to reproduce

experimentally measured climate-vegetation feedbacks (Piao et al., 2013), and overestimating historical tree mortality, most overtly in tropical regions (Yu et al., 2022a). These are clearly major shortcomings, calling into question their suitability to make predictions on the future state of tropical forests. One explanation is that current DGVMs suffer from insufficient realism, built on assumptions and frameworks that may not fully capture the diversity and complexity of tropical ecosystems (Yang et al., 2016). In particular, their aggregations of diverse vegetation types into plant functional types (Scheiter et al., 2013) and simplified (or absent) representation of leaf phenology (Chen et al., 2020; Zou et al., 2023) has been blamed for their highly inconsistent predictions. To illustrate this, Fig. 1.2 compares simulated leaf area index (LAI) and GPP from 16 widely used land surface models to observations for a site in the Brazilian Amazon. LAI is the total one-sided area of leaf tissue per unit ground area. This metric quantifies the amount of leaf surface available for photosynthesis, a key control on GPP. The simulated LAIs either showed little seasonal variation throughout the year or a drop in the dry season which was in contradiction the camera-based observations. Similarly the simulated GPP of the majority of models showed a decrease during the dry season contrary to what was observed with the *in situ* eddy covariance flux estimates.

To address these uncertainties and gain a more mechanistic understanding of how tropical forests will contribute to global carbon dynamics under climate change, it is necessary to track the growth, mortality and foliar dynamics of trees at a larger scale than has been possible from the existing forest plot network. By comparing the patterns of tropical forests around the world it will be clearer which systems can continue to buffer against anthopogenic emissions, which are likely to exacerbate increasing atmospheric CO_2 concentrations and which regions should become the focus for restoration efforts to withdraw atmospheric CO_2 (IPCC, 2022).

1.1.3 Leaf phenology

Leaves modulate the flows of energy, carbon and water between the biosphere and atmosphere. Water escapes from the stomatal pores as the CO_2 required for photosynthesis enters. Leaf area determines the amount of light that can be intercepted by the canopy and leaf photosynthetic capacity (carboxylation potential) determines the rate at which leaves are able to fix carbon, and thus the proportion of incident light that can be usefully harnessed (Hall and Rao, 1999). Photosynthetic capacity peaks when leaves are young, around the time of full expansion, and declines as leaves age (Kitajima et al., 1997). Leaf phenology - the timing of leaf flushing, senescence and abscission - therefore plays an important role in structuring ecosystem processes (Lieth, 1974). The timing of leaf production and loss determines the temporal dynamics of the leaf area index (LAI, total leaf area per unit ground area), which is in turn a key driver of primary productivity (Chen et al., 2012; Manoli et al., 2018) and transpiration rates (Teuling

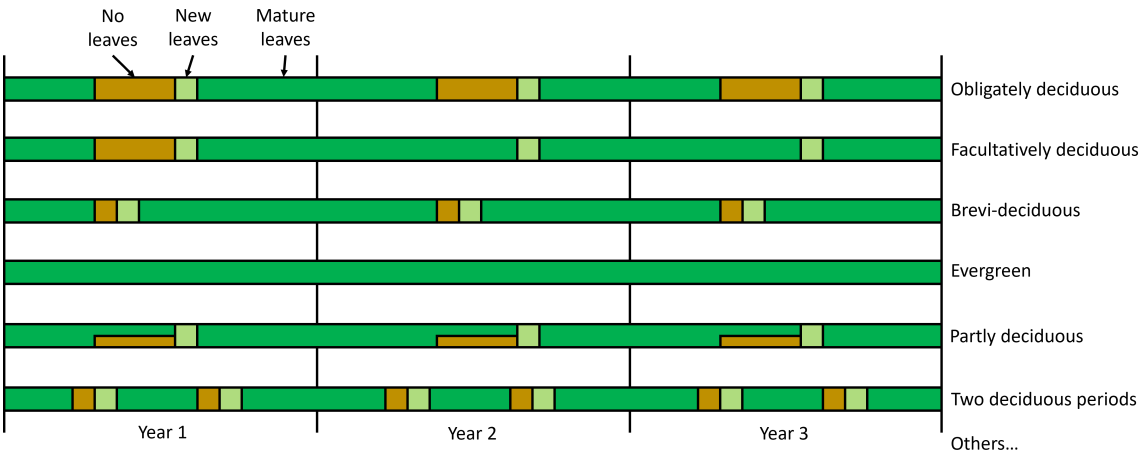


Fig. 1.3 A non-exhaustive illustration of the range of phenological habits present in tropical forests. Adapted from Bohlman (2019).

et al., 2010). Transpiration subsequently influences the availability of soil water resources, local hydrology and regional water cycles (Teuling et al., 2010). Leaf phenology also determines the age distribution of leaves within a canopy, influencing overall photosynthetic rates (Wu et al., 2016, 2017a,b). Additionally, leaf phenology affects multi-trophic species interactions. For instance, leaf emergence impacts herbivore populations which rely on the availability of young leaves (van Schaik et al., 1993), which can have cascading effects on predator populations (Visser and Both, 2005).

Forests in the wetter parts of the tropics are often thought of as aseasonal, but this description masks the dynamic and eclectic nature of tropical leaf phenology (Newstrom et al., 1994). Patterns of whole-canopy leaf phenology vary tremendously across the tropics, from strongly deciduous in the dry tropics to weakly seasonal evergreen in persistently wet tropics (Reich, 1995). Optimised strategies need to balance variation in water availability and insolation. In regions with a long dry period, tropical deciduous forests are common, where it is typical that leaves are shed relatively synchronously during the dry season to protect against water stress and drought (Bullock and Solis-Magallanes, 1990; Lieberman and Lieberman, 1984; Reich and Borchert, 1982, 1984), and fresh leaves put out in anticipation of the wet season (de Camargo et al., 2018; Murali and Sukumar, 1993; Ryan et al., 2017; van Schaik et al., 1993). Wetter forests have a less obvious or clearly defined structure but overall patterns tend to structure canopy leaf production in a way that maximises insolation with fresh leaves put out towards the start of the dry (less overcast) season (Anderson et al., 2011; Doughty et al., 2019; Guan et al., 2015; Lopes et al., 2016; Saleska et al., 2016; Taffo et al., 2019; Wagner et al., 2017). However, within a single wet site there can be a huge amount of variability in terms of leaf

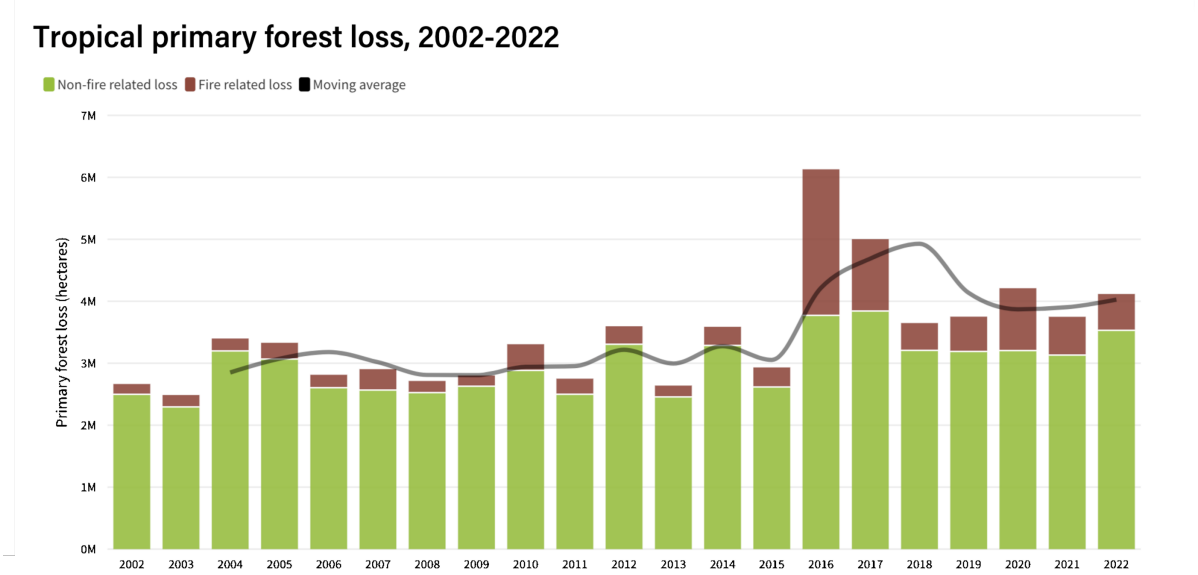


Fig. 1.4 Tropical forest loss 2002-2022. Non-fire related loss can occur from mechanical clearing for agriculture and logging, as well as natural causes such as wind damage and river meandering. The three-year moving average may represent a more accurate picture of the data trends due to uncertainty in year-to-year comparisons. All figures calculated with a 30 percent minimum tree cover canopy density. Source: Weisse et al. (2023).

habit (deciduous-evergreen), and timing and synchronisation of flushing / shedding between species (see Fig. 1.3). Evergreen species, which retain foliage throughout the year, can remain relatively stable or have strong seasonal variation in the rate of leaf turnover (Coley, 1983; Frankie et al., 1974; Kumar et al., 2023). Leaflessness can be part of a predictable annual cycle (obligately deciduous) or only in response to specific conditions or stress (facultatively deciduous) (Harenčár et al., 2022; Stevens et al., 2016). Deciduous trees, can be leafless for just a few days or several months (Loubry, 1994; Williams et al., 2008). Species may synchronise their flushing and shedding while others show completely asynchronous cycles (Borchert, 1980; Morel et al., 2015; Reich and Borchert, 1984, 1988; Wright and Cornejo, 1990). A lack of comprehensive data on the full range of leaf phenology patterns has hindered realistic parameterisation of vegetation models that can be used to simulate the response of tropical forests to a changing climate (Chen et al., 2020; Zou et al., 2023). By better characterising leaf cycling dynamics it will allow for more and improved understanding of the resilience of tropical forests to climate change (see Section 1.1.2).

1.1.4 Deforestation

Humans have cleared forest to create space for agriculture and to extract resources since the Neolithic period; the rate of clearance has tracked population size and technological development (Lewis et al., 2015). The first evidence of agriculture in the tropics dates back some 6,000 years, but the extent of clearance remained relatively small until the arrival of European colonial powers (Ellis et al., 2013; Lewis and Maslin, 2015). Under colonial rule, large areas of lowland forest were harvested for timber and plantations of sugar, tobacco, cotton and oil palm which seeded frontiers of deforestation in the tropical lowlands (Hecht and Cockburn, 2010). Industrialisation accelerated the changes as mechanised processes allowed for more efficient conversion of land. Over the 20th century, industrialisation spread and globalised networks of trade in commodities further accelerated deforestation in productive tropical regions (Lewis et al., 2015; Williams, 2003).

Based on the FAO's Global Ecological Zone dataset, the pre-conversion area of tropical forests has been estimated at 1.46 billion ha (FAO, 2012), of which around 34% has been converted, 30% is in various forms of degradation, and 36% remains intact (Krogh, 2020). Approximately, 380 million ha of tropical tropical forest was lost between 1990 and 2020 alone (FAO, 2020) with the rate of CO_2 emissions from this form of land use change accelerating over that period (0.97 Pg C yr^{-1} in 2001–2005 compared to 1.99 Pg C yr^{-1} in 2015–2019, Feng et al., 2022). In 2021 at COP26, 145 countries vowed in the Glasgow Leaders' Declaration on Forests and Land Use to "halt and reverse forest loss"[1]. The pledge failed to make an immediate impact, as worldwide deforestation (primarily in the tropics) increased by 4% in 2022 (missing the target by 33%; Climate Focus, 2023), and there was a worrying 10% uptick in primary tropical forest loss over the same period (see Fig. 1.4; Weisse et al., 2023). Effective nature-based climate solutions (Griscom et al., 2017) hinge on integrating local and indigenous knowledge (Ajani et al., 2013; Cottrell, 2022; Tengö et al., 2017), and alignment with socioeconomic development to ensure sustainable, equitable, long lasting environmental stewardship (Kanowski et al., 2011; Newton et al., 2016; Nkem et al., 2013). Innovative technical solutions will also be required to target the financial support that has been made available through the UNFCCC framework in a way that can make targets attainable (Climate Focus, 2023; IPCC, 2022; Seymour and Harris, 2019).

Deforestation emerges as a result of complex interactions between human and ecological systems, and its spread is therefore inherently difficult to predict (Geist and Lambin, 2002; Seymour and Harris, 2019). Causes may be separated into those that are *proximate*, such as clearing land for agriculture and infrastructure, and those that are *underlying*, which can include the high level-demographic (i.e. population growth), economic, political, institutional and

[1] see https://ukcop26.org/glasgow-leaders-declaration-on-forests-and-land-use/

cultural forces driving the need to convert land. Agriculture accounts for 70-80% of tropical deforestation (Geist and Lambin, 2002; Gibbs et al., 2010; Hosonuma et al., 2012) and as much as 96% of cumulative deforestation since 1840 (Geist and Lambin, 2002). Through classification of remote sensing data, Curtis et al. (2018) identify that 27% of global forest loss is attributable to permanent land use change for commodity production whereas 24% is from shifting agriculture. Timber extraction is the most prevalent driver of degradation in Latin America and Asia whereas fuel consumption is the most important driver in Africa (Hosonuma et al., 2012). While less substantial in terms of overall scale, mining (legal and illegal) and oil extraction present a unique threat by seeding new frontiers of forest loss in otherwise intact forest (Finer et al., 2015, 2008). Roads and settlements are established to reach newly discovered resources around which agriculture and timber extraction may then establish and expand (Giljum et al., 2022). Deforestation drivers can be highly dynamic (Curtis et al., 2018; Seymour and Harris, 2019) and many developing countries struggle to keep pace (e.g through effective monitoring), citing weak forest sector governance and institutions, and illegal activity as key underlying drivers of forest loss in their jurisdictions (Ken et al., 2020; Kissinger, 2020). Readily available information on the emergence and spread of threats can support more effective enforcement practices and targeted interventions.

1.2 Remote sensing of tropical forests

The advent of remote sensing technology has made it possible to view and track changes across tropical forests in near real time (Fassnacht et al., 2023; Hansen et al., 2013). These technologies have become the backbone of worldwide efforts to assess trends in tropical deforestation (Hansen et al., 2013), and map carbon stocks and fluxes (Baccini et al., 2017; Harris et al., 2021; Saatchi et al., 2011). Satellite data have facilitated this global approach, but assessments are necessarily inferences from the aggregated response of the organisms involved. When there is a subtle signal there can be ambiguity and confusion in evaluating the ecological implications. This was highlighted by the debate over whether the evergreen Amazon underwent a “green-up” during the dry season. While it appeared that Amazonian vegetation became more photosynthetically active during the dry season (Huete et al., 2006; Myneni et al., 2007; Saleska et al., 2016; Wang et al., 2020), some argued that the signal was artifactual, a result of imperfect directional and atmospheric corrections, rather than a meaningful biological response from the trees (Morton et al., 2014; Samanta et al., 2010). Consensus has begun to form around the likelihood that tropical vegetation, when not water limited, will time leaf production to maximise insolation (i.e. for during drier, less overcast periods; Wagner et al., 2017) but without up-close confirmation of the range of leaf patterns

this explanation remains incomplete. Plane mounted sensors have been available for decades but remain prohibitively expensive for most, especially if repeat observation is necessary. UAVs and miniaturised sensors have heralded a new era of forest monitoring with repeatable scanning that can discriminate between individual trees now available for budgets of hundreds of dollars (Sun et al., 2021).

When discussing remote sensing it is important to consider spatial, temporal and spectral resolution (and range), and their associated tradeoffs and costs (Jones et al., 2010). Spatial resolution is typically given as a ground or surface length across an individual pixel of the retrieved data (e.g. 30 m resolution would be a pixel that represents a 30 m x 30 m of the Earth's surface). Temporal resolution refers to the frequency at which a given area is scanned by a sensor. Spectral resolution refers to the number of distinct, separable wavelength bands a scene is observed in. Sensors can be mounted on satellites, aeroplanes and unmanned aerial vehicles (UAVs). Satellite mounted sensors typically have global coverage, a revisit time between days and weeks (not considering geostationary orbits), and can be supported by public and commercial funding due to their broad appeal, but they tend to have a lower spatial resolution than aircraft mounted sensors due to their distance from the Earth, and analysis of the data retrieved may be complicated by cloud cover and the need for atmospheric corrections. On the other extreme, UAVs have a limited range (greater than a traditional forest plot but not generally more than a typical research site), but they are able to fly close to the top of the canopy, reducing atmospheric interference and capturing sub-crown details down to the branch, leaf or even flower level (Araujo et al., 2021). Furthermore, they can be acquired at a modest cost (around £250 for an entry-level RGB ready drone) allowing a very dense time series to be collected at given area. With the exception of Chapter 2, in this thesis I use sensors mounted on aircaft (aeroplane, helicopter, UAV) with the aim of observing and studying individual trees. Here I provide an overview of the remote sensing approaches employed in the research described in this thesis.

1.2.1 Optical and Multispectral

The simplest form of vegetation remote sensing is standard (visible) digital photography which is based on red, green and blue (RGB) bands that approximately correspond to how the human eye perceives colour. Fatefully, the human eye evolved to be centered on and most sensitive to green light in part due to its association with vegetation (Yokoyama et al., 2014), which means RGB imagery can be usefully employed to assess some aspects of vegetation state. As it is based on the most widespread technology it tends to have the greatest spatial resolution for the least cost of all sensors. This allows for clear visual separation between trees and easy transfer of computer vision methods developed for the same modality. Optical remote

sensing may additionally include near-infrared (NIR), and shortwave infrared (SWIR), thermal infrared (TIR) portions of the electromagnetic spectrum, parts that can particularly useful for assessing vegetation health and chlorophyll content (Jones et al., 2010). Multispectral remote sensing is a flexible term that includes RGB and optical domains but may further increase the spectral resolution by including additional bands of longer (NIR, SWIR, micro, radio) or, less commonly, shorter (ultraviolet) wavelengths. It typically consists of no more than 20 spectral bands. In general, cost and payload increase and spatial resolution decreases with the number of spectral bands.

1.2.2 Hyperspectral data and tree species identification

Hyperspectral remote sensing, or imaging spectroscopy, typically extends the number of bands by an order of magnitude (100+) and is usually more costly with a heavier payload as the wavebands are more numerous and sensitive within narrower wavelength ranges, requiring more semi-conductor materials (Kalacska and Sánchez-Azofeifa, 2008). This has limited its uptake in UAV systems although a new generation of miniturised sensors may be about to change that (see e.g. HySpex[2]). In remote sensing of vegetation, it is typical for these sensors to cover wavelengths in the range 400-2500 nm but some spectral regions may be removed from analysis due to atmospheric moisture absorption which strips the regions of any useful signal (Laybros et al., 2020). The narrow waveband sensitivity (typically less than 10 nm wide) allows direct detection of absorption and reflectance markers of specific chemical (including pigmentation) and morphological properties of the target (Kalacska and Sánchez-Azofeifa, 2008) which might be missed by the broader bands of multispectral sensors (typically tens to hundreds of nm). Broadly, the visible region (400-700 nm) is informative of chlorophyll absorption; the red-edge region (700-780 nm) is informative of vegetation health and chlorophyll content; the near-infrared region (NIR, 780-1300 nm) gives information on leaf structure; the shortwave infrared region (SWIR, 1300-2500 nm) provides information on moisture content (Jones and Vaughan, 2010; see Fig. 4.5).

Hyperspectral remote sensing has previously been used for mapping tree species in forests harnessing spectral differences between species (Fassnacht et al., 2016; Ghiyamat and Shafri, 2010). The spectral properties of plants are based on their biochemical and structural characteristics which can vary in subtle or obvious way, and present a spectral “signature” with which to identify a species (Meireles et al., 2020; Ustin et al., 2004). Species, especially those closely related, often exhibit shared chemical, physiological, and morphological traits, influencing their spectral properties (Agrawal, 2007; Cavender-Bares et al., 2016; Meireles et al., 2020).

[2] https://www.hyspex.com/hyspex-turnkey-solutions/uav/

For example, variations in leaf structure and chemical composition, such as chlorophyll and cellulose levels, can impact these spectral characteristics between tree families (Serbin et al., 2016). Despite these family-level similarities, considerable crown-level spectral variability exists within the same family due to genus and species differences, environmental factors (like soil type, water availability, and light exposure), canopy structural variation, and plant health or phenological stage (Ollinger, 2011). Therefore, while plant families may share spectral similarities, these are not definitive for species classification.

Identifying species in diverse tropical forests is challenging due to the number of species and their skewed abundance distributions. Clark et al. (2005) used AVIRIS data to differentiate seven species in a Costa Rican tropical rainforest demonstrating that differentiation between tropical tree species was possible. Féret and Asner (2013)'s work distinguishing 17 species in Hawaii's humid tropical forest, and Laybros et al. (2020, 2019)'s demonstration of an 80% classification rate for 20 species in an Amazonian forest further highlighted the potential to classify more species but still a small fraction of the species diversity that can be present. Notably, discrimination between more species has been achieved in leaf-level hyperspectral species identification, such as 46 species in a Jamaican tropical wetland (Prospere et al., 2014). This success suggests aerial hyperspectral data can extend beyond the limit of around 20 species. However, leaf traits are reflected in spectra in complex ways with spectral regions often encompassing multiple traits (Féret and Asner, 2011; Jacquemoud and Baret, 1990). This complexity increases when spectra are measured remotely as atmospheric effects, canopy structure and morphological characteristic become significant. Indeed, radiative transfer models have shown that signals from canopy structure can dominate over the leaf optical properties and biochemical properties of the vegetation (Béland and Kobayashi, 2024; Knyazikhin et al., 2013). Deciphering these spectral signals to understand evolved traits and species relationships remains a significant challenge in achieving robust, transferable remote identification of species in diverse tropical forests (Schweiger et al., 2021).

1.2.3 Lidar: 3D structure

Lidar differs from the previously discussed approaches in that it is *active*, sending out pulses of light and measuring the reflections to generate a three dimensional point cloud of its target rather than passively receiving reflected solar photons. It has the capacity to penetrate forest canopies and provide three-dimensional data on forest structure (Mazlan et al., 2022). Airborne LiDAR is especially useful for estimating aboveground biomass in tropical forests (Coomes et al., 2017). LiDAR remote sensing has been shown to be capable of quantifying substantial disturbance tropical forest canopies but it has proven more challenging to detect small changes in old-growth biomass at a fine scale including LAI variation (Dubayah et al., 2010). This is

partly due to the difficulty of accounting for extinction through the canopy and difficulty in comparing outputs from different sensors (Cao et al., 2023; Vincent et al., 2023). Tools have been developed for converting lidar pointclouds into PAI estimates based on the interaction of the laser pulses as they travel through the canopy (Vincent et al., 2017, 2021). It is yet to be seen whether these can help tracking changes in leaf quantity through time.

1.2.4 Limitations and the need for high quality ground truth data

Remote sensing allows us to view vast tracts of the Earth's surface, but without ground based reference points it is difficult to make sense of what we are looking at. Data collected from a remote sensor does not translate directly into biological properties, processes and mechanisms (Cavender-Bares et al., 2022). Ground truth data is necessary for calibrating, validating and correcting remote sensing measurements so that can be used to infer the characteristics of the tropical forest, such as vegetation type, canopy height, and land cover. Additionally, as we become more reliant on machine learning as a means to process and interpret remote sensing scenes, the need for open, large, high quality datasets can allow for better accuracy and transferability of models. For instance, without an expanded base of robust, carefully designed field measurements of forest biomass it will not be possible to improve remote sensing based estimates of carbon stocks and fluxes across the tropics (Chave et al., 2019; Mitchard et al., 2014). Spectral features of data cannot be used naively to infer biodiversity but must be calibrated against local surveys (Badourdine et al., 2023; Cavender-Bares et al., 2022). If models are trained on data that represents limited environmental space they will not transfer well to make predictions at other places or times.

Deep learning has the potential to locate and identify species trees from remote sensing data but it is yet to be scaled effectively in the tropics (Beloiu et al., 2023; Ma et al., 2024; Mäyrä et al., 2021). While there is much plot inventory / census data available, most of it is of limited use for training machine learning models based on remote sensing data as it does not necessarily translate directly into what is observed from above the canopy. Tropical forests often have dense, complex canopies and the crowns that are visible from above need to be carefully matched to the inventoried trunks in the field. By building maps of tree crowns that can be overlaid on lidar, multispectral and hyperspectral datasets, it may be possible to teach AI systems how to perform the time consuming tasks of delineating tree crowns and assigning a species identity them. This could considerably expand the number of trees under observation in the tropics beyond what is currently possible with forest plot networks. The few open source benchmark datasets that are available to help produce such methods are restricted to temperate regions (Ahlswede et al., 2023; Weinstein et al., 2020).

1.3 Application of AI to remote sensing of tropical forest ecology

Through a confluence of theoretical, software and hardware advances, deep learning enabled computer vision emerged in earnest in the 2010s. An initial breakthrough came in 2012 when a deep convolutional neural network (CNN) called AlexNet won the ImageNet Large Scale Visual Recognition Challenge by a considerable margin, showcasing the power of the technology and spurring increased interest and investment in the field (Krizhevsky et al., 2017). Key to its success was the use of graphics processing units (GPUs) for training. GPUs were initially designed for gaming but their potential for parallel processing made them ideal for training deep neural networks. Nvidia and other companies started optimizing GPUs for AI tasks in the early 2010s, significantly speeding up AI research and application. Alongside this, the rise of "big data" from social media, cheap sensors, and the broader adoption of the internet provided the vast datasets necessary to make use of the flexibility provided by the depth of the new networks. The trend towards open-source software for machine learning, like Theano (in 2007) (Bergstra et al., 2010), TensorFlow (in 2015)[3], and PyTorch (in 2016)[4], allowed for widespread collaboration and sharing of AI advancements. The maturation of cloud computing platforms such as Amazon Web Services, Microsoft Azure, and Google Cloud Platform provided the infrastructure for scalable AI by offering high computational power on demand. A natural application of this technology was on remote sensing data (Hoeser and Kuenzer, 2020; Ma et al., 2019; Zhu et al., 2017). The volume of available data was growing faster than it could be usefully analysed and the field was calling for approaches that could automate feature extraction and other processing steps. More recently, interest in CNNs for extracting information on vegetation characteristics and dynamics for agriculture, forestry and conservation has been growing (Kattenborn et al., 2021).

1.3.1 Convolutional neural networks

Deep convolutional neural networks (CNNs) preserve spatial structures and can automatically extract features of images making them extremely useful for computer vision applications (Voulodimos et al., 2018). There are several reasons why CNNs have proven so effective in the analysis of images (Goodfellow et al., 2016):

- **Feature Learning**: Image recognition systems prior to deep learning required manual feature extraction. Experts needed to identify features that were key to differentiating

[3] https://www.tensorflow.org/
[4] https://pytorch.org/

between types of images and then program these features into the system. CNNs, however, learn to recognize these features automatically through the training process. This shift from handcrafted feature design to learned features is a core part of what made CNNs revolutionary.

- **Layered Architecture**: CNNs employ a hierarchical, multi-layer architecture that allows the network to process data at various levels of abstraction. For instance, the first layers might detect edges or colors, intermediate layers might identify textures or patterns, and deeper layers might represent more complex features that allow networks to recognize complex objects (this layering is given for illustration but in many real applications the true representations are less intuitive). This hierarchical processing mimics some aspects of the human visual system and allows CNNs to build up an understanding of images in a way that is both efficient and effective.

- **Shared Weights and Locality**: The convolution operation used in CNNs involves a filter or kernel that passes over the input image. The same filter is used across the entire image, which means that the network has fewer parameters to learn (shared weights). This makes CNNs particularly suited for image data, as they can detect features regardless of their position in the input space. Moreover, the use of local receptive fields focuses on local areas, preserving the spatial relationships within the data.

- **Pooling**: CNNs often include pooling layers that reduce the spatial size of the representation, reducing the number of parameters and computation in the network. This down-sampling helps make the detection of features somewhat invariant to scale and translation, increasing the robustness of the model.

- **End-to-End Learning**: With CNNs, an end-to-end learning process is established, where raw data can be inputted into the network, and feature extraction and classification can be handled in one pipeline. This simplifies the process of developing and training models for complex computer vision tasks.

- **Transfer Learning**: CNNs trained on large datasets can be used as starting points for other vision tasks. The features learned by the networks on one task can often be transferred to another task with limited additional training, making CNNs highly adaptable and efficient for a wide range of applications.

CNNs have generated novel insights for ecologists working with Earth observation data (Brodrick et al., 2019) including work on detecting logging trails (Abdi et al., 2022), human settlements (Corbane et al., 2021), deforestation (Torres et al., 2021), forest disturbance (Kislov

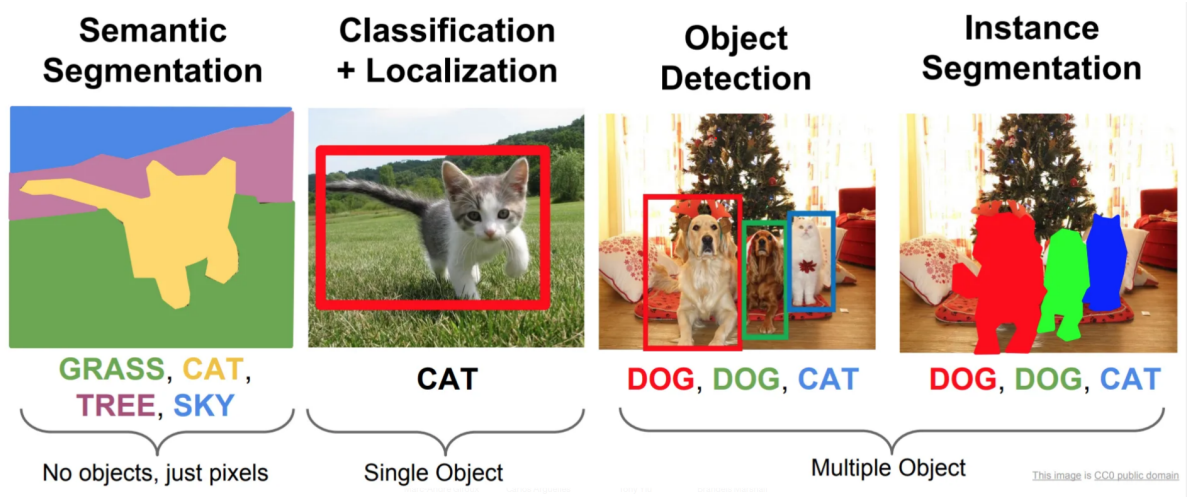


Fig. 1.5 Types of images prediction tasked performed by CNNs. Instance segmenatation designates a class to each pixel of an image. Classification determines what object present in and image and localisation determines where within the image that object is present. Object detection can locate various objects within an image. Instance segmentation not only finds the relevant images but provides exact delineations for each of them. Source: Agarwal (2019).

et al., 2021), quantifying the properties of vegetation (Kattenborn et al., 2021) and mapping the carbon stocks in trees (Mugabowindekwe et al., 2022). However, they may not be as effective as simpler methods when data is limited. Their flexibility comes in part from the fact they can be trained to perform a variety of tasks (see Fig. 1.5). Often it is necessary to classify each pixel in a scene based on its class (semantic segmentation) such as in land use classification but in other cases it might be preferable to locate individual objects such as trees. With respect to individual tree mapping, object detection and localisation is an established method (Weinstein et al., 2019) but the output bounding box can include portions of other trees as well as the target tree which risks mixing of signals. Instance segmentation in which each tree crown in an image is located and exactly delineated is an active area of research (Braga et al., 2020). The Mask R-CNN (He et al., 2017) algorithm has shown promise in tree crown identification and delineation in plantations (Hao et al., 2021; Yu et al., 2022b), pine forests (Hu et al., 2022; Ocer et al., 2020), urban woodlands (Ocer et al., 2020) and simulated tropical forest fragments (Braga et al., 2020). Given the right training data, it seems likely that it could help to scale up tree crown mapping in intact tropical forest landscapes.

1.3.2 Performance evaluation

In evaluating deep learning and classification models, selecting appropriate performance metrics is essential, as each offers distinct insights into a model's strengths and weaknesses.

Accuracy is straightforward, quantifying the overall proportion of correct predictions, though it may be misleading in imbalanced datasets. Precision and Recall address specific types of prediction errors; precision is key when false positives carry high costs, while recall is crucial where missing actual positives (false negatives) is costly (see Fig. 1.6). The F_1 Score helps balance these aspects by combining precision and recall into a single metric, useful when both types of errors are equally important. The AUC-ROC curve, which measures a model's ability to distinguish between classes, is particularly valuable in binary classification scenarios, especially with imbalanced data. However, its effectiveness might be overstated in severely skewed datasets. A more granular approach is provided by the Confusion Matrix, which details the numbers of true positives, false positives, true negatives, and false negatives across classes, offering a comprehensive view of model performance. For evaluating thresholds in predictions, the Precision-Recall Curve is advantageous, especially in contexts where positive class prevalence is low. In regression tasks, Mean Absolute Error (MAE) and Mean Squared Error (MSE) provide error magnitude metrics, with MAE being more robust to outliers and MSE emphasizing larger errors more heavily. Log Loss (or Cross-Entropy Loss) evaluates classifiers by penalizing false certainties, being particularly useful for models outputting probabilistic interpretations. Ultimately, the choice of metric depends on the specific requirements of the task, the nature of the data, and the implications of different error types, often necessitating a combination of these metrics for a well-rounded evaluation.

A range of metrics are employed in this thesis, selected based on the task and the criteria by which we are judging an acceptable performance. To judge performance, a model might be compared to a baseline or benchmark, which can be a simple random guess classifier but more often is a commonly used or state-of-the-art algorithm (see Chapter 2 and Gan et al., 2023). It is crucial to understand the limits of transferability of a model. For example, can a model trained in one location or during a specific period of time be transferred elsewhere in space or time without losing its ability to make accurate predictions. It is rare to achieve a model with perfect accuracy when working with ecological data, but it is important to understand the types of error that are generated so that the weaknesses of a model can be understood. For example, does an algorithm that classifies species based on remote sensing data more often confuse more closely related species than more distantly related species (see Chapter 4)? Understanding errors in this way can be done with a Confusion Matrix. On the other hand, if one is trying to count trees with a deep learning model that detects trees in aerial imagery, an effective model would be one that is not biased towards over-segmenting (splitting a true crown into several predicted crowns therefore counting more trees than there actually are) or under-segmenting (combining several true crowns into a single predicted crown therefore counting less trees than there actually are) (see 3). Exploring this potential bias can be done by systematically comparing errors in the

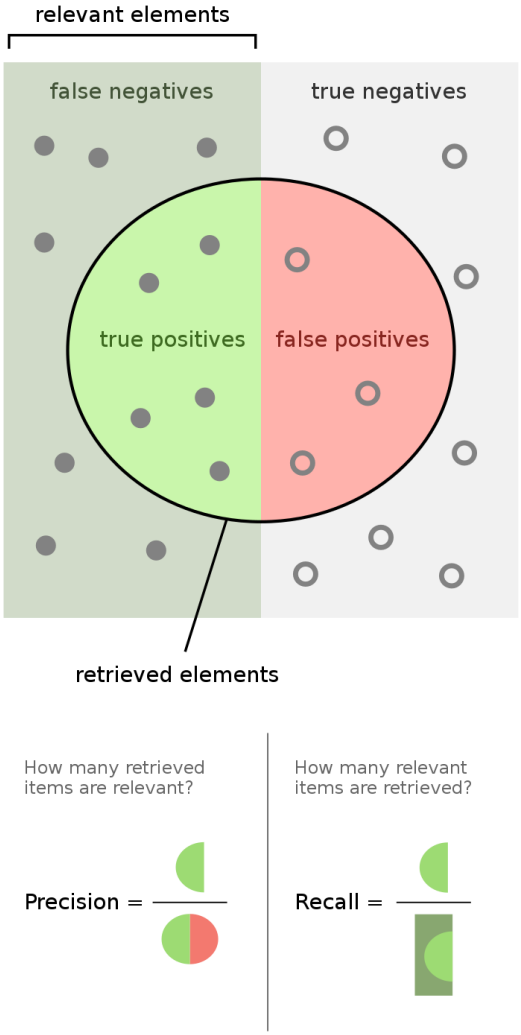


Fig. 1.6 Illustration of precision and recall in terms of false/true negatives/positives for a binary classification problem. Precision is the fraction of relevant instances among the retrieved instances. Recall is the fraction of relevant instances that were retrieved. Achieving a balance between the two is typical for machine learning classification problems. Source: Wikipedia contributors (2024).

predictions to the ground truth data. Crucially, when assessing a model one needs to relate performance back to practical considerations and the domain-specific requirements. A given level of accuracy may be acceptable for one application but wholly insufficient in another. In conservation, for example, there are often costs associated with intervening to protect a given area of land from deforestation (e.g. by sending out patrols). An effective model in this case may be one that can target conservation in such a way so that a minimum threshold of forested area can be protected within a given budget. On a practical note, deep learning models can require extensive resources to train (e.g. powerful GPUs) so an effective model may be one that is not prohibitively expensive to train for a typical conservation practitioner.

1.4 Thesis aims

The overall aim of this thesis was to explore the capabilities and limitations of deep learning methods as applied to remote sensing data to extract information and make predictions on tropical forest dynamics. The term *dynamics* is used in a broad sense to apply to forest loss, tree growth and mortality, and leaf phenology. More specifically, I ask these questions:

1. Can CNNs automatically extract features of satellite and geospatial data that can be useful in making predictions on the spatial location of future deforestation?
2. Can CNNs be used to map individual trees in dense forest landscapes from UAV data?
3. To what extent can tree species be mapped from hyperspectral data across these diverse forest landscapes?
4. Can automatic tree crown maps be used to track growth, mortality and leaf phenology of individual trees across tropical forest landscapes?
5. What patterns of leaf phenology are observed at a wet tropical site?

Each Chapter poses more specific research questions and the explores the topics in detail.

Chapter 2

Using Deep Convolutional Neural Networks to Forecast Spatial Patterns of Amazonian Deforestation

Abstract

Tropical forests are subject to diverse deforestation pressures while their conservation is essential to achieve global climate goals. Predicting the location of deforestation is challenging due to the complexity of the natural and human systems involved but accurate and timely forecasts could enable effective planning and on-the-ground enforcement practices to curb deforestation rates. New computer vision technologies based on deep learning can be applied to the increasing volume of Earth observation data to generate novel insights and make predictions with unprecedented accuracy. Here, we demonstrate the ability of deep convolutional neural networks (CNNs) to learn spatiotemporal patterns of deforestation from a limited set of freely available global data layers, including multispectral satellite imagery, the Hansen maps of annual forest change (2001-2020) and the ALOS PALSAR digital surface model, to forecast deforestation (2021). We designed four model architectures, based on 2D CNNs, 3D CNNs, and Convolutional Long Short-Term Memory (ConvLSTM) Recurrent Neural Networks (RNNs), to produce spatial maps that indicate the risk to each forested pixel ($\sim$30 m) in the landscape of becoming deforested within the next year. They were trained and tested on data from two $\sim$80,000 km^2 tropical forest regions in the Southern Peruvian Amazon. The networks could predict the location of future forest loss with F_1-score = 0.58-0.71. Our best performing model (3D CNN) had the highest pixel-wise accuracy (F_1-score = 0.71) when validated on 2020 forest loss (2014-2019 training). Visual interpretation of the mapped forecasts indicated that the network could automatically discern the drivers of forest loss from the input data. For example, pixels around new access routes (e.g. roads) were assigned high risk whereas this was not the case for recent, concentrated natural loss events (e.g. remote landslides). CNNs can harness limited time-series data to predict near-future deforestation patterns, an important step in harnessing the growing volume of satellite remote sensing data to curb global deforestation. The modelling framework can be readily applied to any tropical forest location and used by governments and conservation organisations to prevent deforestation and plan protected areas.

2.1 Introduction

To achieve the pledge made by world leaders at COP26 to end deforestation by 2030, innovative approaches to forest protection are urgently required. Previous efforts to curb tropical primary forest loss have failed to have a net positive impact (Potapov et al., 2017) despite countries and companies making substantial commitments (e.g. the New York Declaration of Forests in 2014). In regions of the Amazon deforestation is surging (Beuchle et al., 2021) which risks inducing the biome to undergo a critical transition with profound consequences for climate and biodiversity (Lovejoy and Nobre, 2019). While the ultimate causes of deforestation need to be addressed – notably global demand for agricultural and wood products – interventions to tackle proximate causes, such as illegal logging and mining, are also necessary (Asner and Tupayachi, 2017; Finer et al., 2014). One measure identified as effective in reducing deforestation is "targeting protected areas to regions where forests face higher threat" (Busch and Ferretti-Gallon, 2017). However, knowing how to optimally allocate limited resources to tackle diverse threats across vast, difficult to access tropical forest landscapes is challenging. To enable effective, on-the-ground prevention measures, up-to-date information on the location, relative severity and likely evolution of threats is required.

Governmental and NGO commitments to make timely interventions have led to products that give near real time alerts on the location of deforestation events (Hansen et al., 2016; Reiche et al., 2021) , but access to these alerts has had little material benefit in terms of curbing deforestation (Moffette et al., 2021). One issue has been that responses based on near-real-time mapping can only ever be reactionary. As a result, there have been calls to develop early warning systems that inform decision makers of the location of near-term deforestation risk (e.g. WWF, 2020). By developing innovative solutions that harness the growing volume of remote sensing (RS) data, interventions stand a greater chance of preventing deforestation. In the longer term, cost effective conservation plans should not only consider the spatial distribution of conservation features (e.g. species, carbon stocks) and the financial costs, but also account for how threats are likely to evolve (Boyd et al., 2015; Wilson et al., 2007).

Deforestation is difficult to predict as it results from complex interactions within human-ecological systems but characteristic drivers, such as agricultural expansion and ease of access, have been identified (Geist and Lambin, 2002; Lim et al., 2017; Miyamoto et al., 2014; Ritchie and Roser, 2024). Drivers can be mapped to spatially resolved geographical, economic, social and biological variables (e.g. agricultural land value, distance to roads). Within a statistical or machine learning (ML) framework, drivers can be correlated to the likelihood of deforestation at a given location, typically by linking the spatial predictor layers to changes in forest extent over a single time step of several years (see Cushman et al., 2017; Mayfield et al., 2017; Saha et al., 2020 for intercomparisons of different ML frameworks). However, Rosa et al. (2014)'s

review of such approaches concluded that they had a limited or poorly defined ability to forecast locations of deforestation, a key reason being that spatially explicit datasets on the drivers of deforestation are often incomplete, out-of-date or unreliable in tropical forest regions. For example, proximity of roads is widely accepted to be a strong predictor of future deforestation (Barber et al., 2014) but roads in tropical forest landscapes are often unofficial, illegal or not comprehensively mapped (Perz et al., 2007), and can be highly dynamic; roads appear, expand and change course to access new resources (Ahmed et al., 2014). Most studies that have included roads as predictors used static road maps which is inappropriate in fast changing contexts, while others have used unvalidated or speculative methods (such as least cost path finding) to predict the development of roads in tropical forest landscapes (Ahmed et al., 2014; Mena et al., 2017). With many of the decisions that lead to tropical deforestation likely to remain hidden from public view, a fundamentally new approach to making predictions of the spread of deforestation is required. Analogous issues exist for other landscape features that help predict deforestation including the location/spread of human settlements and agricultural land.

The ML frameworks used to date have been limited in how they represent local context. Without the ability to retain the spatial structures of the data, pixels are represented in isolation or local context is represented as averaged neighbourhood metrics. Contrasting approaches that identify threatened regions based on dynamic data (i.e. resolved over numerous time steps) have tended to focus simply on spatial and temporal trends in intensity of deforestation (e.g. the emerging hot spot approach of Harris et al., 2017). While these approaches are easy to implement and update, they are naïve to the drivers of the observed forest loss and therefore unable to reliably predict whether a deforestation front is likely to spread. This risks inefficient targeting of conservation resources. A forecasting approach that could automatically learn and detect the contextual features of the landscape that signal deforestation risk from satellite imagery, and update its predictions accordingly, would bypass the limitations of both exiting ML and simple trend based approaches.

Through implementing modern computer vision approaches, it may be possible for an artificial neural network to automatically learn what features of a landscape are indicative of future deforestation from historic events and thereby avoid the need to rely on problematic data layers. Deforestation often exhibits striking spatiotemporal patterns when viewed from space (e.g. fishbone pattern; de Filho and Metzger, 2006). The emergence of these patterns suggests a degree of predictability while their characteristic spatial and spectral structures can give clues as to the drivers. Over recent years, the quantity of freely available, high-resolution imagery has been increasing at a rate that has exceeded our ability to harness it to support decision making. However, deep convolutional neural networks (CNNs), that can retain spatial structure and be

trained to automatically extract features of images, have revolutionised the field of computer vision (Voulodimos et al., 2018). They have led to radical advances in our ability to make predictions in countless fields and have recently been taken up by RS researchers (Zhu et al., 2017) and ecologists working with Earth observation data (Brodrick et al., 2019) to generate novel insights. This includes work on detecting logging trails (Abdi et al., 2022), human settlements (Corbane et al., 2021), deforestation (Torres et al., 2021), forest disturbance (Kislov et al., 2021) and quantifying the properties of terrestrial vegetation (Kattenborn et al., 2021). Importantly, however, these studies have focused on assessing the current state of systems rather than predicting future states.

Forecasting deforestation is a temporal extension of the landcover classification task. 2D CNNs (designed for working with image data; Kussul et al., 2017) and networks that incorporate an additional dimension (e.g. time) implicitly in their characterisation of data, including 3D CNNs (Li et al., 2017) and recurrent convolutional neural networks (ConvRNNs; Interdonato et al., 2019), have been shown to have state-of-the-art accuracy in detecting and classifying landcover change. Similarly, Convolutional Long Short Term Memory (ConvLSTM) based RNNs have been shown to improve on the performance of ConvRNN for RS classification tasks (Rußwurm and Körner, 2018a). These promising developments suggest that 2D CNNs, 3D CNNs and ConvRNNs architecture types, if trained on large volumes of spatially resolved historic deforestation data, could characterise and learn the features of the imagery associated with changing forest cover and be used to predict the future location of deforestation (something that has not previously been attempted). Critically, they preserve and work with the spatial (and temporal) structures of the input datasets and therefore make predictions based on the local context of the scenes they are presented with.

This paper describes a range of deep CNN model architectures that we designed to predict the risk of forest loss at a given pixel (30 m) when presented with a multi-layered scene (of data from freely available, global datasets) centred on that pixel, with the aim of automatically identifying areas threatened by deforestation. To test these networks, we trained and evaluated their performance on two regions in the Peruvian Amazon, one of the most biodiverse regions on the planet and exceptionally rich in endemic amphibians, birds, fishes, bats, and trees (Bass et al., 2010; Jenkins et al., 2013) but facing a range of threats including deforestation driven by copper and gold mining, logging, agriculture, cattle ranching and crude oil extraction (Finer et al., 2008; Piotrowski, 2019). We trained and tested the models with cloud free satellite data for 2014-2020 and spatially resolved annual deforestation data for 2001-2020 (Hansen et al., 2013). The best performing model was used to forecast deforestation in across regions in 2021, predicting the risk of deforestation for every forested pixel in the landscape.

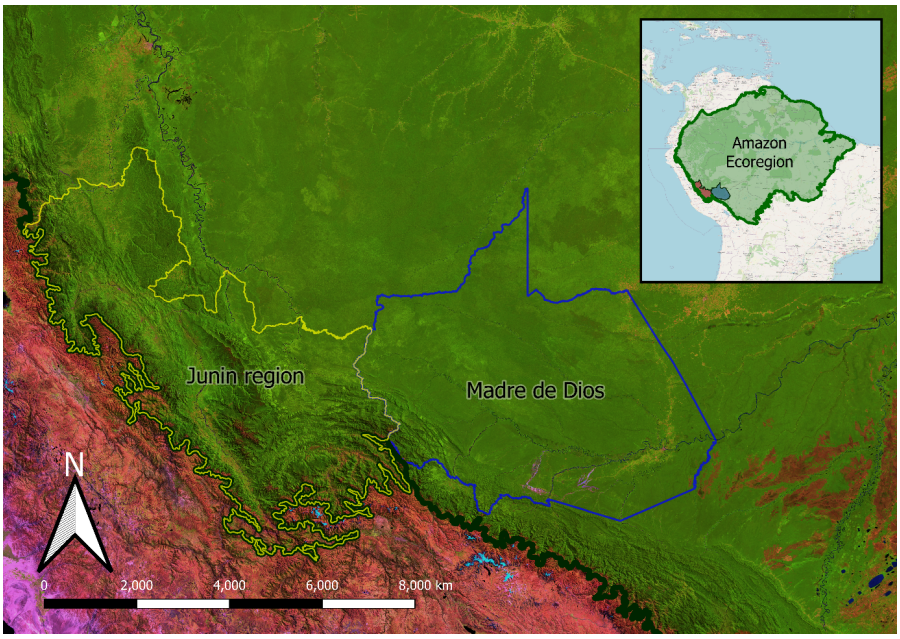


Fig. 2.1 Southern Peruvian study regions. Base image: Landsat composite (bands 5,4,3) 2019 (Hansen/UMD/Google/USGS/NASA).

2.2 Materials and methods

2.2.1 Study regions

This study focused on two regions in the south of Peru (Fig. 2.1). The size of the study regions were chosen to provide as many training points as possible while allowing processing to remain within the computing resources available to us. The data for larger regions could not be successfully loaded on the 4×Tesla P100-PCIE-16GB GPUs and 12×5980 MB CPUs that were available to us.

A preliminary broad intercomparison of a broad set of models focused on the Madre de Dios (MdD) department of Peru. MdD has approximately 97.4% primary forest cover and primary forest loss has steadily increased since 2000 (Hansen et al., 2013; Turubanova et al., 2018). It is almost entirely intact Southwest Amazon moist forest (Olson et al., 2001; Potapov et al., 2017). The Interoceanic Highway, a 2,575km road from the coast of Peru to Brazil, was completed in 2011, enabling access to portions of MdD previously protected from human activity. This increased access supported immigration, which has increased pressure on the forest. Legal and illegal gold mining is prevalent, seeding remote deforestation frontiers in otherwise intact forest (Nicolau et al., 2019). The region exports a range of agricultural products including cotton, coffee, sugarcane, cocoa beans, Brazil nuts and palm oil. Large logging concessions and illegal selective logging (particularly for mahogany) place additional pressure on the forest

there (Chirinos and Ruiz, 2003). The dynamic and diverse pressures made it a compelling trial region.

The most promising models were further developed and additionally applied to a second region. It centred on the Junín department, included surrounding departments and was cropped by the limit of the Amazon Ecoregion. It has approximately 88% forest cover, primarily split between Southwest Amazon moist forest and Peruvian Yungas (subtropical cloud forest) with some Ucayali moist forest (lowland to premontane moist forest) (Olson et al., 2001). While most of its forest is considered primary, there are some highly fragmented regions (Potapov et al., 2017; Turubanova et al., 2018). Drivers of deforestation are similar to those in MdD but with less mining and more hydrocarbon exploration (Finer et al., 2015, 2014). This region provided an opportunity to trial the networks on a similarly sized area but one with greater environmental heterogeneity.

2.2.2 Datasets

We limited ourselves to global, freely available datasets so the approach could be replicated and scaled to any forest location. The forest state labels used in this study were taken from the Global Forest Change dataset (Hansen et al., 2013). It tracks the location of forest loss globally and is collated on an annual basis (latest version: 2001-2020) at ~30m resolution. Loss events are based on a pixel's forest cover (here >30% to 0%) and are neutral to the cause of transition. Management activities, unauthorised resource extraction and loss due to other biotic or abiotic factors are not differentiated, thus observed patterns of loss relate to all possible causes.

The dataset includes a data mask describing areas of land or permanent water which we used to filter valid pixels. It also includes percentage tree cover observed in 2000 which we included as a predictor layer (for information on the forest density) and to filter out pixels with low cover (<30%). From the same source, cloud-free composite Landsat imagery of four spectral bands was available for 2014 to 2020. Near infrared (NIR) and short-wave infrared (SWIR) bands are informative for the remote sensing of vegetation. These were included to allow the computer vision networks to learn features from the spectral signals. We included the Japan Aerospace Exploration Agency's (JAXA) 30 m resolution ALOS Global DSM as a layer to give another dimension for the networks to learn from and exploit the features of the topology. It is an L-band SAR derived product which has better canopy penetration than C-band products and so gives a clearer signal of topology. Historic forest change labels were processed into four one-hot encoded layers which assigned pixels to classes based on how recently (from current time, t) they lost forest cover: (i) 0-1 years from t, (ii) 2-4, years from t, (iii) 5-8 years from t or (iv) more than 8 years from t. A pixel that does not register in these layers was free from loss since 2000. These layers encode the proximity, in time and space, of

Table 2.1 Predictor layers. The first three layers are static and the rest change year-on-year (t). All are 30 m resolution. Abbreviations: top-of-atmosphere (TOA), near infrared (NIR) and shortwave infrared (SWIR).

Variable	Description	Value
datamask	Land surface (0) or permanent water body (1)	0, 1
treecover2000	Percentage tree cover observed in 2000	0 - 100
elevation*	Height above sea level in meters	0 - max(h)
recent_loss1(t)	If forest to non-forest in years [t, t - 2)	0, 1
recent_loss2(t)	If forest to non-forest in years [t - 2, t - 5)	0, 1
recent_loss3(t)	If forest to non-forests [t - 5, t - 8)	0, 1
recent_loss4(t)	If forest to non-forests [t -8, 2000)	0, 1
last_b30(t)†	Normalised TOA reflectance Landsat 7 band 3 (red)	0-255
last_b40(t)†	Normalised TAO reflectance Landsat 7 band 4 (NIR)	0-255
last_b50(t)†	Normalised TAO reflectance Landsat 7 band 5 (SWIR)	0-255
last_b70(t)†	Normalised TAO reflectance Landsat 7 band 7 (SWIR)	0-255

*Tested as an additional data layer but not included in final models as it did not improve performance
†Latest available cloud-free observation at pixel

recent forest loss and thereby assist the models in representing contagion effects. All layers (see Table 2.1) were normalised to take values between 0 and 1 before being fed into the models.

Although additional layers (e.g. distance to roads, settlements) could have been included and may have improved overall accuracy, we wanted to test the ability of the deep CNNs to automatically extract useful features from a set of regularly updated, scalable RS data layers. This tests their utility in poorly mapped or highly dynamic regions. Additional details of the datasets are given in Appendix A.

2.2.3 Modelling process

The aim was to train a classifier that, when presented with a scene centred on a single forested pixel, could predict whether the target central pixel would remain forested (0) or transition to a non-forest state (1) in the year ahead. The model training and testing used 2014 to 2020 data, and forecasts were produced for 2021 (a year for which data were not yet unavailable). This section overviews the model training and testing process. Additional details are given in Section A.4 and corresponding scripts are available at https://github.com/PatBall1/DeepForestcast.

At a time t (years since 2000), the set of valid pixels (J_t) were those that had forest cover. When a pixel transitioned from a forest state to a non-forest state, this change was taken to be permanent. Specifically, a datapoint with a specific spatial location (central pixel), j, and

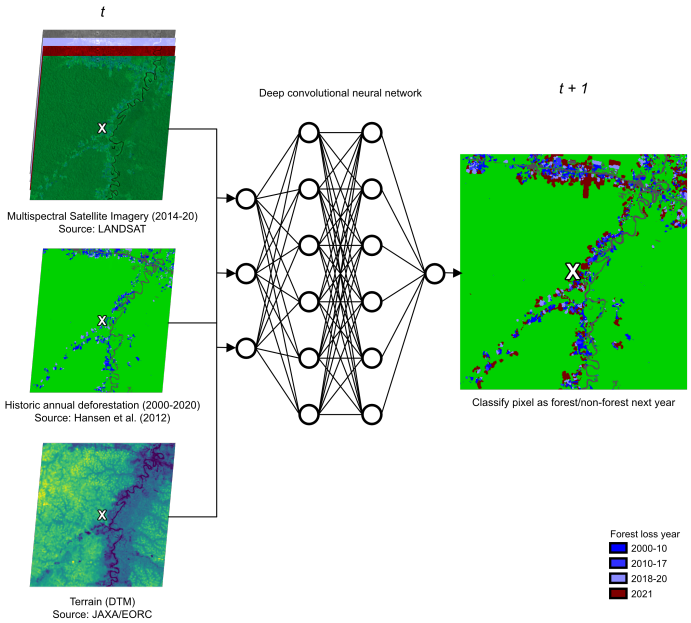


Fig. 2.2 Schematic of the deep learning prediction process. For a given forested pixel at the Centre of a scene, the convolutional neural networks learned to predict the state of the pixel (forest/non-forest) in the year ahead based on multispectral satellite imagery, historic loss patterns and the terrain.

point in time, t, consisted of data from a square $(2r+1)\times(2r+1)$ pixel scene, where r is the number of pixels the scene has in vertical and horizonal directions from the target central pixel, of the layers given in Table 2.1 (stacked bands). This allowed the network to observe what was happening in the neighbourhood of the pixel being considered (i.e. make decisions based on the field of view), which contrasts with typical ML approaches that use information on the pixel in isolation or with averaged neighbourhood metrics that lose the spatial structure information of the data. Each datapoint was associated with a non-forest (0) or forest (1) label from the year ahead ($t+1$; see Fig. 2.2).

For each datapoint, two types of 3D tensors were inputted into the models. The first, S^j, contained the stacked static layers (see Table 1) and was of the shape $S^j \in \mathbb{R}^{s\times(2r+1)\times(2r+1)}$, where s is the number of static layers (here 2 or 3 depending on whether the DSM layer was included). The second type, X_t^j, contained the stacked dynamic layers (see Table 2.1) and was of the shape $X_t^j \in \mathbb{R}^{d\times(2r+1)\times(2r+1)}$ where d is the number of dynamic layers (here 8). The label associated with the data point was defined as $Y^j_{(t+1)} \in \{0,1\}$, and took a value of 1 only if the central target pixel was labelled as non-forest in year $t+1$. By allowing the models to learn the features (and interactions) of the data layers in tensors that are associated with the year-ahead forest/non-forest labels, the models would be able to make predictions of the year-ahead label ($\hat{Y}^j_{t+1}$) when presented with new input tensors. Output $\hat{Y}^j_{(t+1)}$ could vary continuously between

0 (no chance of forest loss) and 1 (certain transition) with a threshold applied to determine the binary classification (what we refer to here as deforestation 'risk').

The input data for each year, with associated year-ahead (t + 1) labels, was partitioned with stratified random sampling, into training (80%) and test (20%) sets. The training set was further split with stratified random sampling into five even splits so that 5-fold cross validation could be performed while tracking model accuracy during the training process. Stratification was based on the proportion of class labels. Less than 0.5 % of each study area transitioned from forest to non-forest in any given year which meant the number of pixels that remained as forest the next year (0) vastly outweighed the number that transitioned to non-forest (1). This extreme class imbalance, if not addressed, would likely have led to very low predictive accuracy for the infrequent class. As suggested by Buda et al. (2018), we modified our training data by under-sampling the over-represented class so that there was at least one positive case (1) to four negative cases (0). Fortunately, there remained several million valid data points available in each study region each year.

A preliminary intercomparison of all model classes was first carried out on the MdD study region using 2014-2018 training and testing data, withholding the latest 2019 labels for the final testing. Using a grid search of hyperparameters, a set of models (differing in architectures and hyperparameters) was trained (with 5-fold cross validation) on 2014-2017 input data based on labels up to 2018. The accuracy of these models was tested on a (within training period) withheld test set (2014-2017 input data, 2018 labels). Model selection was based on AUC (see below). This process was used to evaluate the most effective network depths for each model type.

Training deep convolutional neural networks is resource intensive so, from the initial intercomparison, the most promising model architectures, in terms of model accuracy and training efficiency, were selected to be taken forward to be further refined and tested on the Junín region and well as MdD. Using Bayesian hyperparameter tuning (Snoek et al., 2012; implemented on wandb.ai) and retraining on the 2014-2018 inputs (2019 labels), optimised models were produced. The trials and performances achieved during the hyperparameter sweep is available to inspect at https://wandb.ai/patball/forecasting/sweeps/df5v36lz.

The models were tested on the withheld 2014–2018 inputs (2019 labels) test set before being tested on the year beyond the training period test set (2015–19 inputs; 2020 labels). It was important to test the models' abilities to predict the labels for a year outside of its training period to understand whether the models are transferable through time and therefore able to produce reliable forecasts. The optimised models (with hyperparameters and weights retained) were then updated by continued training on 2015–2019 input data and 2020 labels. The final models were then used to produce deforestation forecasts across the entirety of each study area

for 2021 (available for inspection as a GeoTiff at https://doi.org/10.5061/dryad.hdr7sqvjz). Fig. A.13 summarises the arrangement of modelling processes.

Fresh models, with the same architecture and parameterization as the optimised models, were trained only on the MdD data. To test spatial transferability, the accuracy of this model was evaluated first on MdD and then the Junín region. This model was then used to forecast 2021 deforestation for MdD (available for inspection as a GeoTiff at https://doi.org/10.5061/dryad.hdr7sqvjz). Additional models, with the same parameterization as the optimised models but sequentially trained on data on MdD data then Junín region data were assessed to help understand the effect of expanding the geographic diversity of training cases.

The metrics we used to evaluate the models were area under receiver operating characteristic curve (AUC), precision, recall and F_1 score. AUC integrates model performance over all classification thresholds. As the threshold could be adjusted to suit the user's relative preference towards a higher true-positive rate (recall or sensitivity) or false-positive rate, AUC was insightful to assess overall model performance. F_1 score (harmonic mean of the precision and recall) can communicate accuracy while accounting for variable class imbalance, so it was selected as a metric for model intercomparison over simple prediction accuracy (which can be misleading in cases of class imbalance).

We used the Adam optimizer with weighted cross entropy loss to train the models through stochastic gradient descent (Kingma and Ba, 2014). Weighted cross-entropy loss was used to help address the class imbalance between loss events and pixels remaining forested. The penalty for missing a loss event was set to be double that of incorrectly identifying a loss event at a pixel that remained forested as we felt it preferable for the network to be tuned to avoid missing the relatively rare loss events. To avoid over-fitting, regularisation was implemented through a dropout layer (Srivastava et al., 2014) and early stopping criteria. Dropout randomly sets a fraction of the neurons in the network to zero during forward pass of training, i.e., during each training iteration, a certain number of nodes are"dropped-out" or ignored. This means that their contribution to the activation of downstream neurons is temporarily removed and no weight updates are applied to them during back propagation. Since neurons cannot rely on the presence of particular other neurons, they tend to learn more robust features that are useful in conjunction with multiple different random subsets of other neurons. This can lead to each neuron learning to detect features that are generally useful, rather than features that are specific to the idiosyncrasies of the training data. The models were trained with repeated exposure to the full sampled training set (i.e. multiple epochs).

We used the PyTorch ML framework to develop the models (code available at https://github.com/PatBall1/DeepForestcast). We used GPUs (4 $\times$ Tesla P100-PCIE-16GB) and CPUs (12 $\times$ 5980 MB) with a 12-hour time limit on a high-performance computing cluster

to train and test the models and make forecasts. Details on the computational resources used, including hardware and software, are given in S5. The trials conducted as part of the training process, including hyperparameters used, accuracies attainted, and computing resources used are available at https://wandb.ai/patball/forecasting/.

2.2.4 Model architectures

A variety of MAs were designed and implemented, and described in brief below. Technical details and developed rationales behind the choices of the network architectures can be found in Appendix A. A deep neural network has a set of weights which define the strength of connections between the nodes and the predictions it makes from input data; these are adjusted as the network learns from exposure to training data. A single MA has a corresponding set of variable hyperparameters. Hyperparameters define aspects of how a model handles data, builds interactions, and how it adjusts its weights to learn from the data. A unique set of hyperparameters as well as the architecture defines what we refer to as a model. By varying the hyperparameters and comparing the resultant model accuracies, the most effective set of hyperparameters can be identified (hyperparameter tuning; see Section 2.2.3). A spatial pyramid pooling layer in each network allowed the spatial input window size to be varied and optimised as a hyperparameter (He et al., 2014). We used Batch Normalization which is a layer that normalises each filter to have a zero mean and unit variance (Ioffe and Szegedy, 2015). Ioffe and Szegedy (2015) showed that employing such layers can be beneficial in several ways: the networks train faster as it enables the gradient descent algorithm to take higher learning rates; the convergence of the loss function is significantly less sensitive to how the weights are initialized; it offers some level of regularization by adding a small amount of noise to the data. The output of the network is a single value between 0 and 1 on which a threshold is applied to classify a pixel as becoming deforested or remaining forested (e.g. if a 0.5 threshold is selected, any value below this will be predicted as still forested and any value above will be considered deforested). This meant all pixels could be ranked based on their likelihood of deforestation. The threshold value was set to give a balance between precision and recall but, were there other requirements, it could be adjusted to account for anticipated changes in deforestation rate or to prioritise different amounts of area for protection. More details are given in Appendix A.

Model architecture 1: 2D convolutional neural network

The first and simplest MA we employed (MA1) was a type of 2D convolutional neural network (2D CNN). In a 2D CNN, filters slide (‘convolve’) in two spatial dimensions across a 3D input tensor (with a spatial extent and a depth that corresponds to the number of bands) to extract

spatial features that can be used by the network to make predictions. Inputs were simply the dynamic tensor static tensor stacked at a point in space. Potential temporal features of the data were not implicitly characterised by this type of network. Temporal change was simply captured by pairing the input data for 1 year (at time t) with the forest/non-forest labels from a year ahead (at $t+1$). See Fig. A.10 for a visual representation of the MA.

Model architecture 2: 3D convolutional neural network

The second MA (MA2) was a type of 3D convolutional neural network (3D CNN). Li et al. (2017) demonstrated that 3D CNNs applied to hyperspectral data could achieve state-of-the-art land use classification accuracy. We thought that, by substituting the spectral dimension with the temporal dimension, it would be possible to draw out the spatiotemporal features of forest change. This type of network slides ('convolves') filters in three dimensions (two spatial and one temporal). This meant, for a given central target pixel, the series of dynamic tensors (see Section 2.2.3) could be stacked temporally along the channel axis to form a 4D input tensor and 3D convolutions used to characterise the spatio-temporal features of the input data. The static tensor was passed to a separate 2D convolutional branch. See Fig. A.11 for a visual representation of this MA.

Model architectures 3 and 4: Convolutional long short-term memory recurrent neural network

MAs 3 and 4 were also designed to implicitly handle the temporal characteristics of the input data. Instead of using 3D convolutions, they use a Convolutional Long Short Term Memory (ConvLSTM)-based Recurrent Neural Network (RNN) type architecture. Rußwurm and Körner (2018a) demonstrated the state-of-the-art performance of this network type for crop cover classification from multispectral satellite data. At a given time step, RNNs can remember and make use of information from previous time steps; they are designed to characterise and predict sequential data. LSTM is a specialised RNN design that allows for long time dependencies to be learnt. By integrating spatial convolutional elements within the LSTM component, a ConvLSTM can characterise and predict the sequential evolution of spatial patterns. MA3 had a single ConvLSTM cell, whereas MA4 was 'deeper' as it had a stack of ConvLSTM cells (see Section A.3.5 for details). We compared these two architectures as stacking ConvLSTM cells can improve prediction accuracy in some situations (e.g. Kim et al., 2017). These MAs take the same 4D input tensors as the 3D CNN and pass the static tensor to a 2D convolutional branch. These MAs, while able to achieve comparable accuracies to the first two, were dropped after a broad intercomparison as their training could not be parallelised and so it took far longer

to achieve these accuracies. However, their model classes are available for experimentation from https://github.com/PatBall1/DeepForestcast as their implementation may become more effective as the amount of historic data and computing resources increases.

Baseline: Random forest

To assess the relative benefit of using deep CNN-type methods, we trained, tuned and tested random forest classifiers to perform the same task. The initial data structure was equivalent to that of the 2D CNN, that is, a spatial window around a focal pixel was still used as an input and forest state labels were offset by 1 year. However, the 3D tensors were necessarily flattened into a 1D feature vector so that they could be given to the Random Forest. This meant that the spatial structure was not preserved (as in the case of the CNN type models) but the model was still free to learn which pixels (closer or farther from the central pixel) of which layers were of greatest importance for making predictions. Fivefold cross-validation with a randomised search over the hyperparameters (number of trees, tree depth, number of features) and window size was performed to tune the model.

2.3 Results

2.3.1 Model Performances

Broad intercomparison

The best models from each model class (i.e. selected after hyperparameter tuning) predicted 2018 MdD test set deforestation with similar accuracies (AUC = 0.935–0.944). Precision and recall were approximately equal across MA classes. The best performing MA after hyperparameter tuning was the 3D CNN (MA3; AUC = 0.944). This was followed by the deep ConvLSTM (MA4; AUC = 0.938), the ConvLSTM (MA3; AUC = 0.937) and the 2D CNN architecture (MA1; AUC = 0.935). This suggested that models that could implicitly handle and characterise the temporal with the spatial structure of the input data may be better suited to predicting future deforestation. In other words, those models that could retain and work with the spatiotemporal patterns of the data (how the spatial patterns evolved), rather than just use spatial patterns, tended to perform better. MA4 and MA3, while achieving comparable accuracies, were uncompetitive with respect to training time due to their constraints on parallisation (a well-known problem with RNNs). Therefore, MA1 and MA2 were selected to be taken forward for further development and testing.

The spatial dimensions of the input tensors determined the size of the scene that models were able to view and had an influence on model performance. Model AUC plateaued around an input frame of ∼35 × 35 pixels (∼1 km × 1 km) and began to decrease for input frames larger than 50 × 50 pixels (∼ 1.5 km × 1.5 km). Between these sizes seemed to be the sweet spot where the networks could learn from local context while also remaining focused on the pixel of interest. Full details of the model performances, including hyperparameters and receiver operating characteristic (ROC) curves, are given in S7.

Model development

Further tuning and extended training improved within-year accuracy of the 2D CNN and 3D CNN model classes which performed better than the baseline models. The precision of the baseline model was higher than that of the CNN models but, based on a broad interpretation of the accuracy metrics, this was more a feature of how the decision threshold was set rather than an indication of superior forecasting ability. Across the board, the models predicted less accurately on the Junín region than MdD. It suggests that the Junín region is inherently more difficult to predict on (it has more cloud cover, diversity of forest types and fragmentation). Models that included the topological layer as an input predictor layer were no more accurate than those without, suggesting it was uninformative to the model. To reduce redundancy, this layer was dropped.

There was a drop in model prediction accuracy from the within training period test set and the year ahead test set (see Table 2.2). This was to be expected but it could also indicate some change in the patterns of deforestation or a change in the nature of the input data (it is difficult to produce satellite data that is exactly consistent across dates due to e.g. atmospheric conditions). The greater the complexity of the model, the greater the drop off in accuracy from nowcast to forecast predictions indicating a degree of overfitting. The model that was trained on both regions sequentially was only marginally better at predicting year ahead deforestation than the model that was trained on just the region to be predicted (for 3D CNN). The highest year ahead AUC from the 3D CNN for the Junín region was returned by the model trained exclusively on MdD data. This suggests that data quality is at least as important as volume and spatial range of training data. It also suggests that the spatio-temporal patterns of forest loss between the two regions are somewhat similar and that the models are spatially transferable.

Table 2.2 Within year and year ahead pixel wise classification accuracy of the 3D CNN, 2D CNN and random forest baseline model classes. The reported accuracies are all for unseen test sets. The 2019 labels are within the training period and 2020 are beyond (to test forecast accuracy). The ratio of 0 to 1 labels in the test sets were 4:1. Bold font is used to highlight the best performing models based on F_1 Score and italics show the baseline model performance. The baseline model had a better Precision than the CNN models but this is a result of how the thresholds are set between the models rather than an indication of superior performance.

				Test set 2019				Test set 2020			
Model class	Training period (training labels)	Training region	Prediction region	AUC	Precision	Recall	F1	AUC	Precision	Recall	F1
3D CNN	2014-2018 (2018-2019)	Junín	Junín	0.990	0.903	0.903	0.903	0.839	0.580	0.580	0.580
3D CNN	2014-2018 (2018-2019)	Madre de Dios	Junín	0.879	0.663	0.663	0.663	0.861	0.593	0.593	0.593
3D CNN	2014-2018 (2018-2019)	Madre de Dios; Junín	Junín	0.997	0.904	0.904	0.904	0.849	0.589	0.589	0.589
3D CNN	2014-2018 (2018-2019)	Junín	Madre de Dios	0.961	0.902	0.903	0.903	0.863	0.667	0.667	0.667
3D CNN	2014-2018 (2018-2019)	Madre de Dios	**Madre de Dios**	0.997	0.964	0.964	0.964	**0.890**	**0.715**	**0.715**	**0.715**
2D CNN	2014-2018 (2015-2019)	Junín	**Junín**	0.984	0.865	0.865	0.865	**0.876**	**0.623**	**0.623**	**0.623**
2D CNN	2014-2018 (2015-2019)	Madre de Dios	Madre de Dios	0.882	0.673	0.673	0.673	0.868	0.654	0.6544	0.654
Random forest (7x7)	*2014-2018 (2015-2019)*	*Junín*	*Junín*	*0.806*	*0.688*	*0.434*	*0.532*	*0.803*	*0.679*	*0.430*	*0.527*
Random forest (7x7)	*2014-2018 (2015-2019)*	*Madre de Dios*	*Madre de Dios*	*0.816*	*0.750*	*0.492*	*0.594*	*0.814*	*0.750*	*0.489*	*0.592*

While the 3D CNN forecasted most accurately on the MdD region, the 2D CNN gave the best forecast accuracy for the Junín region. The 3D CNN has the advantage of being able to convolve over spatial and temporal dimensions together (and so learn spatiotemporal features) but the 2D CNN has more datapoints available to learn from as the inputs are for a single year input data (with an offset label) and not stacked across time like with the other model types. More details of the model performances over different hyperparameters is available at https://wandb.ai/patball/forecasting2d/ and https://wandb.ai/patball/forecasting/ for the 2D CNN and 3D CNN, respectively.

2.3.2 Forecasts

The near-future forecasts exhibited some interesting spatial features. To illustrate the 3D CNN's assessment of deforestation risk, we have focused on four recent forest loss events in the two study regions. The 2021 risk forecast was overlaid over Planet/NICFI biannual composite satellite imagery for 2020 (see Fig. 2.3). Fig. 2.3a shows a rapidly growing commercial agricultural area well serviced by a river and roads in Junín. The network signals a very high risk that agricultural clearance will continue to expand into the remaining forest areas. Fig. 2.3b shows a new, remote, unauthorised gold mine in MdD. The network anticipates an immediate expansion of the mining operations into surrounding areas. Fig. 2.3c shows a newly laid access route connecting newly established agricultural settlements to a large river. The network highlights a risk of clustered conversion along parts of its route. Fig. 2.3d shows a large recent forest loss event caused by a landslide. Despite the scale of the event, the network deems its surroundings as relatively low risk from further change of state compared with the anthropogenic causes of loss. A remote landslide is unlikely to result in a progressive loss of forest. The contrasting predictions (in combination with the high reported accuracy) suggest a level of discrimination (or intelligence) in the model's evaluation of transition likelihood. Forecasts are available for inspection in GeoTiff format at https://doi.org/10.5061/dryad.hdr7sqvjz.

2.4 Discussion

2.4.1 Improving spatial forecasts

For the first time, deep CNNs were used to forecast the spatial development of forest loss. From a set of freely available, global predictor layers, we predicted year-ahead forest loss at 30 m resolution F_1 score = 0.62–0.71 relative to the baseline models that achieved F_1 = 0.52–0.60

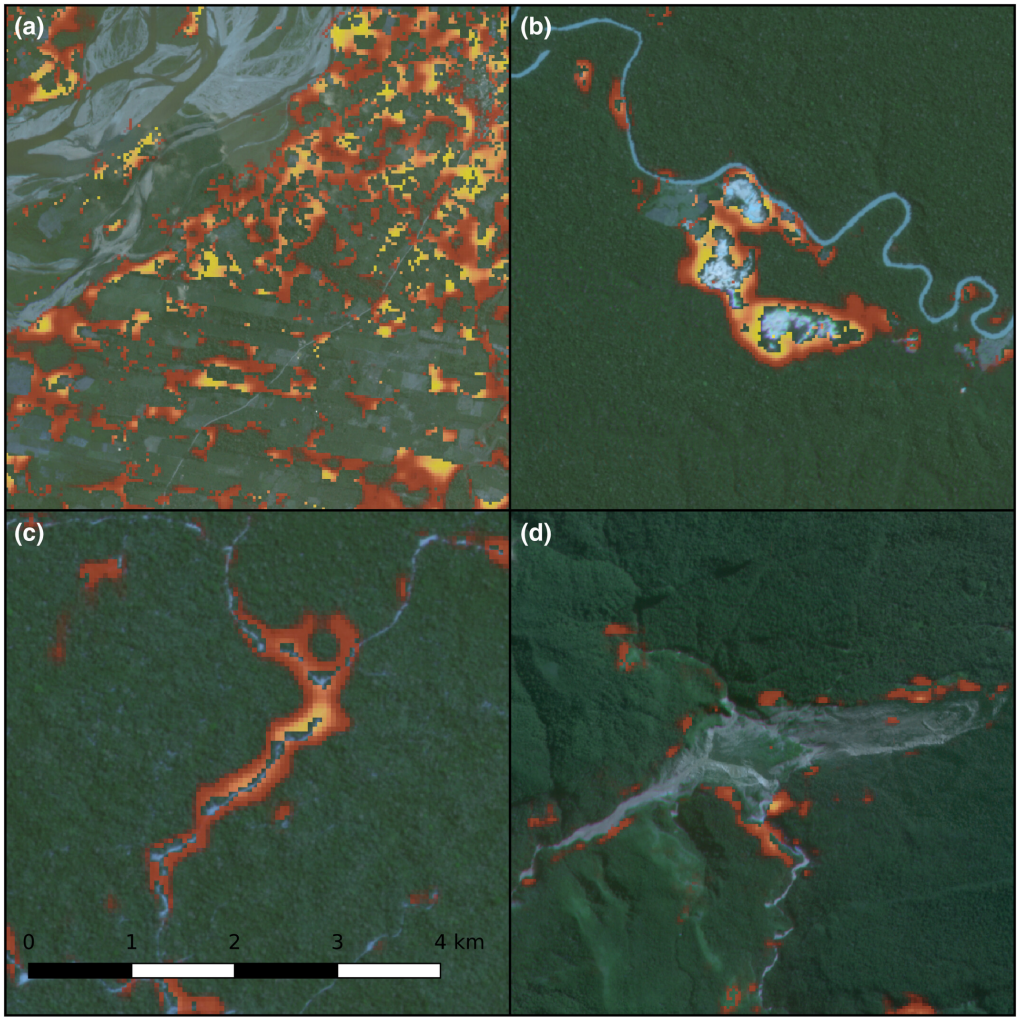


Fig. 2.3 Examples from the 2021 deforestation risk forecasts overlaid on satellite imagery. Yellow indicates the areas most likely to transition to non-forest states fading to red as areas are less likely to transition: (a) expanding commercial agriculture; (b) a newly established, unauthorised gold mine; (c) a new access route connecting undeveloped forest to a river and (d) a recent remote landslide. Source of satellite imagery: Planet/NICFI.

for two ∼80,000 km^2 regions in the Southern Peruvian Amazon while also demonstrating a degree of spatial and temporal transferability of the models.

The apparent ability of the networks to automatically extract landscape features relating to different drivers of forest loss and infer whether loss was likely to continue from a source was encouraging. It highlights the potential for deep learning computer vision methods to process large volumes of RS data to manage risks and make predictions in complex human-natural systems. Previous attempts at forecasting deforestation have often had limited predictive capacity due to a lack of reliable input data on dynamic landscape features. From satellite data and other complementary spatial datasets, our model appeared able to discriminate between one-off loss events (e.g. landslide) from frontiers that are likely to progress over time (e.g. from a newly built road, an unauthorised gold mine). A more rigorous exploration of the decisions of the network would help to determine the degree to which this is true (see below).

The dimensionality of convolutions in the networks - spatial in the 2D CNN case and spatiotemporal in the 3D CNN case - involved a trade-off. The ability of the 3D CNN to exploit spatiotemporal features allowed it to perform most accurately in the case of MdD, the region with less topological variation and ecosystem variety. Indeed, the ability of 3D CNNs to view a pixels spatial and temporal context together has been shown to improve classification accuracy from RS data (Ji et al., 2018; Xu et al., 2018). This implicit handling of the temporal dimension with the spatial dimensions would seem especially important in the case of forecasting. However, stacking the inputs across the time axis reduced the number of individual datapoints available to the network to learn from. The 2D CNN, with a greater volume and temporal variety of data available for it to learn from, performed better on the Junín region. The Junín region has persistent cloud cover and so stacking the data over time may propagate noise and degrade training data. The individual year inputs of the 2D CNN are more resistant to this as potential noise or gaps in the data would not contaminate informative datapoints. The greater volume of data may have helped the network deal with the environmental variability.

Deep learning models have many hyperparameters that need tuning. We demonstrated a degree of spatial transferability of our models, but users should be aware that additional tuning and retraining may be necessary to achieve optimal performance in a new region. The optimal window width to view the scene was 1–1.5 km. For larger windows, the marginal benefit from the additional information provided less benefit than the cost of handling this information. In other words, there was a point at which adding further away pixels to a scene began to distract the network from the area of focus.

We stress that, while the reported accuracy is encouraging, there are reasons to receive these metrics cautiously. The GFC (Hansen et al., 2013) dataset provided all data labels and so

the reported accuracies specifically relate to how well the networks predict future outputs of this classifier and not necessarily the ground truth. The GFC dataset was generated by a bagged decision tree that classifies filtered Landsat data. Our models were likely to have approximated and incorporated behaviours of this classifier and accuracy may be inflated (relative to on-the-ground change) from learned correlations. Biases leading to false positives and false negatives in the initial model have likely been transferred to the forecasting models (Kinnebrew et al., 2022).

Any ML approach is necessarily limited by the fact it is based on historic data and the assumption that observed trends will continue, and therefore does not implicitly account for regime changes in the system. This is relevant here, as political changes have substantial impacts on Amazonian deforestation (Pereira et al., 2020). A changing climate, volatile international economic conditions and advances in technology further challenge the assumption of stationarity. However, as the forecasting system can be dynamically updated it is able to learn emerging patterns which is an improvement on previous approaches. Additionally, the models output a continuous measure of risk across all pixels and, by varying the threshold at which the binary classification is made, it is possible to adjust the overall allocation of forest loss in predictions. It would be possible to couple the spatial evaluation of risk from the CNN system (accounting for landscape scale changes) to a model of overall forest loss based on external drivers to generate forecasts that reflect global trends. The pixelwise risks can be integrated to assess relative risk across a region.

2.4.2 Data sources

We limited our set of data layers to test the networks' ability to automatically extract dynamic landscape features. However, including additional RS data layers as predictors could help to inform the networks and improve accuracy. Of particular interest is night lights, which provide a dynamic signal for human presence and economic activity. At current, the resolutions of available night light products are too coarse to integrate into our system, but this is likely to change in the near future (Levin et al., 2020). The ∼30 m resolution of Landsat derived products is likely to fail to pick up on selective logging. Higher resolution imagery (e.g. Planet NICFI, Sentinel-2) may make that feasible but there is not currently a matching dataset for forest cover labels. The GFC dataset only provides annual information on forest loss and not forest gain which means the forecasts presented here cannot account for reforestation. The Copernicus Global Land Cover collection (Buchhorn et al., 2020) could provide labels for more types of landcover transitions but is only available for 2015–2019 and at 100 m resolution. As this data source develops more intricate predictions that support forest regeneration analyses may be possible. Furthermore, rather than simply predicting transitions between forest and

non-forest states it may also be valuable to make predictions on gradual declines in relation to disturbance events and forest resilience. In this context, it would be possible to train networks similar to those presented here to predict the evolution of greenness metrics such as NDVI (Boulton et al., 2022). The ALOS JAXA elevation layer was found to not improve model performance, a somewhat surprising find in Junín, which has inaccessible mountainous areas. It may be that processing the elevation into a relative slope layer would allow the networks to better comprehend the features of the local topology.

2.4.3 Probing the black box

Deep neural networks are 'black box' models (i.e. the inner logic of the model is not readily interpretable) so it is difficult to say how the networks were making decisions and to what degree they were recognising different potential drivers of loss. There has been work on interpreting deep CNNs to explore their internal reasoning (Li et al., 2021b). By applying interpretation algorithms, it may be possible to identify which parts of a scene are triggering the networks' decision. It may even be possible to generate groups of scenes (e.g. by driver) that typically signal forest loss (see e.g. the Activation Atlas of Carter et al., 2019). Automatic classification of drivers has been demonstrated at coarse resolutions (10 km) with simple decision tree models applied to derived RS products (Curtis et al., 2018). Given the relative sophistication of the CNN technology available, it is conceivable that a network could be trained to classify and map drivers directly from satellite imagery at high resolution, particularly as the amount of high-resolution imagery is rapidly increasing (e.g. Planet NICFI, Sentinel-2). However, a comprehensive training set of labelled deforestation drivers at this resolution does not exist and would require a considerable investment to produce.

Fieldwork to potential transition regions may be another way to better understand the networks' predictions. Of particular interest would be to investigate the cases where the models are failing to make accurate predictions. Identifying specific drivers that lead to misclassification would allow for more focused model training. Active learning algorithms (Settles, 2009) could guide expeditions by identifying regions that would be particularly informative to gather extra data on.

2.4.4 Technological advances

The MAs were inspired by models that have performed well on land cover classification and adapted to fit the specific forecasting task. However, recent architectural innovations could further help to improve performance. Transformer networks (originally from natural language processing) have proven themselves to be the state-of-the-art in handling sequential

data (Vaswani et al., 2017) and recently in some computer vision tasks (Dai et al., 2021). Using self-attention, they avoid the difficulties in parallelising RNNs, and so can be trained more efficiently. Integrating transformer components into the networks may prove to be an effective way of handling the spatiotemporal features of the data (especially as the available time series grows). This could help to extend the networks forecasting horizon. Further improvements could also be gained by refining the training process to focus it on contentious datapoints. The under sampling of the non-change class was random and many training points would be uninformative as they are in very remote locations. A method that directs the training to cases that sit close the decision threshold between the classification states would address this. We have provided open-source code and benchmark results, and hope further improvements can be integrated through collaborations.

2.4.5 Applications

The forecasting approach presented here was designed to be flexible, useful to a range of potential users and scalable to forests globally. All pixels in a landscape can be ranked according to the likelihood of loss allowing for spatial prioritisation. An organisation may have resources to police a certain proportion of a region and the forecasts would allow them to monitor the most threatened areas. The forecasts could also direct interventions to emerging frontiers of loss in otherwise isolated and intact landscapes. On the other hand, when planning a permanent protected area, it may be preferable to avoid areas that are most imminently threatened but address areas that may undergo transition if left unprotected. In either case, it is beneficial to understand how threats are likely to evolve. Governmental and non-governmental organisations can use the tools presented here to understand how the landscapes in which they operate will change into the future and refine their protocols for managing and responding to deforestation risk. The effectiveness of deep CNNs in the current context suggests that they should also be applied to other ecological forecasting tasks based on RS data such as forest fire forecasting (see e.g. Santopaolo et al., 2021).

Chapter 3

Accurate delineation of individual tree crowns in tropical forests from aerial RGB imagery using Mask R-CNN

Abstract

Tropical forests are a major component of the global carbon cycle and home to two-thirds of terrestrial species. Upper-canopy trees store the majority of forest carbon and can be vulnerable to drought events and storms. Monitoring their growth and mortality is essential to understanding forest resilience to climate change, but in the context of forest carbon storage, large trees are underrepresented in traditional field surveys, so estimates are poorly constrained. Aerial photographs provide spectral and textural information to discriminate between tree crowns in diverse, complex tropical canopies, potentially opening the door to landscape monitoring of large trees. Here we describe a new deep convolutional neural network method, *Detectree2*, which builds on the Mask R-CNN computer vision framework to recognise the irregular edges of individual tree crowns from airborne RGB imagery. We trained and evaluated this model with 3,797 manually delineated tree crowns at three sites in Malaysian Borneo and one site in French Guiana. As an example application, we combined the delineations with repeat lidar surveys (taken between 3 and 6 years apart) of the four sites to estimate the growth and mortality of upper-canopy trees. *Detectree2* delineated 65,000 upper-canopy trees across 14 km^2 of aerial images. The automatic method in delineating unseen test trees was attained an F_1 score of 0.64 and for the tallest category of trees (from the height of the tallest tree at a site to a height exactly divisible by 5 m such that the category contains at least 10 trees) an F_1 score of 0.74. As predicted from previous field studies, we found that growth rate declined with tree height and tall trees had higher mortality rates than intermediate-size trees. Our approach demonstrates that deep learning methods can automatically segment trees in widely accessible RGB imagery. This tool (provided as an open-source Python package) has many potential applications in forest ecology and conservation, from estimating carbon stocks to monitoring forest phenology and restoration.

3.1 Introduction

Intact tropical forests are an important component of the global carbon cycle: they are major carbon stores and significant carbon sinks (Pan et al., 2011). However, the strength of the carbon sink is diminishing as a result of global warming (Brienen et al., 2015; Hubau et al., 2020) and there are concerns that forests are reaching a tipping point beyond which they could switch irreversibly to open savanna systems (Chai et al., 2021). Forecasting the future of tropical forests is challenging, because little is known about the ways different species will respond to changing climate, or the resilience provided by that diversity (Fisher et al., 2018; Gallup et al., 2021; Koven et al., 2020; Restrepo-Coupe et al., 2021). To understand the likely responses of forests to further climate change, ecosystem models need to represent growth and mortality processes of individual trees more accurately than is currently the case (Kellner et al., 2019; Piponiot et al., 2022; Zuidema and van der Sleen, 2022).

Remote sensing of individual upper-canopy trees can improve estimates of forest carbon (Dalponte and Coomes, 2016) and provide a means of tracking growth and mortality over large spatial scales. Traditional monitoring approaches rely on measuring stem dimensions in permanent inventory plots, and periodically revisiting those plots to assess recruitment, growth and mortality (Chave et al., 2019). However, such plots cover a small fraction of the overall tropical forest extent (~0.0002% of tropical forests are sampled by the main plot networks) and their locations are often dictated by ease of access rather than by robust statistical sampling designs (Davies et al., 2021; ForestPlots.net et al., 2021; Marvin et al., 2014). Furthermore, upper-canopy trees store the majority of biomass carbon in tropical forests, but relatively few of them are sampled in inventory plots due to their relative rarity (Coomes et al., 2017; Lutz et al., 2018; Meakem et al., 2018). This under-sampling is particularly problematic when assessing impacts of climate change, because upper-canopy trees are most vulnerable to periods of water shortage (Gora and Esquivel-Muelbert, 2021; Stovall et al., 2019) which are increasing in frequency (IPCC, 2021). Remote sensing has the potential to overcome these sampling challenges by providing wall-to-wall maps that can be used to monitor millions of upper-canopy trees.

Remote sensing of individual trees has mostly focused on airborne lidar data, which is used by, among others, the forestry industry to map trees at landscape scales (Zhen et al., 2016). Delineating individual trees from airborne lidar datasets is most successful for conifers, because their apical dominance results in clear local height maxima that make tree crowns easily distinguishable (Dalponte and Coomes, 2016; Hastings et al., 2020), but complex tropical canopies have presented a far greater challenge for lidar delineation (Aubry-Kientz et al., 2019). Tropical forest canopies are often densely packed with partially interwoven crowns which point-cloud clustering algorithms can struggle to distinguish (Aubry-Kientz et al., 2021).

Furthermore, lidar surveys require expensive aircraft (airplanes, helicopters or high-end drones) and sensors (that together require tens of thousands to hundreds of thousands of pounds to purchase or rent) whereas standard RGB imagery can be collected cheaply with drones (roughly £250 to purchase entry-level equipment to several thousand pounds for high-end options).

Automatic delineation of trees in RGB photographs can harness colour and texture information to distinguish trees, even if they are structurally similar (Almeida et al., 2021; Iglhaut et al., 2019). Most current methods of individual tree identification from RGB imagery use bounding boxes (Fig. B.3) (Santos et al., 2019; Weinstein et al., 2019, 2021), but more exact delineation of the edges of tree crowns would provide information on crown area and lateral growth, and avoid mixing signals from neighbouring vegetation. Recent advances in neural network approaches to computer vision provide opportunities to recognise individual trees from standard digital photographs taken from drones. A class of machine learning algorithms called deep convolutional neural networks (CNNs) is revolutionising vegetation science through its ability to exploit spatial structures and automatically extract high-level features from image data (e.g. analyses of satellite imagery) (Kattenborn et al., 2021; Mugabowindekwe et al., 2022; Zhu et al., 2017). In the field of computer vision, exactly segmenting individual objects of interest from an image is known as *instance segmentation*. The Mask R-CNN algorithm has shown promise in tree crown identification and delineation in plantations (Hao et al., 2021; Yu et al., 2022b), pine forests (Hu et al., 2022; Ocer et al., 2020), urban woodlands (Ocer et al., 2020) and forest fragments (Braga et al., 2020). Mask R-CNN has features that could allow it to overcome the challenges of delineating crowns in complex tropical canopies by discriminating based on the spectral and textural signals which are rich due to the phylogenetic diversity.

Here, we describe *detectree2*, a system that automatically detects tree crowns from aerial RGB imagery. We adapted Facebook AI's Mask R-CNN algorithm, (the *Detectron2* release), which has models that have been pre-trained on a wealth of available image data that can be transferred to new tasks (He et al., 2017; Wu et al., 2019b). We trained and evaluated *detectree2* on four tropical forest sites. In total, 3797 manually delineated tree crowns were used of which 1530 spatially separated crowns were reserved to evaluate the model. The size of this dataset meant we had to consider ways of boosting the efficiency of model training such as data augmentation. We evaluated the performance with F_1 Scores, which quantify the skill of the method in delineating individual tree crowns accurately. We expected a model trained at one site to drop in performance when transferred to making predictions of crowns at the other sites and that supplying a greater variety of training data would boost performance. As an example ecological application, we deployed the trained model across 14 km^2 of airborne RGB imagery, automatically delineating 65,786 tree crowns. For context, this area is approximately 40% the total area of forest inventory plots in the main plot networks across the tropics (Davies

et al., 2021; ForestPlots.net et al., 2021). We then combined these crowns with repeat airborne lidar data to investigate the growth and mortality rates of upper-canopy trees in relation to their height. Regional and global syntheses of forest inventory data suggest that growth slows down and mortality rates increase with tree size (Coomes et al., 2003; Hurst et al., 2011; Iida et al., 2014; Muller-Landau et al., 2006; Richardson et al., 2009). We therefore expected to find the tallest trees we sampled to have lower growth rates and higher mortality rates than shorter trees. The *detectree2* Python package is available to install and apply on new regions[1].

3.2 Materials and Methods

3.2.1 Study sites

The analyses were conducted at four locations across three tropical field sites:

1. Sepilok Forest Reserve (East and West), Sabah, Malaysia (5°50'N, 177°56'W)
2. Danum Valley Conservation Area, Sabah, Malaysia (4°57'N, 177°41'W)
3. Paracou Field Station, French Guiana (5°16'N, 52 °55'W)

Danum Valley hosts lowland tropical rain forests dominated by dipterocarp species that are among the tallest forests on the planet (Shenkin et al., 2019). The available data from Sepilok included ecologically distinctive areas to the East and West. Sepilok West consists mostly of tall forest (similar to Danum), while Sepilok East is a heath forest growing on shallow soils overlying sandstone, containing smaller, more densely packed trees (Coomes et al., 2017). All three sites in Malaysia experience a similar climate with approximately 2300 mm rainfall per year with the wettest months between November and February (Nilus et al., 2011).

Paracou

As the primary research described throughout this thesis was conducted in forests at Paracou Field Station, French Guiana (5°16'N 52 °55'W) (see Fig. 3.1), here I give a more comprehensive description of the site. The lowland tropical moist forests grow mostly on shallow ferralitic soils underlain by a variably transformed loamy saprolite (Gourlet-Fleury et al., 2004). The mean annual rainfall is approximately 3200 mm with a three month dry season from mid-August to mid-November. September, the driest month, will typically have 50 mm of rainfall (Bonal et al., 2008; Wagner et al., 2011) which makes the climate just within tropical monsoon regime

[1] https://github.com/PatBall1/Detectree2

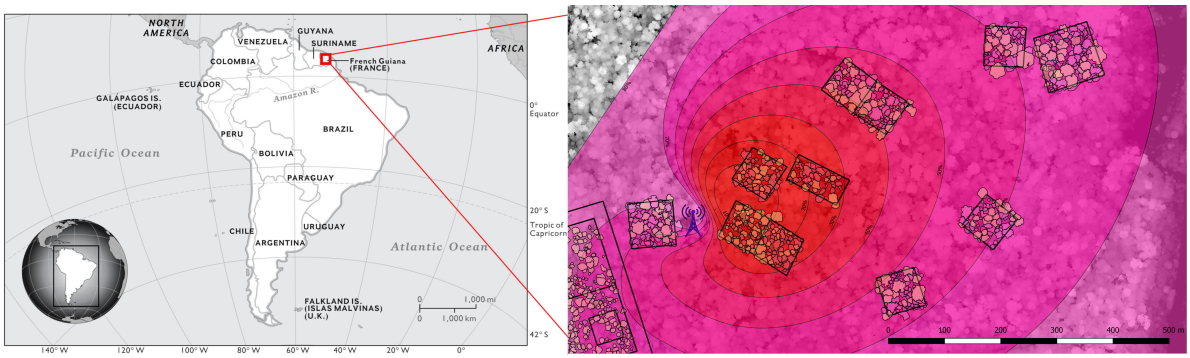


Fig. 3.1 The location of the study site with the flux tower footprint in pink and manual tree crown polygons. We observed the phenology of the tree crowns in this region with a view to eventually link the phenological patterns to productivity calculated from the measurements on the flux tower (see Chapter 5).

of the Köppen climate classification system (tropical rainforests have a minimum monthly precipitation of 60 mm; Köppen, 1884).

The field station has 27 permanent plots ranging in size from 0.5 ha to 25 ha which contain approximately 76,000 trees of DBH $\geq$ 10 cm consisting of over 800 different species (Gourlet-Fleury et al., 2004). In these plots, inventories are taken every 1-5 years with the species, precise geographic location and DBH of each trunk recorded. We generated a set of manually delineated crowns which were validated with subsequent fieldwork. By comparing *in situ* observations with the remote sensing data, we either matched the identified crowns to individual trees present in the site inventory (Gourlet-Fleury et al., 2004), or in cases where crowns were located outside of the known plots, we engaged botanists to assign the appropriate species. We updated the crown outlines and confidence scores based on the field observations and noted where there were liana infestations in the crowns. The fieldwork was focused on approximately 100 hectares of forest lying within the footprint of a flux tower used to measure gas exchange between the ecosystem and atmosphere. The footprint is defined as the region of forest around the tower which influences the CO_2 and water fluxes recorded by its sensors, and depends on wind speed and direction (see Section D.1 for details on how the footprint was defined).

3.2.2 Remote sensing data

Airborne RGB surveys were conducted at all four sites using crewed aircraft (Table 3.1). Repeat lidar surveys were also conducted at all four locations (see Table 3.1, noting different sensors and altitudes between flights in Sabah). We analysed RGB imagery from 3.85 km^2 of Malaysian forest, with a ground resolution of 10 cm. In Paracou, we sampled 10.2 km^2 of imagery, with an 8 cm ground resolution. The raw imagery was orthorectified, georeferenced and collated

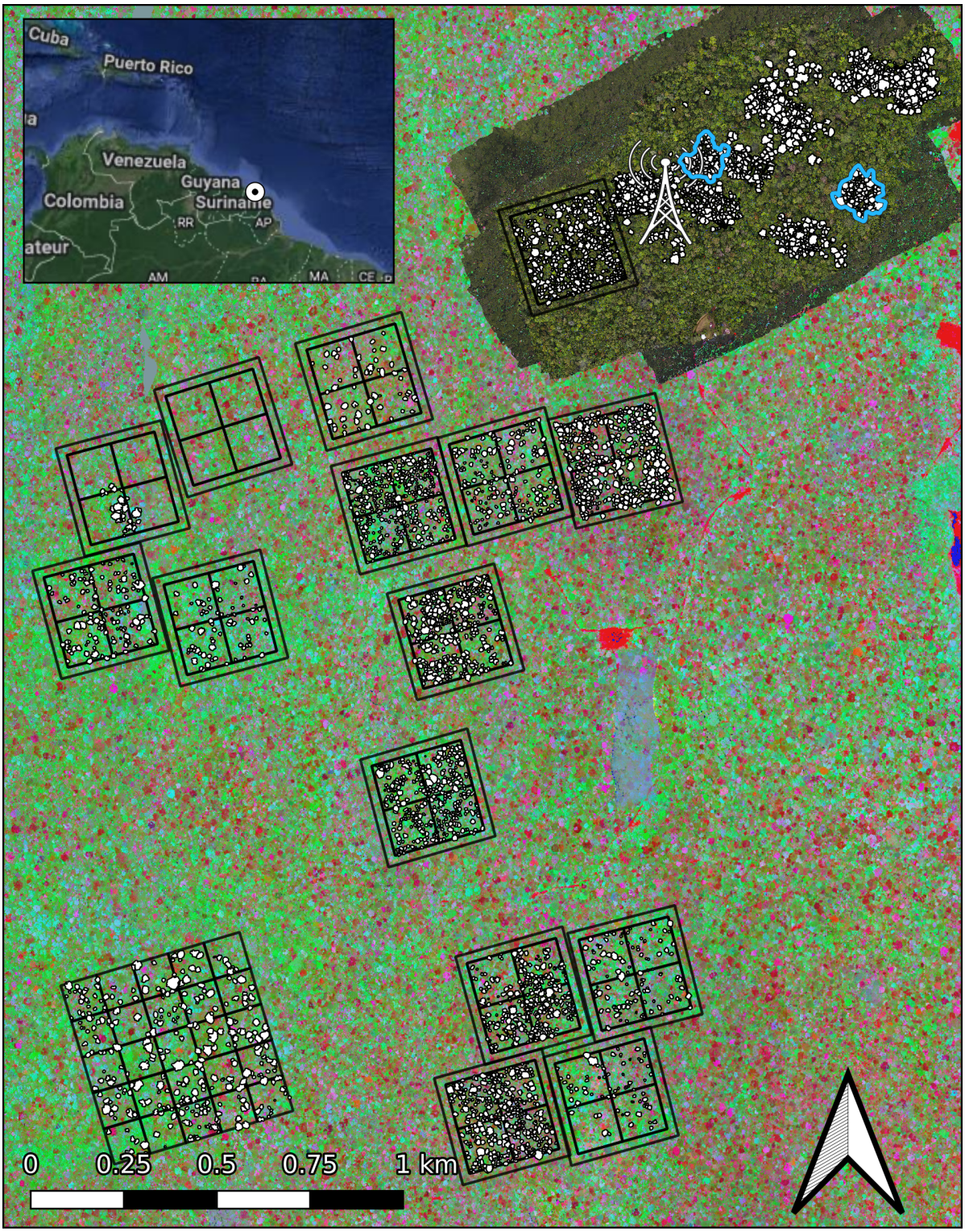


Fig. 3.2 Site map of Paracou, French Guiana, with crowns and hyperspectral imagery. The manually delineated, labelled crowns are in white. The colourful background scan that covers the entire site is a representation of the hyperspectral data (selected projected PCA bands). The repeat survey UAV-RGB region is shown in the northwest around the site's flux tower. Within this region the segmentation test areas are delineated in blue - the crowns within these areas were excluded from all training of the segmentation delineation. The black boxes show the plots in which inventories are conducted.

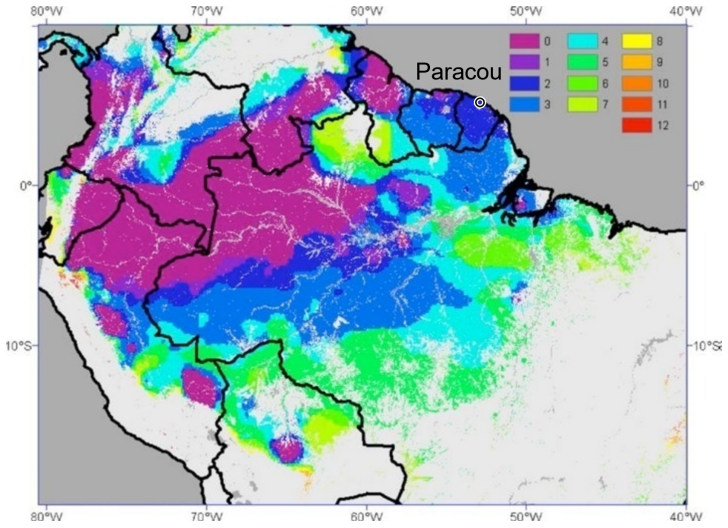


Fig. 3.3 The distribution of dry season length (as the number of months per year with < 100 mm precipitation) in tropical South America taken from Xiao et al. (2006). Added to the map is Paracou Field Station which receives an average of 3200 mm annual rainfall but experiences a moderate dry season from mid-August to November which may constrain and structure the leaf phenology seen at the site.

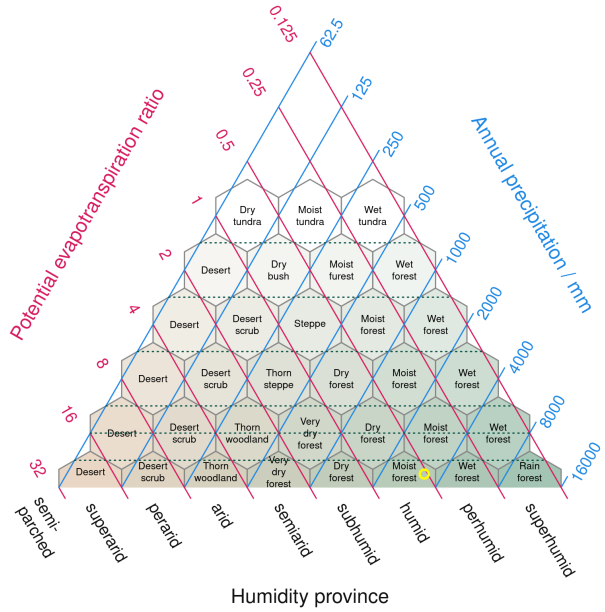


Fig. 3.4 Location of Paracou (yellow circle) within the Holdridge life zone classification scheme. Based on meteorological observations at the site from 2000-2022, Paracou can be classified as a tropical moist forest with a brief dry period.

Table 3.1 Remote sensing data sources used in the study. The exact location of the sites is described in Section 3.2.1. Resolution is given as ground resolution for the RGB imagery and as the processed CHM resolution for the lidar scans. Beam divergence is given at the $1/e^2$ points. Sepilok West and Sepilok East were separated for analysis due to the different characteristics of the Sepilok forest in these two areas.

Scan dates	Region	Modality	Resolution	Pulse density	Beam divergence	Scanning angle	Altitude	Sensor
23-Oct-2014	Danum	RGB	10 cm	-	-	-	796 m	Leica RCD105
01-Nov-2014	Danum	Lidar	1 m	5 pls m^{-2}	< 0.22 mrad	± 14°	2000 m	Leica ALS50-II
19-Feb-2020	Danum	Lidar	1 m	35 pls m^{-2}	< 0.5 mrad	± 30°	200 m	RIEGL LMS-Q560
10-Oct-2014	Sepilok	RGB	10 cm	-	-	-	796 m	Leica RCD105
05-Nov-2014	Sepilok	Lidar	1 m	16 pls m^{-2}	< 0.22 mrad	± 14°	2000 m	Leica ALS50-II
15-Feb-2020	Sepilok	Lidar	1 m	42 pls m^{-2}	< 0.5 mrad	± 30°	200 m	RIEGL LMS-Q560
19-Sep-2016	Paracou	RGB	8 cm	-	-	-	800 m	IXA180 Phase One
19-Sep-2016	Paracou	Lidar	1 m	35 pls m^{-2}	< 0.25 mrad	± 30 °	800 m	RIEGL LMS-Q780
15-Nov-2019	Paracou	Lidar	1 m	35 pls m^{-2}	< 0.25 mrad	± 30 °	800 m	RIEGL LMS-Q780

into homogeneous mosaics using structure from motion in AgiSoft Metashape (AgiSoft, 2021; Westoby et al., 2012) in Sabah. In Paracou, the imagery was orthorectifed using TerraPhoto to the Canopy surface model derived from simultaneously acquired lidar data.

3.2.3 Manual tree crown data

To train and evaluate our automatic delineation approach, we created a manually labelled dataset of trees at all four sites. We generated our delineations using both RGB and lidar data and, in the case of Paracou, supplementary hyperspectral layers. We used several techniques to improve the accuracy of crown delineation, including manipulating the contrast and saturation of the RGB image to exaggerate differences between the crowns, and using a mask of the lidar data to remove irrelevant parts of the RGB imagery. These techniques meant that the majority of tree crowns were separable by eye, however it should be noted that in rare cases, tree crowns were difficult to delineate with certainty and the labeller's best estimate was used.

The Paracou tree crown dataset was constructed and validated over four field missions. As it is used extensively in the rest of the thesis, I elaborate on this process here. The data was collected in blocks typically corresponding to the inventoried field plots (Gourlet-Fleury et al., 2004), and was developed initially at a computer remotely and then validated and edited on a tablet in the field. Initially, at a computer, the co-registered remote sensing data (including RGB, multispectral, lidar CHM and hyperspectral layers), the inventory data (with geolocated, species labelled trunks represented as points scaled by DBH) and site boundaries (including plot perimeters) are loaded into QGIS. By comparing between the remote sensing data layers, I would make a provisional set of polygons representing the boundaries of the individual tree crowns. Based on the size of the crowns relative to the trunks, I would provisionally assign

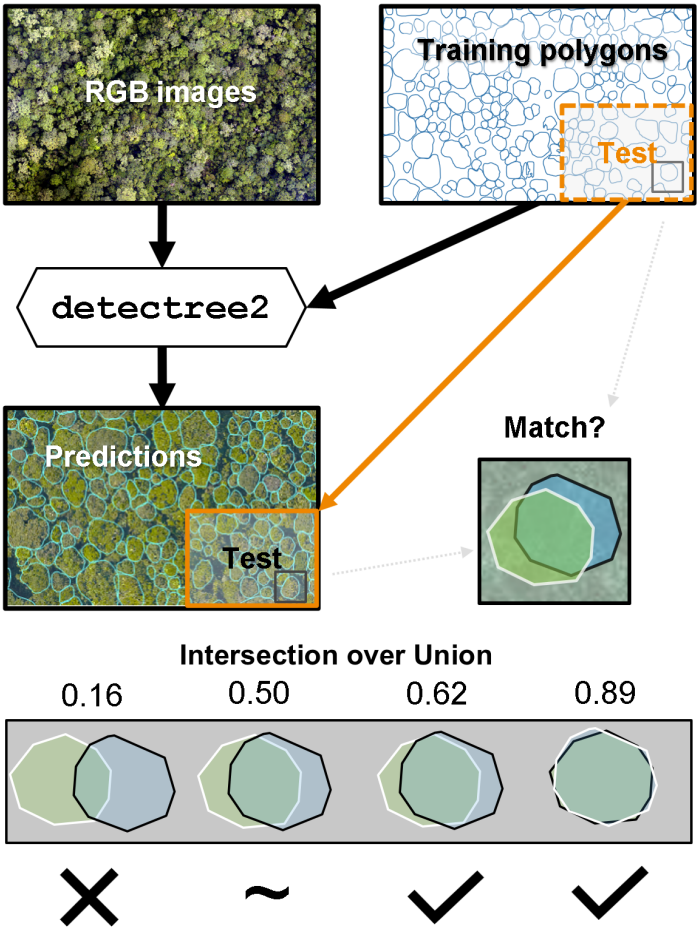


Fig. 3.5 The automatic tree crown delineation workflow. Manually delineated crowns are randomly split into training and test sets (though the figure suggests that the sets were determined geographically, this is purely for visual clarity). The Mask R-CNN framework combines the training set with the RGB imagery to learn how to delineate automatically from RGB images. A set of automatic predictions are produced across the entire RGB image and compared to the test set to evaluate the performance of the automatic delineations. Intersection over union (IoU) is used to determine when an automatic crown has been successfully matched with a manual crown.

a unique tree identification number from the inventory to each crown. The unique identifier would allow each crown to be assigned a species as recorded in the inventory. I also assigned two confidence scores to each polygon each ranging from 1-4 where 1 represented absolute certainty and 4 represented nearly complete uncertainty. The first score, *Crown Integrity*, represented the certainty with which the crown boundary was correctly delineated (i.e. a single whole crown rather than multiple crowns or a partial crown). The second score, *Trunk Match*, represented the confidence with which the crown had been correctly assigned to an individual in the inventory. After as much of the area of interest as possible had been labelled, the remote sensing, tree inventory and crown polygon datasets were transferred to a portable tablet (using the QField software package) for editing and validation in the field. Two to five people would go to the plot of interest, one inspecting and editing the dataset on the tablet and directing the activity, and the rest inspecting the canopy. When labelling trees that fell outside of the forest plots it was necessary to have a local botanist present to identify the species of the trees. By relaying information between the individual on the tablet and those inspecting the canopy, the polygon boundaries and tags could be edited based on observations to achieve a precise, ground validated tree crown dataset. The complex, multilayered canopy and propensity for trees to grow off vertical meant that having several people inspecting crowns from different positions while another who could inspect the bird's-eye view from remote sensing data was valuable for achieving a reliable, comprehensive dataset to train and test our methods on. For this reason, the Paracou tree crown dataset was considered our "gold standard".

We trained and tested our model with a total of 3797 manually delineated tree crowns across Paracou (1267), Danum (521), Sepilok West (1038) and Sepilok East (971). The crowns from Paracou were validated in the field with an expert local botanist, whereas the crowns in Malaysia were drawn by inspection of the remote sensing products. Four individuals performed the manual delineations which provided the network with variability in the inputs.

3.2.4 Data preparation

The orthomosaics and corresponding crown polygons were tiled into squares of approximately 100 m x 100 m to be ingested into the network (40 m core area, 30 m overlapping buffers). These dimensions reduced the chance of a crown being only partially represented in the dataset while also being small enough to hold in memory during training. To be included in the training and test sets, a minimum crown polygon area coverage of a tile was set at 40%. Including overly sparse tiles was likely to lead to poor algorithm sensitivity while being too strict with coverage would have limited the amount of training and testing data available.

If training and test crowns are close to one another, spatial autocorrelative effects are likely to inflate the reported performance (Kattenborn et al., 2022). In other words, if training crowns

are geographically close to test crowns it is likely that prediction accuracy will be misleadingly high as there is more similarity in remote sensing acquisition parameters and environmental conditions between the train and test data than would be expected by random chance. To avoid this, individual tiles (rather than individual crowns) were assigned to training and test sets ensuring spatial separation. Approximately 10% of the tiles from each site were reserved at random for testing. To avoid contamination of the test set, tiles with any overlap with the test tiles (including with the buffer) were excluded from the training set. The training tiles were further randomly partitioned into 5-folds for cross validation. This allowed for the tuning of parameters and the implementation of early stopping (see Section 3.2.6) without exposing the test set. Details of the data processing are described in Section B.3.

3.2.5 Model architecture and parameterisation

Instance segmentation combines object detection with object segmentation. Once an object has been detected in a scene, a region of interest (as a bounding box) is established around the object. A "segmentation" is then carried out to identify which pixels within the region of interest make up the object of interest (and which lie outside; see Fig. B.2 for an example).

We adapted Facebook AI's Mask R-CNN architecture as it was the best in class algorithm upon release for instance segmentation when tested on the Microsoft COCO (Common Objects in Context) benchmark (He et al., 2017; Lin et al., 2014) and has since been updated (as Detectron2) with improved training efficiency, documentation and transferability for integration into bespoke tools (Wu et al., 2019b). We adapted the Detectron2 computer vision library to handle geospatial inputs/outputs and perform the delineation of individual tree crowns. The library performs instance segmentation by generating object "masks" which exactly circumscribe the objects in the image (see Fig. B.2 for an example prediction). It also has a "model zoo"[2] from which specific model architectures with weights from a variety of pre-training regimes can be loaded. Taking a pre-trained model (weights) and retraining it to perform a novel task (e.g. delineating trees from aerial imagery) is an example of *transfer learning* which can drastically reduce the amount of training data required to achieve acceptable performance on the new task (Weiss et al., 2016). We selected the R101-FPN configuration[3] as it has "the best speed/accuracy tradeoff" of the architectures available (Wu et al., 2019b). Each object predicted by Detectron2 is associated with a confidence score which relates to how sure the network is in its prediction. A suitable threshold can be selected to optimise accuracy or balance precision and recall. Additional details are given in Section B.4 and for full technical

[2] https://github.com/facebookresearch/detectron2/blob/main/MODEL_ZOO.md

[3] https://github.com/facebookresearch/detectron2/blob/main/configs/COCO-Detection/faster_rcnn_R_101_FPN_3x.yaml

specifications, one should refer to the original papers and the Detectron2 repository (He et al., 2017).

3.2.6 Training and model selection

Training, tuning and model selection were performed with the five folds of training data tiles (see Section 3.2.4). To test the effect of volume and diversity of training data on performance we employed three training regimes: (1) Training on data of a single site, (2) Training on all sites ("combined"), (3) Training on all sites and then trained with a fixed training period on the single site. The idea behind (3) was to train on the full available data and then 'hone' the delineator based on the local context.

Typically, a deep CNN would require several thousand training examples to learn a new task. This is a challenge in the case of tree crown delineation as manual delineation is time consuming. The burden was reduced by transferring a model trained on a different instance segmentation task (Lin et al., 2014). Additionally, the training data were augmented by applying several randomly applied transformations to the training cases including vertical and horizontal flips, rotation, and varying the saturation and contrast of the image.

The model hyperparameters (see Table B.1) were tuned with a Bayesian hyperparameter sweep implemented on wandb.ai[4]. This is an automated process that allows an automated agent to iteratively adjust hyperparameters to optimise accuracy. The best performing models and optimal confidence threshold for a given model (see Section 3.2.5) were selected based on the F_1 score (see Section 3.2.7) on the validation fold. See Section B.4 for more details on model architecture, training and validation. The Colab (Jupyter) notebooks in the GitHub repository[5] illustrate the best practices for training and selecting models.

3.2.7 Performance evaluation

After tuning and training, the best performing models were taken forward to be evaluated against the test tiles. Matches between predictions and manual crowns (i.e. true positives) were identified by assessing the degree of spatial overlap between possible pairs. A minimum area threshold for valid crowns was set to 16 m^2. This removed fewer than 2% of manual crowns and introduced a level of consistency between sites and between the effort given by the manual delineators. The threshold was small enough to remove most inconsistencies between the lower area thresholds for inclusion of the human delineators while still including a representative

[4] https://wandb.ai/detectree/tune/sweeps/
[5] https://github.com/PatBall1/Detectree2

distribution of crown sizes for the analysis of the variation in performance by tree height. Crown overlap was calculated as the area intersection over union (see Fig. 3.5):

$$IoU(A,B) = \frac{A \cap B}{A \cup B} \tag{3.1}$$

where A is the automatically delineated crown area and B is the manually delineated crown area. An IoU of an overlapping pair of more than 0.5 was considered a match. This is a commonly used threshold in similar studies (e.g. Aubry-Kientz et al., 2019; Hao et al., 2021) that allows for small discrepancies in alignment and outline. While not rigorously tested here, it tends to give a match that feels intuitively acceptable by human interpreters (see Fig. 3.5). These true positives, the unmatched predictions (false positives) and the unmatched manual crowns (false negatives) were used to calculate the precision, recall and F_1 score of the automatic predictions.

Despite the best efforts of the manual delineators and selecting for tiles with high manual crown coverage, the manual crowns were inevitably an incomplete representation, so recall (fraction of relevant instances retrieved) was an insightful metric. However, to ensure balance with precision we used the balanced F_1 score metric to assess and compare the accuracy of the models. This approach is not biased by tree crown area and is widely used in tree crown segmentation studies (Braga et al., 2020; Hao et al., 2021). See Section B.5 for more details on the evaluation metrics.

To evaluate the performance of *detectree2* across tree heights, we assigned a height to each test crown (based on the median pixel value of the initial CHM within the crown) and arranged them into 5 m height bins. The shortest bin (0-5 m) at each site was iteratively merged with the next shortest bins until more than 10 individuals were represented (e.g. at Paracou 17 trees with a median height of 22.97 m fell into the lowest bin of 0-25 m). An equivalent process was used to define the highest bin at each site. The median tree height was calculated within each bin.

3.2.8 Transferability across sites

To determine whether models were able to generalise across different tropical forest areas, we evaluated the performance of the models when trained on one site and transferred to others. We compared these performances against the effect of using the "combined" training regimes described in Section 3.2.6.

3.2.9 Application to monitoring growth and mortality

We applied our best models for each site to their entire tiled orthomosaics (excluding the very edges where distortion is prominent) to generate site wide crown maps. We combined these crown delineations with repeat lidar surveys to determine the height changes in individual trees in our four sites. We determined the relationship between tree height and tree growth by fitting a robust least squares regression (Audibert and Catoni, 2011) to the data. Robust least squares was chosen to minimise the effects of outliers and mortality events on the regression. We note that here we are measuring the vertical growth of trees, instead of the growth in diameter at breast height (DBH) which is traditionally measured in forest inventory data. The link between our measured changes and those of the inventory measurements is not explicitly tested and an allometric relationship is assumed.

To estimate mortality rates, we needed a suitable metric to identify mortality events. We took a statistical approach defining a mortality event as a negative change in height of more than three standard deviations below the robust least squares fit. This allowed for the possibility that a mortality event may uncover another layer of vegetation rather than the forest floor. This choice was ratified by manual inspection of trees meeting this threshold, and confirming that they constituted mortality events. Annual rates were determined by dividing by the time between lidar scans.

Differences in lidar scanning parameters (pulse density, scanning angle, flight height etc.) can bias height estimates (Roussel et al., 2017). For this reason, we resisted reporting a direct comparison of reported growth and mortality rates between sites. As our focus here was on demonstrating the use of *detectree2* for locating crowns, we considered that a detailed exploration of the potential biases from the lidar data beyond the scope of the current paper.

3.2.10 Computation

Training deep CNN models can be computationally expensive and benefits from the availability of GPUs. Model training and evaluation was performed on the Google Colab (Pro) platform which employed Intel(R) Xeon(R) CPU at 2.30GHz with 12.8 GB RAM and Tesla P100-PCIE-16GB GPUs. On this platform, model training always completed within 2 hours.

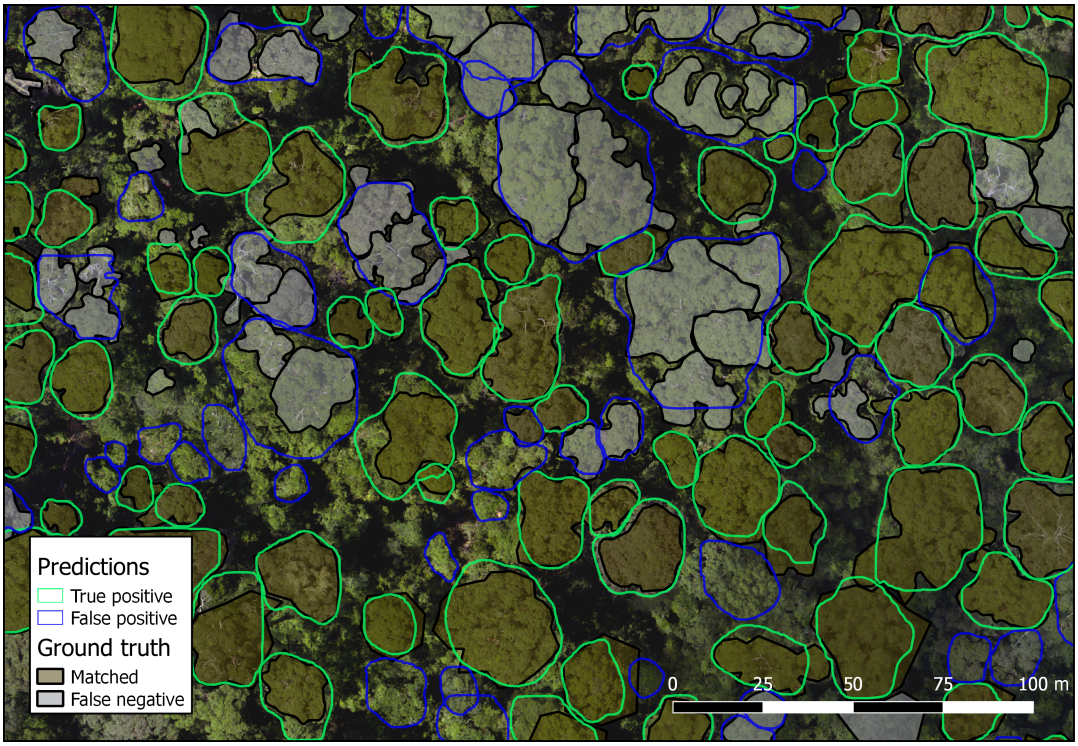


Fig. 3.6 An area of predicted crowns (transparent) overlaid on ground truth crowns (shaded with black outlines) at Danum. Colors and shading are used to indicate whether individual crowns have been successfully delineated. Some examples of under-segmentation (where a single prediction encompasses multiple ground truth crowns) and over-segmentation (where multiple predictions false try to split a single ground truth crown) are visible.

Table 3.2 Precision, recall and F_1 score of *detectree2* tree crown delineations by site as measured against the manual crowns of the test set tiles. The unweighted means of the metrics across individual sites are given as a summary overall performance.

	# test trees	Precision	Recall	F_1 score
Paracou	381	0.595	0.543	0.568
Danum	278	0.713	0.662	0.687
Sepilok East	167	0.612	0.653	0.632
Sepilok West	704	0.640	0.656	0.648
Average (sum)	*(1530)*	*0.640*	*0.629*	*0.634*

3.3 Results

3.3.1 Performance by site and tree height

Detectree2 located and delineated trees with F_1 score > 0.56 across all sites (see Table 3.2). It performed better in the tall dipterocarp dominated forests of Danum and Sepilok West and worse in the more compact forests of Sepilok East and Paracou. Indeed, Danum, the site with the best performance, had the greatest proportion of the tallest class of trees of any of the sites (see Table B.2 for full results). There was no apparent relationship between the amount of training data available at a site and the performance of the automatic delineator suggesting forest structure was the key determinant of accuracy. Where predictions were not accurate, it was slightly more likely to be from under-segmention (0.23-0.45) than over-segmentation (0.13-0.23) across all sites (see Table B.3; Clinton et al., 2010).

Across all sites, accuracy improved with tree height (see Fig. 3.7). This is likely due to the increased crown visibility of tall trees in the RGB images. Paracou has the least well differentiated canopy of all sites (it is relatively flat with crowns frequently interweaving) which may explain relatively poor performance there.

3.3.2 Performance between forest types

Danum and Sepilok West have tall dipterocarp dominated forests whereas Paracou and Sepilok East have a more compact forest structure. As we expected, performance degrades when testing a model on a different forest type to the one it was trained on (see Fig. 3.8a). For example, the performance at the forests of Sepilok West is significantly degraded when a model trained on Sepilok East or Paracou is used. In contrast, there is no drop in performance for predictions at Danum when the Sepilok West model is used and there is even a slight increase in performance for Sepilok East predictions when the Paracou model is used. In general, the model that was

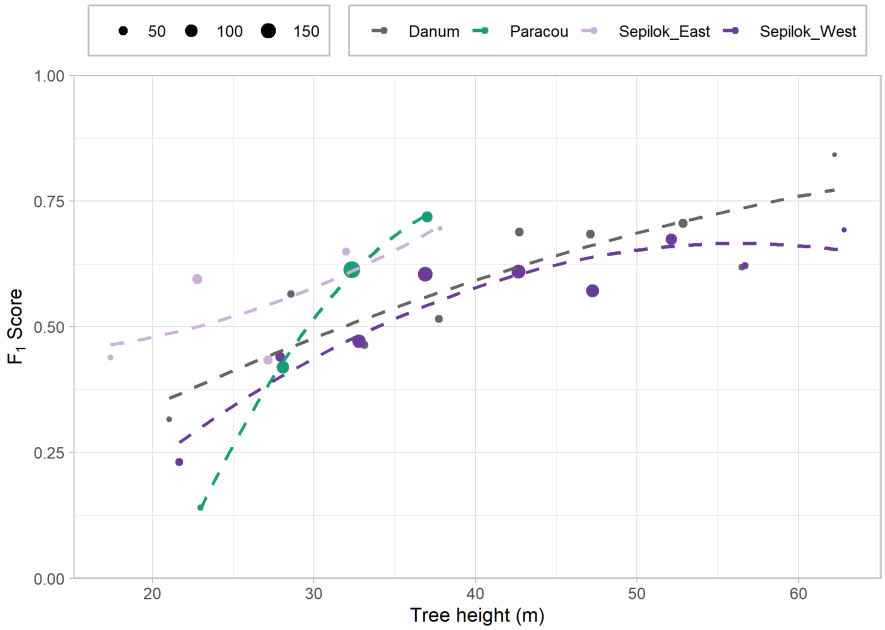


Fig. 3.7 F_1 scores of the tree crown delineations at the four different sites across tree heights. Bins of 5 m width were used to calculate F_1 score and corresponding median tree height. Point area is scaled by the number of test trees in the bin. Quadratic best fit lines (dashed) have been added to highlight the trends.

trained on all sites at once ("combined") outperformed the models that were trained on just a single site with the exception of Paracou (see Fig. 3.8b). Across the board, the best performing models were those that were exposed to data from all sites before being trained for a fixed number of iterations at the site to be predicted on. This suggests that providing a broad range of input data helps the networks to learn the key visual features but further tuning for local context helps maximise performance.

Application: Growth and mortality

One application of *detectree2* is to study tall tree growth and mortality rates. To do this, we overlaid *detectree2*'s tree crown predictions at the start date for each site on repeat lidar data (as canopy height models described in Section B.1) to retrieve the tree height dynamics over time.

We were able to estimate the relationships between tree height and tree growth for each site by fitting robust least squares linear relationships between the two variables for Danum, Paracou, Sepilok East and Sepilok West Fig. 3.9. The regression coefficients and intercepts are given in Table B.5. The growth rate decreased with tree height in all sites.

We assumed trees had died when their height decreased substantially. To evaluate this quantitatively, we fitted a robust least squares to the height change, against the original height

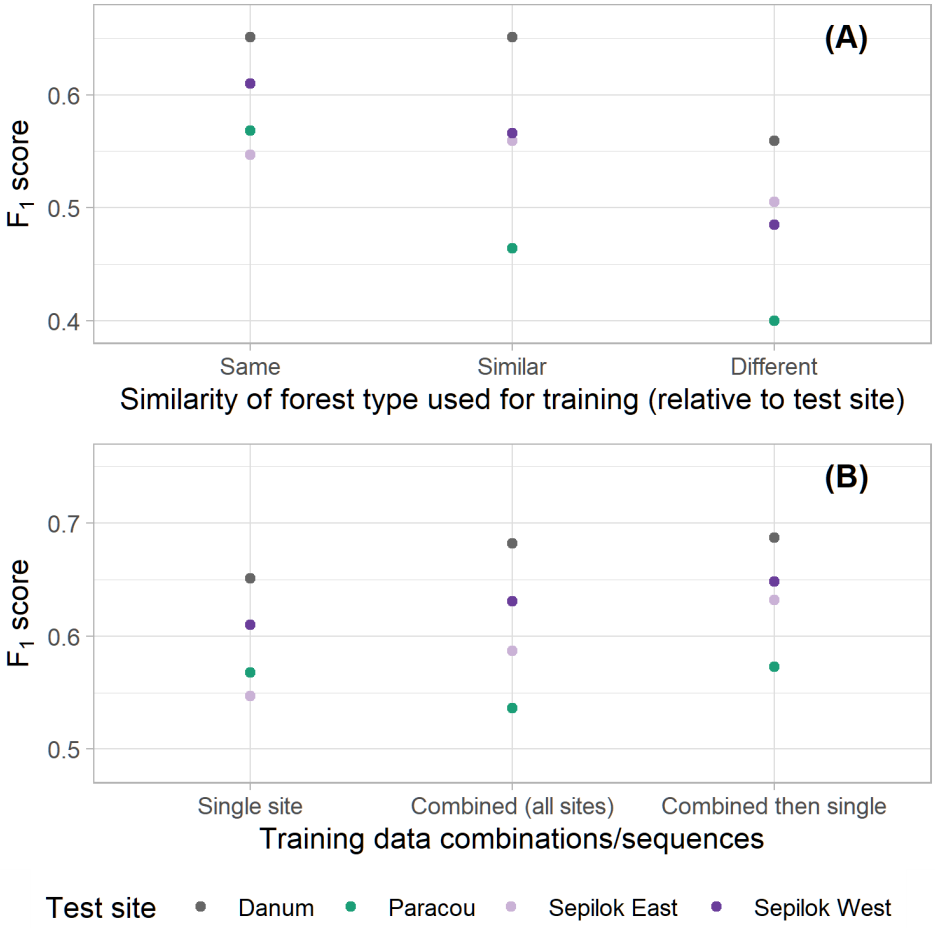


Fig. 3.8 Sensitivity of *detectree2* delineation accuracy to the variety of training data used. In **(A)**, "Same" indicates that training and testing took place at the same site. Sepilok West and Danum are "similar" forest types in that they are tall dipterocarp dominated forests in contrast to Sepilok East and Paracou that are shorter forests with a larger number of trees per hectare. As each site has two "different" sites and an average was calculated for the F_1 score. **(B)** shows the change in performance that occurs through employing different combinations of the training data. That can be just a single site, all sites at once ("combined") or all sites at once followed by a limited number of iterations on the site to be tested on.

of the tree, taking trees that were 3 standard deviations below the mean of the regression fit to be mortality events. The robust least squares regression differs from ordinary least squares as outliers contribute less to the regression fit. Therefore the robust least squares weights the fit towards those trees which did not suffer large height loss, and by taking the threshold to be 3 standard deviations we aim to identify only those trees that are outside the assumed normal distribution of typical tree growth and measurement error. Furthermore, as the robust least squares still incorporates outliers when fitting the data, three standard deviations was considered sufficient to identify mortality events. Fig. 3.9 illustrates how certain trees were identified as mortality events and some visual examples of mortality events from Paracou are given alongside (see Fig. B.12 for the other sites).

The mortality rates increased with tree height Fig. 3.10. The given uncertainty estimates were determined by bootstrapping. Table B.6 gives details of the growth and mortality rates.

3.4 Discussion

3.4.1 Improved tropical crown delineation

Accurately delineating trees in remote sensing data is a long-standing problem in ecology and conservation, and would enable us to efficiently monitor large areas of forests. *Detectree2* addresses this problem, delineating individual trees in aerial RGB imagery with a precision of 0.64 and recall of 0.63. We used *detectree2* to automatically delineate 65,786 trees across three tropical forests. We found that the accuracy of *detectree2* increased with tree height, meaning that the tall trees which store the most carbon are also the most reliably delineated. However, its relatively poor performance on shorter trees and inability to detect below canopy trees means that the sample of trees it detects is not representative of the full tree height distributions present.

Detectree2 performed well across a range of challenging, dense, closed canopy forests. It is able to exactly delineate highly irregular crowns within the jigsaw of the canopy rather than simply identifying a bounding box. This opens up new opportunities for tracking dynamic processes including growth and demographics (as demonstrated here) as well as phenology (where bounding boxes would risk mixing signals). Furthermore, *detectree2* is relatively accessible since it requires a low number of manually delineated trees as training data compared to other methods (Braga et al., 2020; Weinstein et al., 2019). These advantages are partly due to *detectree2* being built on a state-of-the-art pre-trained model. While direct comparison is impossible due to the different test data and the differing tasks (instance detection vs. segmentation), our method performs comparably to the results of Weinstein et al. (2019), which

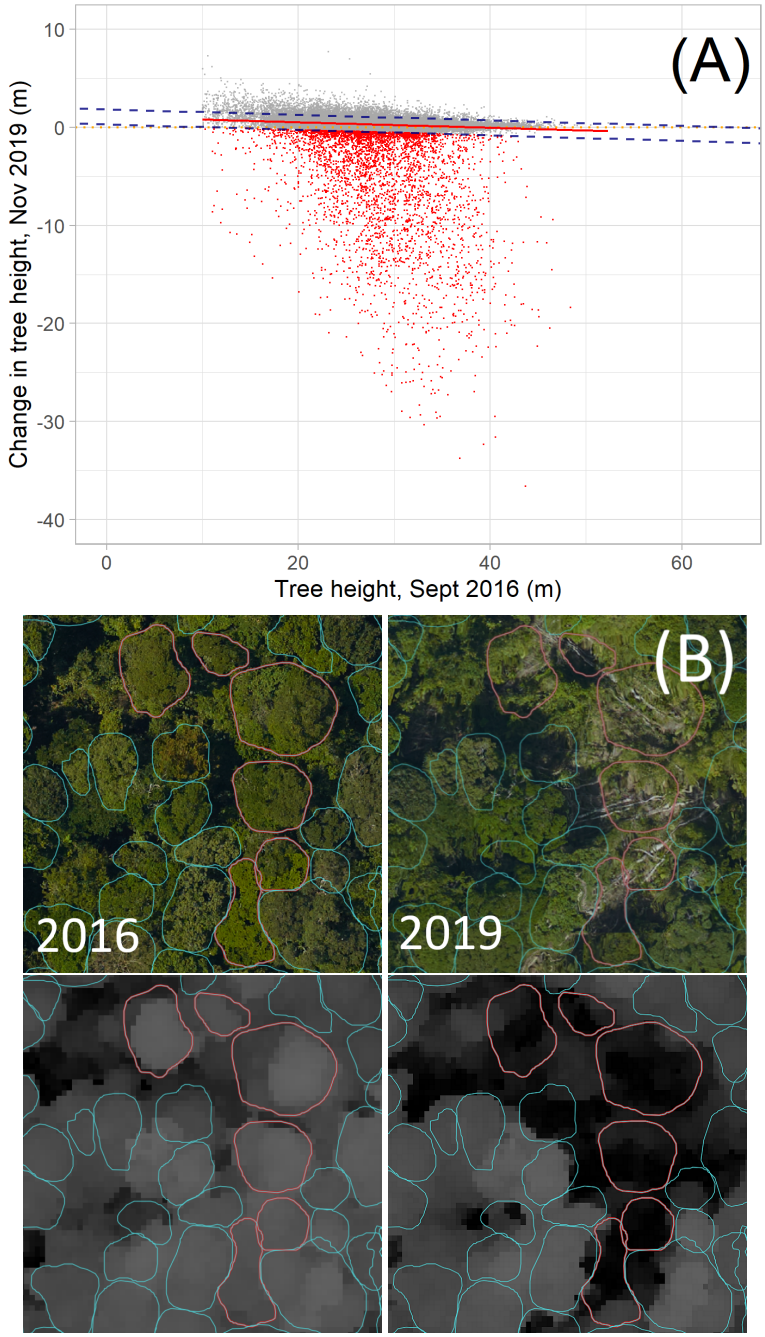


Fig. 3.9 Estimated tree growth and morality at Paracou with an visual example from the remote sensing data. **(A)** shows the robust least squares fit for change in height and tree height for Paracou. The dashed lines indicate three standard deviations either side of the best fit and red points below the lower bound indicate likely mortality events. **(B)** illustrates how predictions were overlaid on lidar data, and shows mortality events clearly visible in the lidar and the RGB imagery. The 2016 imagery is shown on the left, 2019 on the right. Crown delineations are based on the earlier imagery.

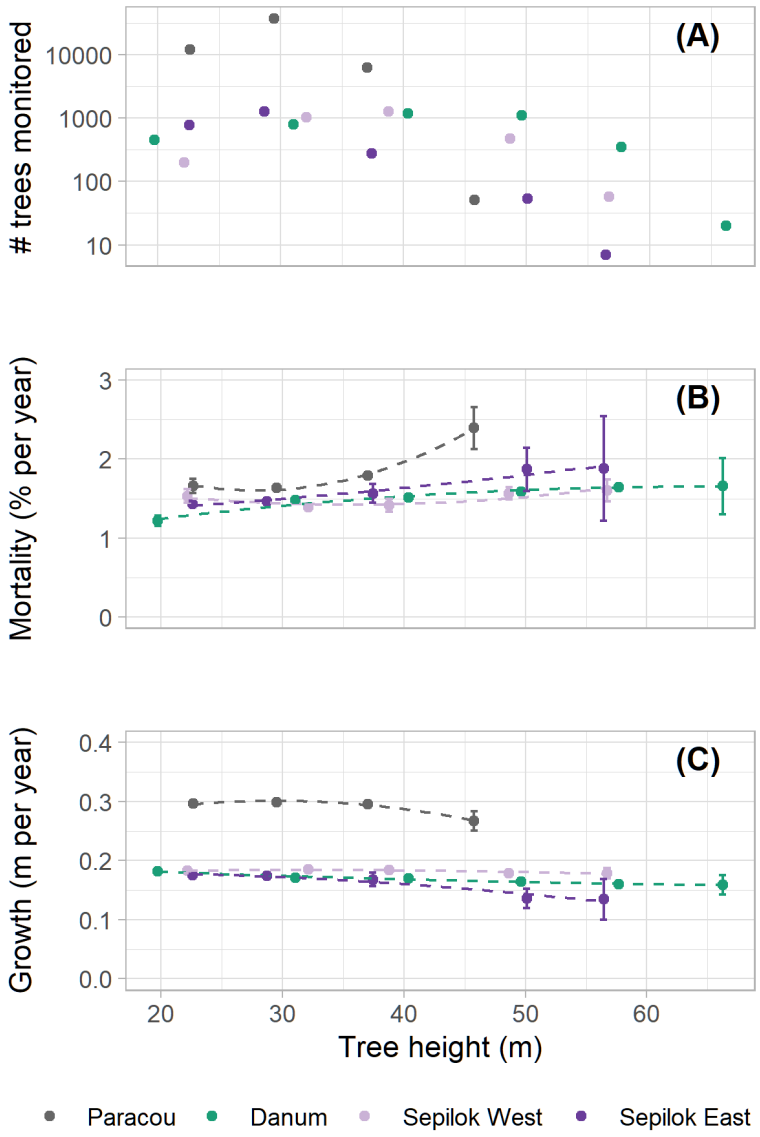


Fig. 3.10 Inferred mortality and growth relative to tree height across the sites. **(A)** shows the distribution of tree heights per site. **(B)** shows the mortality rates of trees of different heights in each site, and **(C)** gives the growth rate of trees split by height bin. Due to biases in tree height measurements that can arise from differences in lidar scan parameters we advise against a direct comparison of growth and mortality rates between sites. Uncertainty estimates of mortality and growth were determined by bootstrapping with 1000 resamples using the *boot* package in R. Quadratic best fit lines (dashed) have been added to highlight trends.

reported a tree crown recall of 0.69, a precision of 0.61 and an F_1 score of 0.65. Our results did not match the Mask R-CNN performance reported in Braga et al. (2020) but this study is based on semi-synthetic images (i.e. constructed by stitching together existing images) of forests and so is not directly comparable.

3.4.2 Generalisability across sites

There was no obvious relationship between the amount of training data available at a site and the accuracy attained there. Rather, forest type and tree height distribution seemed to be the key factors for determining accuracy. The well differentiated forest at Sepilok West and Danum were the easiest to delineate while the lowest accuracy was in Paracou which has little variation in the height of the visible canopy. Furthermore, at Paracou it is common to observe crowns mixing and growing into each other which makes visually separating the crowns challenging. This in turn is down to soil type and other biogeographic factors.

We found that the accuracy dropped when transferring a model trained on one forest type to predict on another. However, we found that *detectree2* can be quickly trained to perform well on new areas of forest using around 10 images (each ~1 ha in scale) with all visible tree crowns manually delineated. This manual delineation represents approximately 4 hours work. The best performing models were those that were exposed to training data from all the sites and then "honed" with a limited number of training iterations on the site to be predicted on. This suggests that our trained models (provided freely with the Python package[6]) could be transferred to a new site with very little manual data or training iterations.

We note that the manual delineations were done by different people focusing on different parts of the sites. There was no clear effect of different delineators on the results but this would be somewhat confounded with site differences.

3.4.3 Application: Growth and mortality rates

Tall trees store the majority of forest carbon and dominate many important forest nutrient cycles. However, they are rare and therefore poorly represented in traditional field inventories (Hurst et al., 2011) which makes estimating their growth and mortality rates particularly challenging (Coomes et al., 2003; Iida et al., 2014; Muller-Landau et al., 2006; Richardson et al., 2009). Tall trees are also particularly sensitive to the effects of climate change, such as increased wind speeds and drought (Gora and Esquivel-Muelbert, 2021), and as such, tracking their dynamics over time is increasingly important. Recent remote sensing studies are bringing new insights into disturbance patterns by mapping the gaps left in the forest canopy after a tree (or multiple

[6]See https://github.com/PatBall1/Detectree2

trees or branches) have fallen (Araujo et al., 2021; Cushman et al., 2022; Huertas et al., 2022). Tracking individual trees over time instead of gaps will make it easier to interpret our results in an ecological context and also to compare the results more directly to the available field inventory data.

Across all sites taller trees had higher mortality rates and lower growth rates. This aligns with large scale analyses of field-based studies (Iida et al., 2014). The apparent higher growth and mortality rates in French Guiana as compared to the sites in Malaysia was potentially a result of biases introduced to the variation in scan parameters (flight height, pulse density, time of year etc.) and so the values should not be directly compared across sites. Inventory data shows that mean DBH growth for trees at Paracou was 1.2 mm/yr (Wagner et al., 2010) compared to 0.9 mm/yr in Sepilok East, 1.1 mm/yr in Sepilok West and 0.5 mm/yr in Danum (Ordway et al., 2022; Piponiot et al., 2022). We note that these inventory measured DBH growth rates may not directly correlate to the height growth measured in this study. Another caveat is that we defined mortality as a drop in height of more than a statistically determined threshold. We do not verify directly that the tree has died, although it is likely that it has snapped or uprooted. Further analysis would help to understand the discrepancy in observed height change at Paracou in comparison to the other sites but is not the focus of the current study. Nevertheless, we believe this example application demonstrates the utility of *detectree2* in expanding the sample of trees under observation.

3.4.4 Future methodological developments and applications

Detectree2 performs impressively when delineating tall trees but it fails to delineate a significant proportion of trees (31%-46% of visible crowns and no sub-canopy crowns). There is considerable scope to increase the quantity and variety of training data by labelling more trees by hand. An approach to compensate for shadowed regions may also support the detection of trees otherwise obscured by their neighbours.

The fact that *detectree2* can be quickly trained to perform well on a new type of forest and imagery demonstrates that it is a useful tool for forest management. Many conservation or restoration projects have access to low-cost imagery from drones or satellites. *Detectree2* would allow them to quickly quantify and track the number and size distribution of trees across an entire landscape. In combination with other remote sensing data sources, this could allow for improved carbon stock and dynamics estimation. Estimating carbon stocks in forests has traditionally been done using area-based methods which discard considerable granular information at the individual tree level (Coomes et al., 2017).

We focused on aerial RGB imagery which is the cheapest and most widely available imaging source for tropical forests. We also benefited from the variety of pre-trained models that come

with this data type. However, different data sources may provide additional information that would help to discern differences between crowns. In particular, multi-spectral imagery that typically includes additional bands in the near-infrared is commonly used to study differences between trees due to the optical properties of vegetation (Knipling, 1970). Alternatively, the canopy surface (a raster expressing the height of the canopy) is commonly used in traditional segmentation techniques (e.g. watershed algorithms) and is produced with photogrammetry as a step in generating an orthomosaic. Including this as a layer would add an additional dimension of information that could help to distinguish fine differences in structure. It would be straightforward to include additional (or different) bands to the *detectree2* framework but it would forego the utility of the pre-trained models. Therefore, it is likely that significantly more training data and computational resources would be required to train a model (from scratch) to the desired performance.

Ideally, we could apply this approach to satellite imagery to perform global analyses. Preliminary tests suggest that *detectree2* can accurately delineate trees in RGB imagery at 2 m resolution (see Section B.8) which is equivalent to modern high-resolution satellite imagery. If this proves possible it will help answer many long-standing questions in forest ecology as well as provide an important tool for forest management. A "random resizing" augmentation step would further help improve generalisability across resolutions and incorporating "small object" detection features (Tong and Wu, 2022) would improve the sensitivity to shorter trees.

While we studied the delineation of a single class (*tree*), *detectree2* can be trained and make predictions on multiple classes. A next step will be to test the ability of the network to identify and map species, and to assess liana infestation occurrence in crowns. Previously, hyperspectral data has been employed to address this problem but with limited success due to the phylogenetic and spectral diversity of lianas and relatively low spatial resolution of hyperspectral imagery (Grabska et al., 2020; Wessel et al., 2018). The availablility of *detectree2* as an open-source Python package means other research groups can test its efficacy on their own research questions.

Chapter 4

Harnessing temporal and spectral dimensionality to map and identify species of individual trees in diverse tropical forests

Abstract

To understand how tropical rainforests will adapt to climate change and the extent to which their diversity imparts resilience, precise, taxonomically-informed monitoring of individual trees is required. However, the density, diversity and complexity of tropical rainforests present considerable challenges to remote mapping and traditional field-based approaches are limited in scale. This study introduces a new approach for mapping tree species using a convolutional neural network method (*detectree2*) to segment tree-crowns from aerial photographs and Linear Discriminant Analysis (LDA) to identify species from hyperspectral data (416 - 2500 nm). We build upon previous work in two ways. Firstly, we aimed to improve the accuracy of crown delineations by surveying the same patch of forest with UAV-RGB ten times over six months and fusing multi-date information on the location and shape of individual trees. Secondly, we extended the scope of species identification to include more species than has been previously attempted. We trained and tested our algorithms on subsets of a database of 3600 ground truth, labelled tree crown polygons representing 250 species in French Guiana that we had delineated by hand and field verified. We assessed how well our segmentation approach could locate and delineate individual tree crowns and how well our classification approach predicted the species of those crowns. We extracted information on waveband importance for species separation from our classification model. Based on an existing phylogeny of the trees in our dataset, we tested for phylogenetic signal across the hyperspectral bands and probed how species were being classified by comparing the phylogenetic signal to the importance of bands for separating species. The accuracy of delineations increased gradually as additional dates of tree crown maps were stacked and combined. Stacking increased the F_1-score from 0.69 to 0.78. The overall (microaverage) F_1-score for species classification was 0.76. A total of 64 species were predicted from the hyperspectral data with F_1-score > 0.7. The performance for classifying a species increased with the number of crowns in the database available for that species: just 10 training crowns were needed to achieve an expected F_1-score = 0.7 for crown level classification. With this new approach, we assessed that 70% of tree crown area at landscape scale was accurately mapped. The most important wavebands for discriminating species were dominant and narrowly clumped on the NIR side of the red edge region (748 - 775 nm). While most wavebands showed some phylogenetic signal, waveband importance

for species classification was negatively correlated with phylogenetic signal. Our integrated approach makes a significant contribution to the ongoing development of efficient and accurate methodologies for mapping canopy tree species in tropical forests, providing a framework for mapping trees in diverse tropical forests that is far more comprehensive than its predecessors.

4.1 Introduction

Moist tropical forests are renowned for their species richness. The ability to map tree species using remote sensing data is of value to ecologists and in other disciplines (Fassnacht et al., 2016). It enables real-time assessment and monitoring of biodiversity and species compositions (Shang and Chisholm, 2014) and can provide insights into habitats (Jansson and Angelstam, 1999; Kennedy and Southwood, 1984; Pausas et al., 1997), tree community dynamics and ecosystem function (Chambers et al., 2013; van Ewijk et al., 2014). Such data can support conservation efforts, by locating species of interest (Baldeck et al., 2015), improving estimates of stored carbon (Bredin et al., 2020), tracking invasive species (Chance et al., 2016; Sabat-Tomala et al., 2020), helping to manage water stress (Asner et al., 2004; Watt et al., 2021) and the spread of pests and disease (Chan et al., 2021; Liu et al., 2021), and improving our understanding of migration patterns (McGrath et al., 2009). Maps of canopy trees are also valuable for assessing the extent to which the extraordinary diversity of tropical forests influences ecosystem processes (Reichstein et al., 2013), including resilience to climate change (Corlett, 2011; Malhi et al., 2008) through the presence of species with varied climatic tolerances (Lewis et al., 2009). However, although it is now firmly established that diversity can promote ecosystem stability (Loreau and de Mazancourt, 2013), it remains unclear whether such diversity is necessary, because of functional redundancy among species (Biggs et al., 2020). Hence, an integrative understanding of the distribution of tree species diversity is vital in understanding the impacts of anthropogenic change and crafting effective strategies for the sustainable management of these globally significant ecosystems (Goetz et al., 2009). From a commercial standpoint, remote sensing of tree species can improve the speed, scope and precision of forestry inventories (Laybros et al., 2020; van Aardt and Wynne, 2007) which aid in calculating available resources and planning for sustainable harvests (Vauhkonen et al., 2014). Together, these insights can support evidence based policy for sustainable forest management, balancing human activity and ecological preservation (FAO, 2020).

Mapping individual tree crowns and identifying their species at large scale in diverse and dense tropical forests presents a significant challenge (Asner and Martin, 2011), requiring accurate delineation of tree crowns and classification of pixels within those crowns. The first step of locating and delineating individual tree crowns in densely packed tropical forests from

above is difficult as crowns can interweave in complex ways. Remote sensing of individual trees has mostly focused on airborne lidar data, but this is most successful for temperate and boreal forests (Dalponte and Coomes, 2016; Hastings et al., 2020) while complex tropical canopies have presented a greater challenge (Aubry-Kientz et al., 2019). RGB photographs offer colour and texture information to distinguish trees, even if they are structurally similar, but agreement on tree location and crown shape between human analysts can be variable (see Section C.1). Mask R-CNN (He et al., 2017) based approaches have shown promise in their ability to harness the colour and texture information of RGB images to precisely separate irregular edges of neighbouring tree crowns (Ball et al., 2023; Gan et al., 2023). However, differences in illumination (resulting in variably shadowed trees), the sway of trees and branches, phenological discrepancies and irregularities in the orthomosaicking process often reveal inconsistent arrangement/delineation of crowns for aerial imagery across dates. Methods that can draw on consensus between datasets and are tested against a robust ground truth are required.

Assigning species labels to the tree crowns is also challenging due to high species diversity, the rarity of the majority of species and the prevalence of common photosynthetic pigments. Tropical forests have approximately log-normal species abundance distributions: a few species are common, while most are represented by only a few individuals per hectare. Indeed, a pan-Amazonian study based on 640,000 trees (DBH $\geq$ 10 cm) measured in 1100 ha of forest plots, 36 % of species were modelled to have a population size of fewer than 1000 individuals across the whole of the Amazon, while 1.4 % of species were estimated to account for half of all trees (ter Steege et al., 2013). This skewed distribution presents challenges for biodiversity conservation efforts, as it means that many species are always at risk of extinction due to their small populations. It also complicates efforts to identify and study individual tree species using remote sensing techniques, as the large number of rare species can be difficult to reliably detect and classify. While there is vast species diversity, the presence of chlorophyll and other photosynthetic pigments is ubiquitous across trees. This further complicates remote species identification, as distinguishing between individual species based on their pigmentation alone is not feasible. Closely related species often share certain chemical, physiological and morphological characteristics that can influence their spectral properties (Cavender-Bares et al., 2016; Meireles et al., 2020). For instance, the leaf structure and chemical properties (e.g. concentrations of chlorophyll and cellulose) can vary between families which in turn can affect their spectral properties (Serbin et al., 2016). However, there can still be considerable variability in the spectral properties of plants within the same family due to differences at the genus and species-level, as well as due to environmental factors (like soil type, water availability, and light exposure) and plant health status or phenological stage (Ollinger, 2011). So, while it is

possible that members of the same plant family might exhibit some similarities in their spectral properties, these are not definitive, and a comprehensive classification of plant species based on spectral data would likely need to take into account more specific characteristics and use sophisticated machine learning algorithms to capture these complex patterns (Fassnacht et al., 2016).

Hyperspectral remote sensing (imaging spectroscopy) is often applied to mapping tree species in forests (Fassnacht et al., 2016; Ghiyamat and Shafri, 2010). The spectral properties of plants are influenced by numerous factors, including their evolved biochemical and structural characteristics (Meireles et al., 2020; Ollinger, 2011; Ustin, 2013; Ustin et al., 2004). Many studies have shown that subtle differences in spectral reflectance arising from these biophysical differences can be detected in hyperspectral data, allowing species to be mapped and their health monitored in low diversity (often temperate) systems (Fassnacht et al., 2016). However, distinguishing species in diverse tropical forests has proven more challenging; Clark et al. (2005) used hyperspectral data from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) to differentiate seven species in a tropical rainforest in Costa Rica, paving the way for further research in this domain. Féret and Asner (2013) were able to distinguish 17 species in the lowland humid tropical forest of Hawaii. Laybros et al. (2020, 2019) showed that a classification rate of around 80% was achievable when classifying within a pool of 20 well represented species in an Amazonian forest in French Guiana. Greater scope has been shown with proximate, leaf level hyperspectral species identification (e.g. 46 species at a tropical wetland in Jamaica (Prospere et al., 2014)) highlighting the potential for a broader scope with aerial hyperspectral data at the crown level. However, leaf traits are incorporated into spectra in complex ways (Féret and Asner, 2011; Jacquemoud and Baret, 1990; Ollinger, 2011) and multiple traits can overlap in a given spectral region (Curran, 1989). This is further complicated when spectra are measured from a distance as signals of relevant traits are confounded with structural characteristics. Radiative transfer models have shown that signals from canopy structure can dominate over the leaf optical properties and biochemical properties of the vegetation (Béland and Kobayashi, 2024; Knyazikhin et al., 2013). However, elsewhere they have been used to show that incorporating the fraction of non-photosynthetic vegetation as leaf brown pigments rather than as woody material gives the better match between simulated and observed canopy spectra (Ebengo et al., 2021). The degree to which these uncertainties place limits on the predictability of tree species from hyperspectral data is as yet unclear. Structural properties such as leaf density, clumping and angle distribution, as well as leaf biochemistry, may, to some extent, be considered functions of species. Rather than attempt to reconstruct a physical representation of the canopy, we test the separability of species through a data driven analysis of the spectra at a single site and point in time. Unpicking the spectral signals

in relation to the evolved traits and how species relate to each other is a key challenge in progressing towards robust, transferable remote identification of species in diverse tropical forests (Schweiger et al., 2021).

Applying computer vision approaches in a way that combines information across time may lead to improved detection and segmentation of canopy trees (Martin et al., 2018; Shamaoma et al., 2023). While RGB data lack the spectral resolution of hyperspectral data, its spatial resolution is typically superior, especially when acquired from a drone that can fly close above the forest canopy. This means that textural information can be observed and outlines of individual tree crowns can be distinguished precisely. The relatively low cost of the sensors and UAV systems also means that regular surveys of the same area become feasible. Differences in the canopy (including phenological states), atmosphere, and illumination can mean that predicted crown maps vary considerably across dates. To address this temporal variability, we implemented a consensus-fusion approach to combine tree crown polygons detected on different dates. The tree crown polygons from individual dates were spatially matched and the vertices of their parameters averaged, generating output crowns maps that represented the inter-date consensus on the location and shape of tree crowns. Our hope is that this approach would help to mitigate the effects of temporal fluctuations, enhancing the consistency and accuracy of crown delineation.

In this study, we propose a novel approach to rainforest tree mapping that (1) generates precise delineation of individual tree crowns and (2) classifies the species of each tree. The automatic delineation of individual tree crowns was performed with *detectree2*, a tool based on the Mask R-CNN architecture (Ball et al., 2023; He et al., 2017). For the first time, we combine delineated tree maps generated from repeated airborne surveys and analyse whether identifying consensus across stacked tree-crown maps improves segmentation accuracy. Our species identification model was trained and tested using a database of 4,000 manually delineated tree crowns, each verified through several field missions. This dataset, in combination with our novel approach, sets the stage for a more accurate and inclusive tree mapping and species classification system. This two-step approach leverages the strengths of both technologies - the spatial resolution of UAV RGB imagery and the spectral resolution of hyperspectral data - achieving a high level of accuracy in tropical forest mapping.

We address the following research questions:

1. Can the accuracy of tree crown maps from aerial imagery be improved by combining information from segmentations at different dates to build a consensus-based map?

2. Which type of machine learning classifier can most accurately predict the species of tree crowns from hyperspectral data?

3. How many mapped individuals of a given species are needed to achieve a 'good' classification accuracy?
4. Which wavebands are the most important for determining species?
5. Do spectra observed within tree crowns exhibit phylogenetic signal?
6. Are the most important wavebands for species classification those that have the strongest phylogenetic signal?
7. Are closely related species more often confused in their classification than distantly related species?

Through a novel integration of ground validated manual tree crown generation and machine learning algorithms, we developed a robust and accurate methodology to delineate tree crowns and predict their species. Our methodology combines traditional fieldwork, advanced machine learning techniques, and high-resolution remote sensing data, offering an innovative approach to species-level forest mapping.

4.2 Materials and Methods

4.2.1 Study site and overview of methods

The research was conducted in forests at Paracou Field Station, French Guiana (5°16'N 52 °55'W) (see Fig. 3.2). The lowland tropical rainforests grow mostly on shallow ferralitic soils underlain by a variably transformed loamy saprolite (Gourlet-Fleury et al., 2004). The mean annual rainfall is approximately 3200 mm with a three month dry season from mid-August to mid-November where rainfall is typically less than 50 mm per month (Bonal et al., 2008; Wagner et al., 2011). The field station has 27 permanent plots ranging in size from 0.5 ha to 25 ha (see Fig. 3.2) which contain approximately 76,000 trees of DBH $\geq$ 10 cm consisting of over 800 different species (Gourlet-Fleury et al., 2004). In these plots, inventories are taken every 1-5 years with the species, precise geographic location and DBH of each trunk recorded. The ten most common species account for just over 30% of the inventory's individuals. 90% of the species present have been placed within a time calibrated phylogeny by Baraloto et al. (2012). Full details of the site are given in Section 3.2.1.

Our study uses a combination of remote sensing data from a UAV-mounted camera and a plane-mounted imaging spectrometer (Table 4.1). These data were co-registered to a LiDAR-derived Canopy Height Model (CHM) with affine transformations based on matched manual

control points, ensuring precise alignment. A CNN approach to UAV-RGB data recorded with 10 surveys over 6 months was used to locate and delineate the individual tree crowns present in the landscape. Hyperspectral imagery was used to classify the species of the crowns. The imaging spectrometer measures reflected light in a broad range of wavelengths across the electromagnetic spectrum, often with hundreds of bands, each corresponding to a narrow wavelength range (Ustin et al., 2004). In this study, the measured wavelengths were in the range 400-2500 nm and there were 400 spectral bands. Each tree species absorbs and reflects light in a specific way across these wavelengths, generating a spectral 'signature' that can be used to distinguish between species (see Fig. 4.5). In the context of tropical forests, the application of hyperspectral data is particularly advantageous due to the high species diversity and variability in physical (including crown structure) and chemical characteristics among trees.

The predictions were evaluated against strict, unseen test sets of manual tree crowns that were not exposed to the algorithms during training, providing a robust assessment of model accuracies.

4.2.2 Remote sensing data acquisition and co-registration

Table 4.1 Remote sensing data sources used in the study. The exact location of the sites is described in Section 4.2.1. Resolution is given as ground resolution for the RGB orthomosaic and as the processed CHM resolution for the lidar scans. Altitude is given as height above canopy.

Scan date(s)	Modality	Resolution	Altitude	Spectral range	Sensor
23-Oct-2020—06-Apr-2021	RGB	5 cm	70 m	421-617 nm (3 bands)	1" CMOS (Phantom4 Pro)
19-Sep-2016	Hyperspectral	1 m	900 m	416-992 nm (160 channels)	Hyspex VNIR-1600
19-Sep-2016	Hyperspectral	2 m	900 m	930-2500 nm (288 channels)	Hyspex SWIR-384
15-Nov-2019	LiDAR	0.5 m (CHM)	800 m	1550 nm (active)	RIEGL LMS-Q780

UAVs (DJI Phantom 4 Advanced and DJI P4 Multispectral) were employed to collect high-resolution RGB imagery, with a scan approximately every three weeks over a 6-month period (10 surveys in total) of the region shown in Fig. 3.2. We used an earlier airborne LiDAR dataset to assist in the positioning and alignment of data, providing a baseline layer for integrating and interpreting other remote sensing data.

The RGB orthomosaics were compiled from the raw geotagged UAV photographs using structure from motion (SfM) photogrammetry in AgiSoft Metashape. The software aligns overlapping images to produce a sparse point cloud, refines it into a dense point cloud, and subsequently constructs a 3D mesh. This mesh, integrated with original photo textures, facilitates the creation of a Digital Elevation Model (DEM). The DEM, combined with the aligned images, allows for the generation of an orthomosaic, a georeferenced image free from perspective distortions. Supplying images across several dates in single blocks to the first steps of the SfM

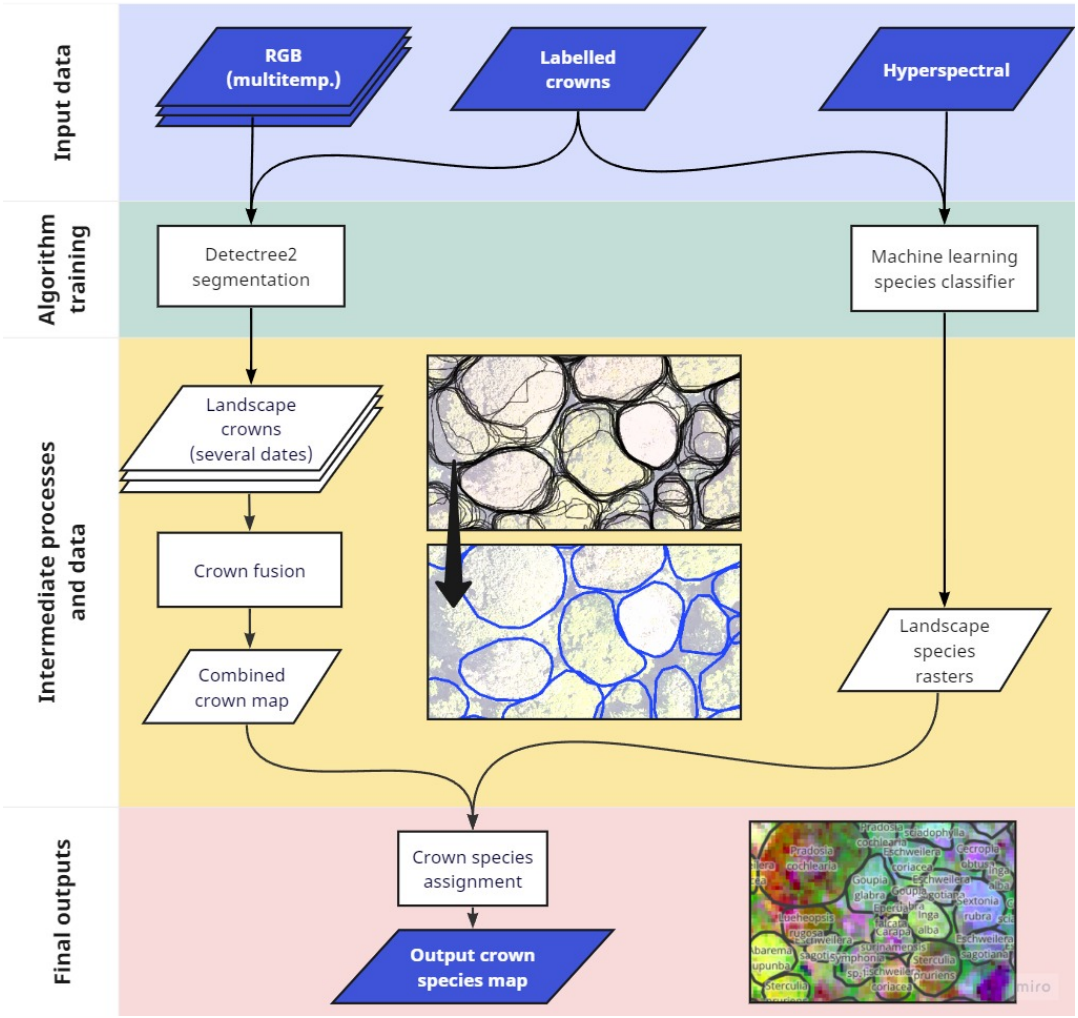


Fig. 4.1 Simplified schematic of the crown mapping approach showing input data and the intermediate steps to producing a labelled tree crown map. In the centre is an illustration of the process of temporal polygon fusion. The top image shows the overlapping tree crown polygons predicted over multiple dates. Where polygons have a high degree of overlap, each will still have a slightly different shape due to differences in the RGB orthomosaics through time. The bottom image depicts the polygons after the fusion process, effectively averaging the positions of the vertices of the original polygons and discarding those without good consensus through time. The averaging is weighted by the confidence score of the input polygons.

processing improves spatio-temporal coherency (Feurer and Vinatier, 2018b). Following this approach, instead of processing each date separately, five date blocks were supplied for the alignment and initial sparse point cloud formation establishing a common geometry between dates. The dates were then separated for the dense matching steps and final orthomosaic generation.

Hyperspectral pre-processing is described in detail by Laybros et al. (2020, 2019) and summarised here. Two sensors mounted to an aircraft side-by-side were used to cover the full 416-2500 nm wavelength (see Table 4.1). To merge the data from the two hyperspectral sensors without degrading the spatial resolution of the VNIR imagery, we resampled the SWIR imagery to 1 m using nearest-neighbour interpolation. Images were orthorectified and georeferenced at 1 m spatial resolution with the PARGE software using the canopy Digital Surface Model produced from the LiDAR point cloud. Bands in the SWIR with a low signal to noise ratio due to water absorption were removed leaving 378 of the 448 total bands (see Fig. 4.5). Per pixel illumination was calculated using the shadow detection method of Schläpfer et al. (2018). Spectral information used to train and make predictions with the species classifiers was extracted from the overlapping flight lines rather than from a mosaic. This allows for valuable information to be retained as multiple views of individual crowns within the overlapping flight lines which has been shown to improve the classification performance (Laybros et al., 2019). Reflectance spectrum normalization was applied to each pixel. The normalization consisted of dividing the reflectance value of each band by the spectrum of a pixel, by the sum of all reflectance values, which has been shown to improve tree species classification (Dalponte et al., 2014). Some machine learning classifiers are sensitive to the scale in which each feature (band in this case) is supplied with features that have a higher absolute variability tending to dominate. To address this we applied the 'standard' scaling approach which standardises features by removing the mean (centering on zero) and scaling to unit variance (see Fig. 4.5).

Accurate co-registration of data from RGB and hyperspectral imagery was important to ensure spatial alignment. We used the LiDAR-derived Canopy Height Model (CHM) as the baseline layer, with all other data being registered to it. This choice was due to the CHM's stability and precision in representing the physical landscape, providing a solid reference for co-registration. Eight control points were manually assigned across the different datasets, using identifiable features within the LiDAR CHM, such as the flux tower, roads and dominant trees, as primary reference points and an affine transform was applied based on these. This co-registration process ensured that the crowns represented across the datasets corresponded to the same geographical location, serving as the foundation for subsequent analysis steps, including tree crown delineation and species classification.

4.2.3 Field-derived tree crown database

To train and validate our models we generated a set of hand delineated, 'ground truth' crowns with species labels; this database was built and curated between 2015 and 2023 and validated over eight field missions in this period. An initial delineation of tree crowns was performed in QGIS using a combination of RGB, multispectral, hyperspectral, and LiDAR remote sensing data (see Table 4.1). The LiDAR CHM was used as the foundational base layer on which the crowns were drawn as it provides the greatest stability and spatial precision for the outlines. We overlaid the RGB, multispectral and hyperspectral data layers and examined and compared between them to use as much of the spectral, textural and shape information available to us as possible. Where the crowns fell within the inventory plots, an initial guess as to which individual the crown belonged to, based on the location and size of the trunk, was assigned to the polygon. Two provisional confidence scores were assigned the polygons: (1) a 'crown integrity' score describing the certainty with which the outline defines the complete crown of a single individual, (2) a 'trunk match' score describing how confident we were that a crown had been correctly assigned to an individual in the inventory. Where there were changes to the crowns through mortality or branch fall events, the date of change was encoded so that the crowns could be filtered to match the remote sensing data source that they are paired with. Subsequent fieldwork further refined and updated the tree crown delineations. By comparing *in situ* observations with the remote sensing data, we either matched the identified crowns to individual trees present in the site inventory (Gourlet-Fleury et al., 2004), or in cases where crowns were located outside of the known plots, we engaged botanists to assign the appropriate species. We updated the crown outlines and confidence scores based on the field observations and noted where there were liana infestations in the crowns. Additional details can be found in Section 3.2.3.

4.2.4 Automated delineation and fusion of results from repeat surveys

The field-delineated tree crowns were partitioned into a training set and a testing set based on their geographic location (see Fig. 3.2). The regional partitioning ensured clear spatial separation between training and testing datasets, thus negating the inflation of reported accuracy induced by spatial autocorrelative effects providing more reliable and independent assessment of model performance (see Kattenborn et al., 2022).

We used *detectree2*, a tool based on the Mask R-CNN deep learning architecture for automated tree crown delineation (see Chapter 3), which has been shown to outperform another leading method for tree crown detection (Gan et al., 2023). *Detectree2* was trained on the manual crown delineations and corresponding RGB images from the training dataset (see

Fig. 3.2). Researchers seeking to map trees in their landscape of interest may have some ground truth crowns and RGB surveys to train a model on or they may not have training data available. To reflect this, we tested models trained under different regimes:

1. The first was a 'base' model. This was not trained on the UAV RGB imagery and was just exposed to the plane mounted data and crowns from the range of sites described in Chapter 3. This meant it had been exposed to the Paracou forest but with a different sensor, different resolution imagery and four years separation. This pre-existing model is openly available for anyone to use[1].

2. The '1 date' model took the 'base' model and further trained the model on just the first date of the UAV RGB imagery and manual crowns.

3. The '5 date' model took the 'base' model and further trained the model on the first five dates of the UAV RGB imagery and manual crowns.

Comparing the performance of these models provides an idea of what level of accuracy might be expected for researchers aiming to map their landscape with different amounts of training data.

Does the fusing of crown maps across dates improve segmentation accuracy? The trained models were then used to detect and delineate the tree crowns across the entire region of the UAV RGB scans for all 10 images in the range 23-Oct-2020 to 06-Apr-2021. The RGB images were first tiled, predicted on and recombined to generate a set of polygons representing tree crowns from each date (see Chapter 3 for details). Each predicted crown polygon was associated with a confidence score (0-1) indicating the reliability of each tree crown prediction. Where spatial overlap between predictions existed ($IoU \geq 0.2$), the most confident prediction was retained and the less confident predictions removed. The predictions at the individual dates were then combined into 'fused' delineations. This aimed to find an inter-date "consensus" on crown location and shape across dates. The fused sets of crowns went from combining just two dates up to combining the full ten dates. This was done to determine the marginal benefit of adding additional dates of data, each of which comes with an associated cost of the survey.

To combine the polygon sets across dates, the individual date polygon sets were first concatenated into a single set. Each polygon was then compared to every other polygon in the combined predictions set to identify 'significant' matches. A significant match was defined as another polygon with which the intersection area with the current polygon was at least 75% of its union area with the current polygon (i.e. $IoU \geq 0.75$). Accordingly, in the largest combined predictions set (10 dates), a single polygon could have a maximum of nine matches (assuming a

[1] https://github.com/PatBall1/detectree2

strong agreement across dates) and a minimum of zero matches (where there is no confirmation of the polygon across any of the other dates).

Where polygons had been assigned to a matched, inter-date group, the group of polygons was then 'averaged' to produce a single, representative polygon. This averaging was performed by first normalising each matched polygon to have the same number of points on its boundary (300) and then computing the average location of each corresponding point across polygons. The confidence scores of the initial predictions were used as weights in the averaging process so that the most confident predictions had the strongest influence on the resulting shape. If just one significant match was found, the original polygon was kept as was. An outputted polygon would be assigned a 'summed confidence' (i.e. the sum of each individual polygon confidence score) to provide an estimate of the reliability of each prediction, taking into account both the confidence of the original predictions and the degree of agreement between them. It was also assigned a 'combination count' to track how many polygons had been fused to generate it (and on how many dates the tree crown had been delineated). The resulting set of averaged polygons provided a spatial-temporal integration of the tree crown predictions, with each polygon representing the average location and outline of a tree crown across multiple time points. The full algorithm is available to inspect on GitHub[2]. The algorithm parameters, including confidence thresholds, IoU threshold for matching, and number of vertices on a normalised polygon, were tuned on the training crowns prior to testing on the unseen test regions - the optimised values are given above.

To evaluate the performance of the segmentation algorithm, we measured the overlap between predictions and reference crowns. An IoU of an overlapping pair of more than 0.5 was considered a match. This is a standard threshold used in the comparison of tree crown segmentation algorithms (Aubry-Kientz et al., 2021) that allows for small discrepancies in alignment and outline. These "true positives" as well as the unmatched predictions (false positives) and unmatched manual crowns (false negatives) were used to calculate the precision, recall and F_1 score of the predictions.

To determine whether combining tree crown segmentation predictions across dates improves the segmentation accuracy through consensus building, we assessed the F_1-score of each combination of dates, from each single date prediction to the combination of all ten dates. By bootstrapping different date combinations we estimated a mean and standard deviation for the F_1-score for each level of inter-date combination (single date through to ten dates).

[2] https://github.com/PatBall1/detectree2/blob/f996564bfcbaed1ff0ef13a63ea3e62f47252731/detectree2/models/outputs.py#L439C10-L439C10

4.2.5 Tree species classification

Once we had mapped tree crowns as objects in the landscape it was necessary to assign them a species label. The aim was to train a classifier that could, based on the spectral bands of the hyperspectral data, classify tree crowns by species. Within the 3600 label crowns in the database, 178 species had at least two crowns which meant these species could have at least one crown for training and one crown reserved for independent testing. The remaining crowns were not selected for training of the models but their proportional representation in the population was accounted for in the final evaluation of the performance metrics. That allowed for a representative estimate of the landscape level classification accuracy (as compared to performance metrics evaluated within a subset of species in the landscape).

Due to the scarcity of data, particularly for the less well represented species, it was necessary to take a pixel based approach for training and prediction that would then be aggregated for crown level predictions. Taking a pixel level approach was also likely to improve the spatial transferability across a landscape which may have different atmospheric and illumination perturbations. Each pixel was labelled with a class (species) and contained a reflectance value for each hyperspectral band. Pixels with an illumination of less than 60% were discarded. The problem was approached as a supervised learning problem so pixels without a class label were not included in model training.

Whereas the crown delineation required a regional partitioning of crowns between train and test sets to give a robust estimate of performance, for tree species identification, a species-stratified crown level partitioning was more suitable. Due to the diversity and mixing of the forest the average distance between crowns of the same species was large enough to not require additional spatial constraints that would control for spatial auto-correlation.

Which machine learning classifier has the greatest predictive power? Classifier models were trained using the hyperspectral data extracted from the delineated tree crowns. In line with commonly practised methodologies, we evaluated several algorithms such as Multi-Layer Perceptrons (MLPs), Linear Discriminant Analysis (LDA), Random Forest (RF), Linear SVM, k-Nearest Neighbours and Logistic Regression due to their widespread implementations and adaptability. The hyperspectral pixel-data proved challenging to train on, having 378 spectral bands and spanning 178 species, 101 genera, and 41 families. The dataset exhibits significant imbalance; for instance, the most populous tree species boasts 55,448 pixels while the least populated possesses a mere 223. Fig. C.3 offers a histogram of the pixel count by species, highlighting the problem. While classifiers were ultimately evaluated at crown level accuracy via pixel-wise majority voting, the scarcity of crown labels required training at the pixel level. Such an approach, however, encounters pitfalls: adjacent pixels may exhibit local effects that risk leakaging between training and validation sets when naive cross validation is applied. An

uninformed stratified split at the pixel level yields classifiers with high pixel-level accuracy but low crown-level accuracy on unseen data. To counter this, a stratified group k-fold (k=5) cross-validation strategy was adopted. Here, pixels were grouped by crown and stratified by training target, ensuring that any singular crown only appeared in either the test or training set, avoiding the aforementioned issue.

Further complicating matters, the data exhibits significant inter-crown and intra-species noise. Fig. C.4 depicts the spectral value distribution for the two most expansive tree crowns of species *Pradosia cochlearia*, illustrating noteworthy variance. Meanwhile, Fig. C.5 shows the standard deviation across each spectral band of the top ten species (by pixel frequency) in the training dataset, alongside the standard deviation spanning all pixels from those species. Note that these spectra underwent standardized scaling, ensuring unit variance across all pixels.

Models were trained using a group k-fold (k=5) cross validation on training pixels, maintaining the invariant that pixels from the same crown could appear in either the training or test fold but not both. Models were also trained on a simple k-fold (k=5) cross-validation, to test performance of models trained using a naive pixel-wise approach (i.e. would fitting to local features impact the transferability between crowns). After cross-validation, models were trained on the full training set and used to predict a species class across the entire landscape. Crown-level classifications were determined by majority vote of pixels comprising the crown[3]. Evaluations were conducted using a held-out test set with crown labels, utilizing the (micro-average) F_1-score for model performance comparisons. The test set contained a random sample of 20% of crowns for each species except where a species had just four individuals or less, in which case a single test crown was randomly selected. Additionally, precision, recall, and the macro-average F_1-score were recorded, the latter assisting in gauging performance for less prevalent classes.

How does classification accuracy depend on number of training crowns? To understand how many individual training crowns from a species were required to achieve a good classification accuracy we modelled the relationship between F_1-score and training crown number with a beta regression. As scores could take 0 or 1 values, this was done as a Bayesian zero-and-one-inflated beta regression with the *zoib* R package (Liu and Kong, 2015; Liu and Li, 2016). We assigned a threshold of F_1-score ≥ 0.7 for a 'good' performance. At this value we could be confident that we correctly identify the majority of the individuals of a species present without assigning trees of a different species a label of that species. This would allow us to expand our sample in a way that could strengthen the phenological signal without masking

[3] A more sophisticated approach averaging pixel wise class predictions was tested but this gave almost identical results to the simpler majority-vote approach

it with noise (discussed in Chapter 5). The expected minimum number of training crowns required to achieve this performance was recorded.

Which wavebands are the most important for determining species? To determine which wavebands were most important for distinguishing species, we selected the best predictive model (based on test set F_1-score) and extracted feature importance in a method suitable for that model type. In the case of LDA, this was done by extracting the coefficients for each waveband across all classes. Each coefficient gives the importance of that feature (in this case, wavebands) for discriminating between classes. A larger absolute value of the coefficient means that the feature is more significant in differentiating between classes. A coefficient near zero suggests that the feature does not contribute much to the separation of the classes. The average absolute coefficient value at each band across classes gave an overall importance value for classification, aggregating how important each feature is for separating all classes in the dataset.

Do spectra observed within tree crowns exhibit phylogenetic signal? To test for phylogenetic signal in the hyperspectral data, we calculated the mean and standard error of the standardised reflectance for the pixels belonging to each species, and using the time calibrated phylogeny of Baraloto et al. (2012) and the *phytools* R package (Revell, 2012), calculated Pagel's λ (Pagel, 1999), and Blomberg's K and significance test (Blomberg et al., 2003) on each of the bands of the hyperspectral data included in the species classification. Pagel's λ and Blomberg's K are statistical measures that quantify the degree of phylogenetic signal in continuous trait data. Pagel's λ measures the extent to which trait evolution follows a Brownian motion model, scaling from no phylogenetic signal (0) to complete Brownian motion (1). Blomberg's K quantifies how the observed trait variance among species deviates from what would be expected under a Brownian motion model. A K value greater than 1 suggests that closely related species are more similar than would be expected under a Brownian motion model of trait evolution, a K value less than 1 suggests less similarity among related species than expected under a Brownian motion model of trait evolution.

Are the most important wavebands for species classification those that have the strongest phylogenetic signal? We tested the correlation (Spearman's rank ρ) between the strength of the phylogenetic signal (λ and K) of each spectral band to the feature importance values of each band. This was to test whether those bands that were helpful in discriminating species also contained a stronger phylogenetic signal.

Are closely related species more often confused in their classification than distantly related species? To test whether more closely related species were more likely to be confused with one another than distantly related species we tested the correlation between pairwise phylogenetic distance (as Myr of independent evolution between pairs of plant taxa as provided

in the time calibrated phylogeny of Baraloto et al., 2012) and pairwise mis-classification rate. This was done with the rank coefficient as the distributions were highly non-normal.

4.3 Results

4.3.1 Diversity of Paracou's canopy

We labelled crowns of 3,539 individuals to species-level. The dataset comprised 250 unique species, 131 genera and 45 families. We removed 72 of these species from subsequent analyses because they were only encountered once in our survey, so could not be included in the classifier training and testing (see Fig. 4.2). The 72 species represented just 2% of the crowns we recorded. The filtering process left 178 species for further analysis.

4.3.2 Tree crown segmentation

Does the fusing of crown maps improve segmentation accuracy? The accuracy of the tree crown segmentation was improved by combining multiple dates of tree crown segmentation predictions and retaining crowns that had good confidence and agreement between dates (see Fig. 4.3). We compared the accuracy of segmentation from using a single time step and multiple time steps with an unseen test set of 169 test crowns across two spatially separate test zones. The accuracy of delineations increased as more dates were combined. The best performing model overall was the one trained on five date's worth of imagery. The accuracy of its delineation was boosted significantly by combining information from different time steps from mean F_1-score of 0.68 for a single date prediction, to a peak at the combination of nine time steps with a mean F_1-score of 0.78. In terms of total crown area, approximately 86% of the test region had well located and delineated crowns (see Fig. 4.4). Accuracy tended to increase with tree crown area.

The 'base' model and the model trained on additional single date of UAV data had a comparable performance at a single date prediction and combinations of less than seven dates (see Fig. 4.3). However, after combining seven dates, the combined delineations of the 'base' model became substantially better, surpassing the single time step prediction accuracy of the model trained on five dates. The model trained on a single time step failed to improve in accuracy by combining dates to the same degree as the other two models, suggesting it became overfitted to the limited available data. The 'base' model was trained on a range of non-UAV RGB imagery (see Chapter 3) without the additional focused training on UAV data; this seemingly led to better temporal transferability than was achieved with the additional focused training on just a single date.

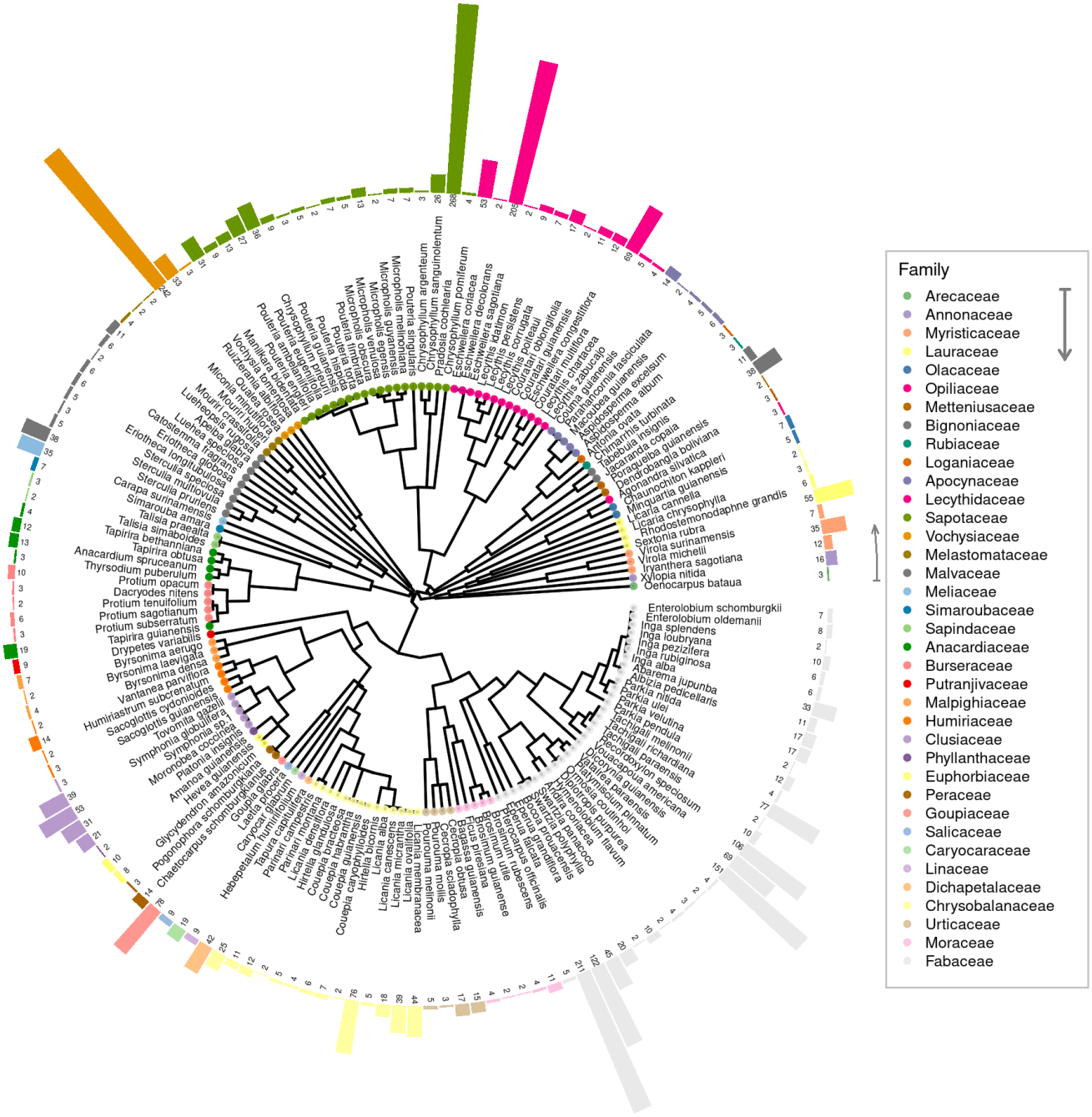


Fig. 4.2 Abundance of species recorded in the field mapped onto the phylogeny of Baraloto et al. (2012). Bars represent the total number of individuals sampled in the field. The legend is arranged in the order families appear on the tree (anti-clockwise from 3 o'clock).

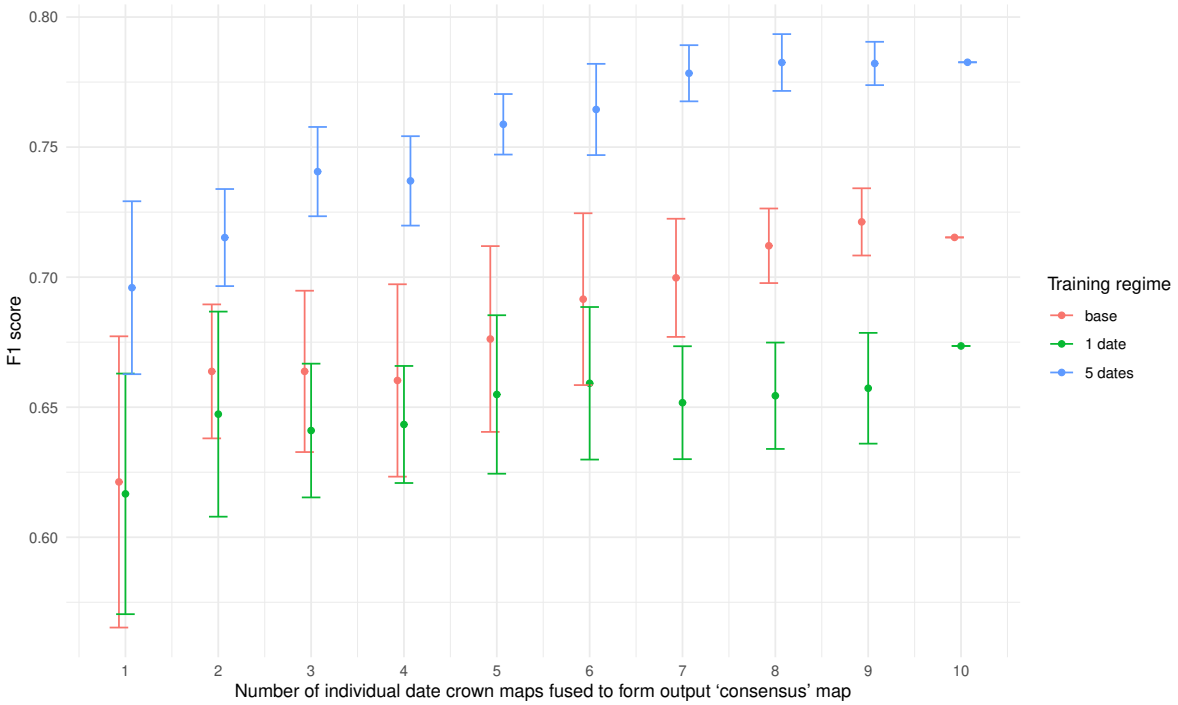


Fig. 4.3 The performance of the tree crown delineations with the number of individual date crown maps that were combined to form the output 'consensus' map. Researchers may or may not have training crowns available and a varying number of RGB surveys. To reflect this we tested three models trained under different data regimes: (1) the 'base' which was freely available online (trained on different crowns and imagery) (2) the '1 date model' - the base model then trained on manual crowns with a single date of RGB imagery; (3) the '5 date model' - the base model then trained on manual crowns with five UAV-RGB surveys of the same location. The mean and standard deviation of the F_1-score was calculated with bootstrapping. Note that the 10 date combination could not be bootstrapped as there was only a single combination of dates possible.

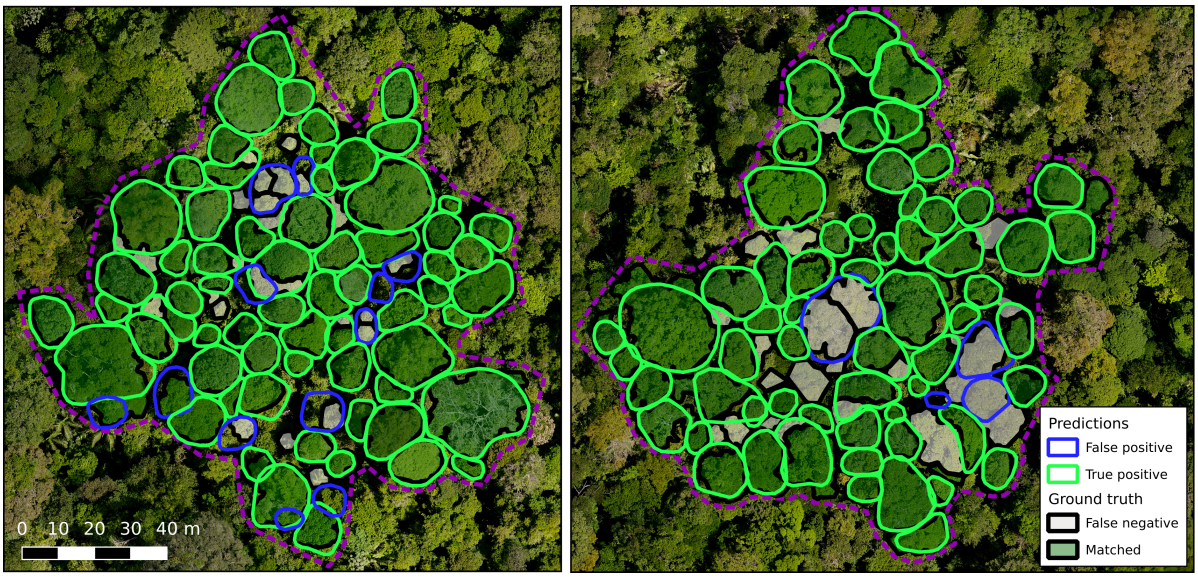


Fig. 4.4 Predictions and ground truth crowns in the unseen test regions. The reference set of crowns are shown with black border and their fill colour depends on whether a crown was matched (green) or unmatched (light grey) with a prediction crown. The predictions that match with a reference crown have a green border and those that did not match a reference crown have a blue border. A match was granted in the case that a reference crown and a prediction crown had $IoU \geq 0.5$.

4.3.3 Species classification

Which type of machine learning classifier can most accurately predict the species of tree crowns from hyperspectral data? The LDA classifier performed best at classifying the species of test set of tree crowns from their hyperspectral signal (micro average F_1- score = 0.76; see Table 4.2). Of the 178 species included, 64 of the species (36%) were classified with an F_1-score of over 0.7. The more flexible MLP and SVM classifier failed to match the performance of the LDA classifier, highlighting that the LDA's approach to separating the classes led to more robust transferability between crowns. Furthermore, the LDA classifier took 20 seconds to train whereas the MLP and SVM took several hours.

As a percentage of total crown area of the test set, about 81% was assigned with the correct species. By combining the percentage of the total crown area that was well located and delineated (86%) and the percentage of the total crown area that was assigned the correct species (81%), we estimate conservatively (given that incorrect areas of each step are more likely to coincide than not) that 70% of the landscape's crown area was mapped correctly. For reference, the crowns of the top 20 most abundant species make up less than 60% of the total crown area of the reference dataset.

How many mapped individuals of a given species are needed to achieve a 'good' classification accuracy? We distinguished 64 species with an F_1-score of at least 0.7, which is

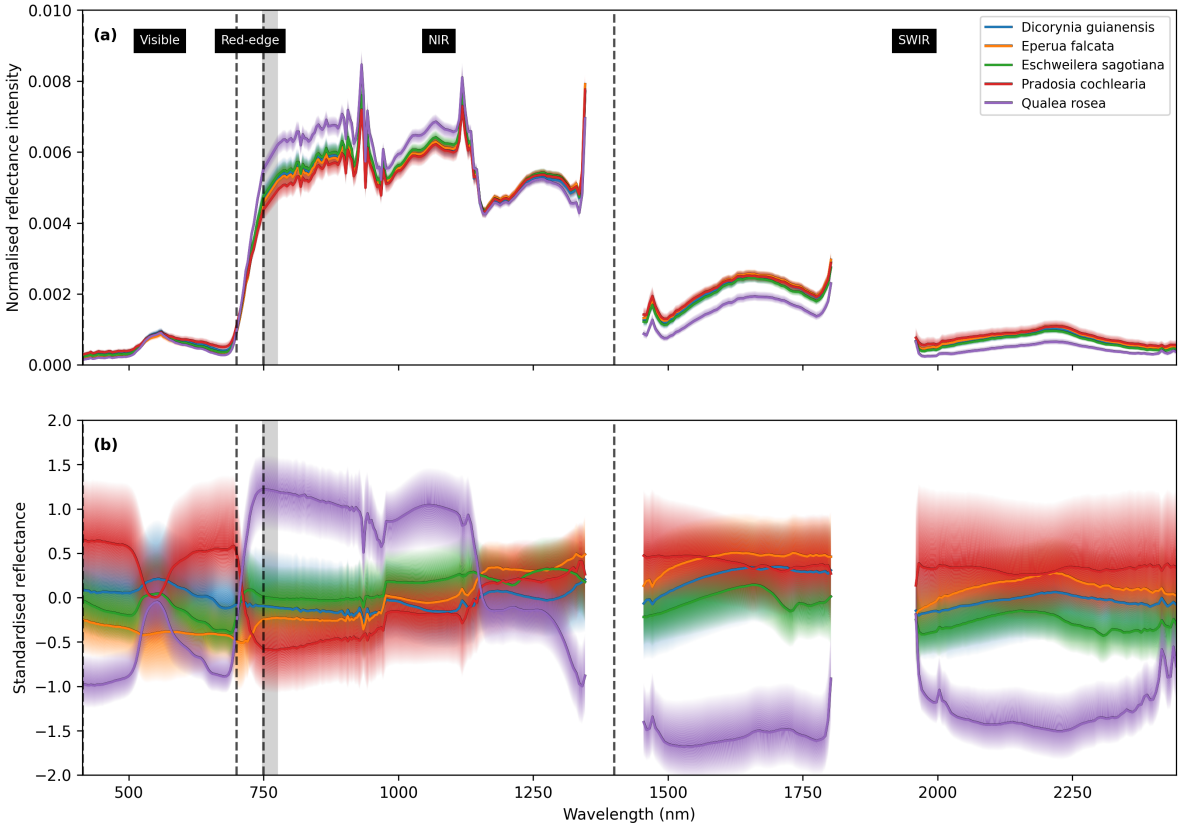


Fig. 4.5 The observed spectrum of five common species which together comprise 31% of the total crowns in this study. The median line of pixels values is plotted and the IQR is shaded to show spread. (a) Shows the reflectance spectrum where pixels have been normalised by dividing the reflectance intensity by the summed reflectance over all bands. (b) The mean value is subtracted from the standardised reflectance, and the resulting value is divided by the standard deviation for each band across valid tree crown pixels, so that all bands are shown on the same scale. The plots illustrate how species might be identified from spectral information (*spectral signatures*). Shadowed pixels were removed prior to standardisation. The two gaps in the SWIR region are the result of removing bands influenced by air humidity. A grey band at 748 nm to 775 nm shows the spectral region in which bands were most important for classification.

Table 4.2 The accuracy statistics for the classification models based on the unseen test set of crowns (across 178 species).

Classifier	Precision	Recall	F_1-score (micro)	F_1-score (macro)
LDA	0.74	0.75	0.75	0.52
Logistic	0.72	0.74	0.73	0.41
LinearSVM	0.67	0.70	0.70	0.35
MLP	0.64	0.70	0.70	0.34
QDA	0.55	0.58	0.58	0.26
RandomForest	0.44	0.42	0.42	0.13
kNN	0.33	0.37	0.37	0.08

a more species than previous studies that do not cover more than 20 species (Laybros et al., 2020, 2019). The accuracy of classification increased with the number of training crowns in the class (see Fig. 4.6) in agreement with similar patterns observed by Baldeck and Asner (2014) and Féret and Asner (2013). There was a sharp increase between 1 and 10 training crowns, after which the performance improved more gradually. Species with at least 10 training crowns could reasonably be expected to be classified with an F_1-score of 0.7.

Which wavebands are important for determining species? Eight bands between 748 and 775 nm, right on the edge of the "red edge" transition between the red and near-infrared ranges, dominated in terms of relative feature importance for separating species (see Fig. 4.8). The next most important region was 640-660 nm which fell in the red part of the visible spectrum. The 560-575 nm (green), 1630-1680 nm (SWIR) and 1000-1100 nm (NIR) regions were also relatively important for discriminating between the species.

Do spectra observed within tree crowns exhibit phylogenetic signal? Most standardised and normalised bands showed a statistically significant phylogenetic signal ($p \leq 0.05$; see Fig. 4.8). Out of a total of 378, Pagel's λ test gave 310 significant bands while Blomberg's K test gave 239 (whereas one would expect 19 significant bands by random chance).

Are the most important wavebands for species classification those that have the strongest phylogenetic signal? From inspecting Fig. 4.8, there was no obvious relationship between feature importance and phylogenetic signal of the bands. However, both λ ($\rho = -0.14$, $p = 0.005$) and K ($\rho = -0.16$, $p = 0.001$) had a statistically significant negative correlation with feature importance suggesting that the more phylogenetic signal in a band the less helpful it is to discriminating between species.

Are closely related species more often confused in their classification than distantly related species? By comparing phylogenetic distances between pairs of species with their

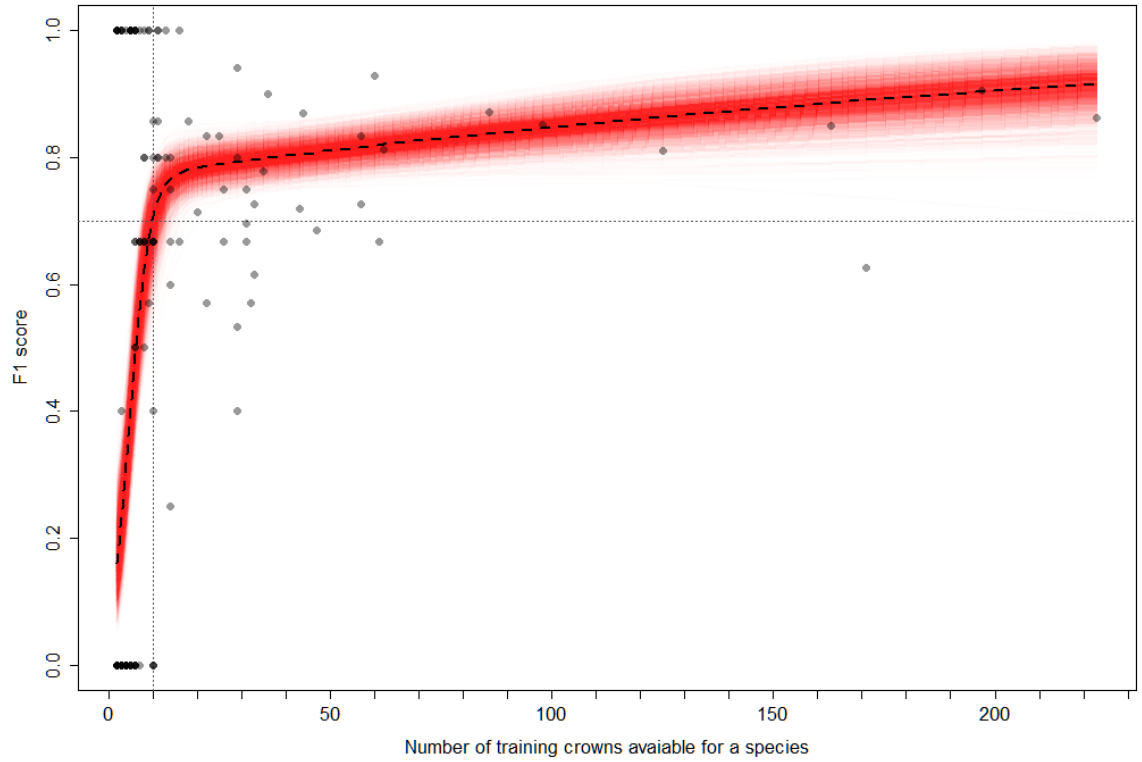


Fig. 4.6 Classifier performance for individual classes in relation to the number of training crowns in the class. Each dot represents the F_1-score for the classification of an individual species. A zero-and-one inflated beta regression was performed to find the expected F_1-score by number of classes (the black dashed line). Dots are slightly transparent so where they appear darker there are several overlapping species with that score. This is common with low numbers of training crowns as there are fewer discrete scores that can be attained. The dotted line at x=10 shows the minimum number of training crowns required before an expected classification performance of F_1-score = 0.7 is acquired for a species.

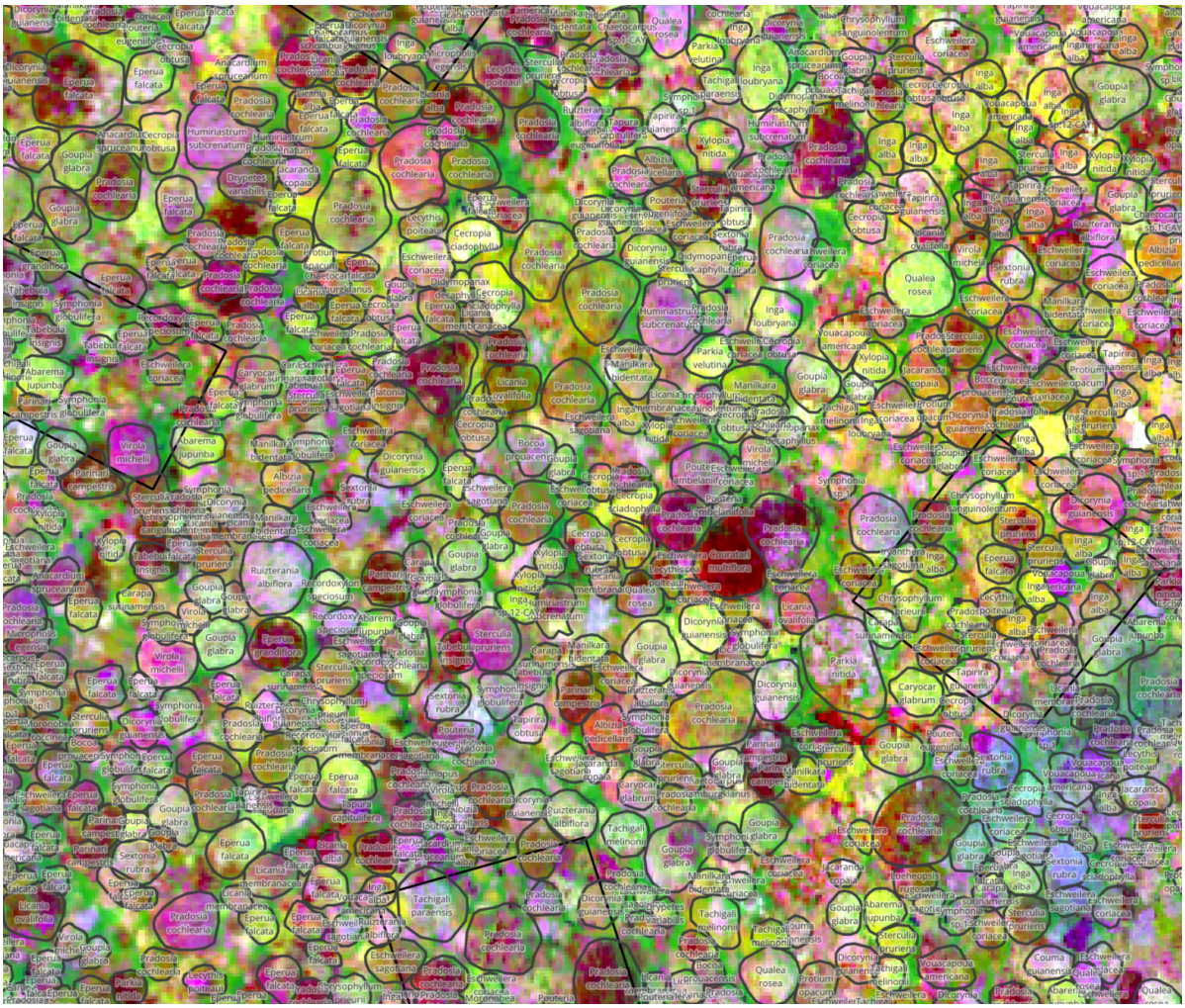


Fig. 4.7 A portion of the finalised crown map overlaid on three selected bands of the PCA projection of the HSI. The PCA projection of the HSI is purely a means to visualise the hyperspectral data and was not used at any stage of the analysis. Crowns identified are outlined in black and labelled with predicted species. The black squares are forest plots.

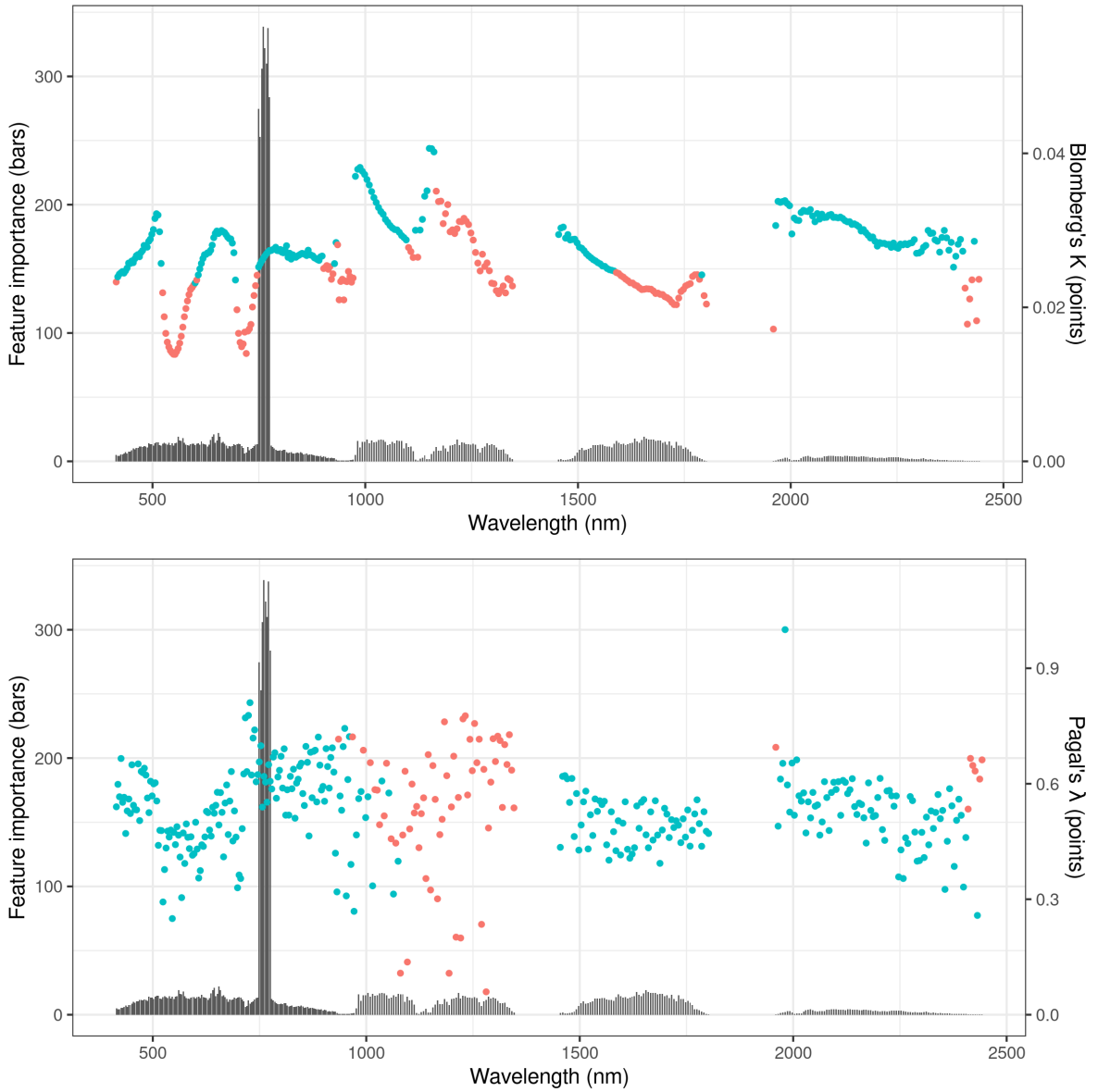


Fig. 4.8 The relative feature importance (bars) and phylogenetic signal (points) of each band used in the classifications. The blue circles indicate that the phylogenetic signal is statistically significant (at $p \leq 0.05$) while orange circles show values that are not statistically significant.

pair-wise mis-classification rates, we determined that the more closely related the species the more likely they were to be confused with one another in the classification ($\rho = -0.0437$, $p = 0.0008$). The confusion matrix of the species is given in Fig. 4.9. Note where confusion between species occurs within genera.

4.4 Discussion

Mapping tropical rainforest canopy species is essential for gaining a granular understanding of large-scale ecological processes. We have extended the scope and reliability of species identification from aerial hyperspectral data in diverse tropical forests, surpassing prior levels of identification accuracy for a far greater variety of species. Previous studies have managed to map around 20 species with accuracy (Féret and Asner, 2013; Garzon-Lopez and Lasso, 2020; Laybros et al., 2019) with greater scope in diverse tropical forest species classification only coming from direct leaf spectroscopy (Harrison et al., 2018; Prospere et al., 2014). Since tropical forests typically contain several hundred species per hectare (Duque et al., 2017; Lee et al., 2002), of which around 30%-60% make it to the canopy (Bohlman, 2015), many of the crowns have been left unidentified. By extending this number to aim to map comprehensively (178 species with 64 species $F_1 > 0.7$) we get much closer to complete landscape mapping of upper canopy trees. In our study, the top 20 species covered less than 60% of the total crown area. Assuming perfect segmentation and classification accuracy, this would be the upper limit of accuracy (by area) that could be achieved at a landscape scale if only those 20 species were included in the mapping process. In comparison, by expanding the pool of species, we were able to accurately map over 70% of the total crown area of the landscape. This more complete coverage is not the result of more sophisticated species classification algorithms, as we found that the well-established LDA approach transferred better between crowns than more flexible methods (SVM, MLP; in agreement with Féret and Asner, 2013). Instead, improvements were achieved by (1) creating a large reference database of labelled geo-located crowns; (2) improving training/testing datasets by careful mapping and identification of trees in the field; (3) high-quality hyperspectral imagery that accurately co-aligned with tree crown maps, allowing species to be distinguished from upper red edge bands; (4) improved methods for segmenting tree crowns, which is vital for mapping across landscapes. We discuss these improvements, the fundamental basis for species identification based on knowledge of the evolution of plant physiology and morphology, then consider the obstacles to achieving affordable, easy-to-use and transferable approaches to tree species identification.

Abundant high quality field data. The field dataset was developed, curated and ground validated over a number of years. The careful mapping of what was observed from above

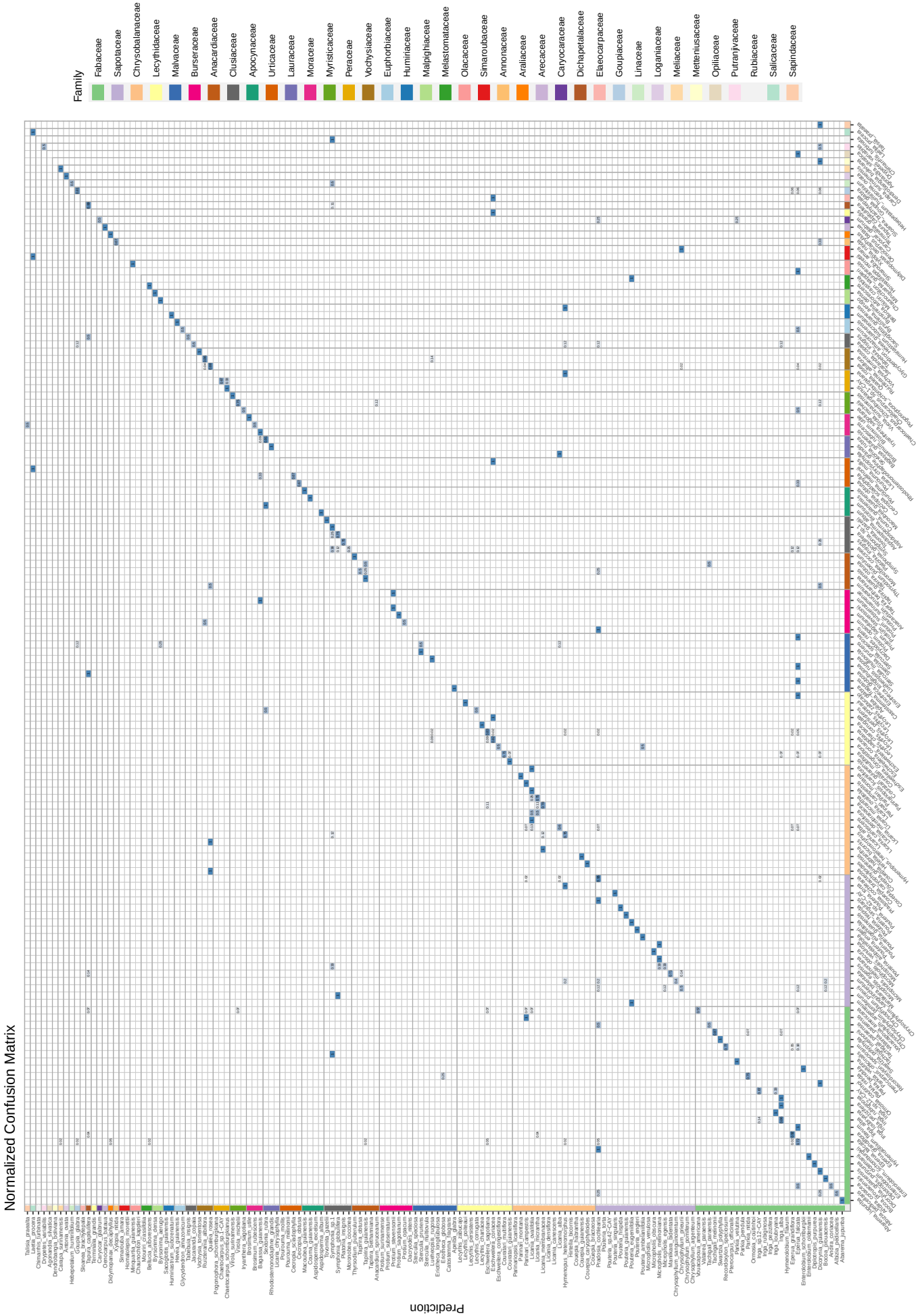


Fig. 4.9 Normalised confusion matrix for predictions on the test set crowns. The families are ordered by the number of species included in the study that they contain. Deeper blue values show higher proportions (closer to 1) and the faint colours show proportions closer to 0.

to the reality on the ground was labour intensive but the resulting database of 3600 crowns provided a robust basis upon which this study could be built. Without substantial, high quality ground datasets, remote sensing is limited in the inferences is it able to make (Chave et al., 2019; Davies et al., 2021). Despite this, there are few studies that provide benchmark data upon which tropical tree species mapping approaches can be tested (Laliberté et al., 2020).

The importance of high-quality hyperspectral imagery. Aerial hyperspectral data captures spectral intensities across hundreds of contiguous, narrow wavelengths, allowing signals of biochemical and morphological properties in foliage, such as chlorophyll, water content and leaf structure to be observed (Clark and Roberts, 2012). This depth of information is indispensable for the remote identification of species, particularly in diverse tropical forests where conventional, broader band multispectral imagery falls short (Zhang et al., 2006). The hyperspectral images used in this study were collected close to nadir, ensuring minimal distortion and noise. Additionally, these images were co-aligned with ground-truthed tree crown maps, thereby significantly enhancing the reliability and precision of species identification. Through this approach we identified the narrow upper red edge band range of 748 to 775 nm as by far the most important region for discriminating species (Fig. 4.8).

Improved segmentation of tree crowns. The emergence of CNN methods has allowed for considerable progress in automatic processing of images across a range of fields. By integrating geospatial features with the sophisticated Mask R-CNN architecture (He et al., 2017), the *detectree2* Python package (Ball et al., 2023) can harness subtle spectral and textural clues to delineate trees precisely achieving state-of-the-art performance on tree detection in aerial RGB data (Gan et al., 2023). Even humans struggle to agree where trees are located when looking at the same data layers (Section C.1). By allowing a machine to learn on a carefully validated manual (field verified) dataset, it has synthesised past human attempts to produce quicker and more consistent predictions. Accuracy of segmentation can be significantly increased by combining maps of segmented tree crowns over time. However, forest canopies can appear very different across dates due to atmospheric perturbations, differences in illumination (resulting in variably shadowed trees), the sway of trees and branches, phenological changes, death of individual trees and branches and irregularities in the orthomosaicking process. UAV RGB sensors are cheap and provide a source of repeated, high resolution scans over the course of a few months (during which there are few mortality events and little growth). We found that, by combining predictions across dates, a model that had never been exposed to the specification of imagery that it was predicting on (in this case UAV RGB) in training could reach a comparable accuracy to a model trained on a high volume of the specific imagery (Fig. 4.3). There is no substitute for gathering high quality training data but, if this is not available, repeat predictions with a pre-trained, freely available model can give excellent tree crown delineations on a new

site. This can support the establishment of new studies to track tree growth, mortality and phenology over large areas. This was the first time that information across dates has been combined to improve the accuracy of tree crown delineations.

Towards a fundamental understanding of species differentiation. We found upper red edge bands in the 748 to 775 nm range were by far the most important bands for discriminating species (Fig. 4.5), a finding at odds to Laybros et al. (2019) but aligned with Badourdine et al. (2023)'s assessment of band importance for assessing canopy taxonomic diversity at the same site. This pattern has not been picked up explicitly in other previously published studies but is not precluded by them as importance has generally been assessed over broader regions of the spectrum (with red edge to NIR generally being considered important) (Clark et al., 2005; Dalponte et al., 2012; Fassnacht et al., 2016; Hennessy et al., 2020; Marconi et al., 2022). Leaf traits are incorporated into spectra in complex ways (Féret and Asner, 2011; Jacquemoud and Baret, 1990) and multiple traits can superimpose in a given spectral region (Curran, 1989). The "red edge", defined loosely as the 700-750 nm region, is widely recognised as an important region for classifying vegetation and is linked to chlorophyll content, leaf area, water content and overall plant health (Boochs et al., 1990; Filella and Penuelas, 1994; Gitelson et al., 2003; Hennessy et al., 2020; Horler et al., 1983; Thomas and Gausman, 1977). However, the red edge is usually defined as the 700-750 nm range. Here we found that the region immediately beyond the red edge (748-775 nm) was particularly sensitive to differences among species in a tropical forest. The "upper red edge" is at the transition zone between chlorophyll absorption (in the red) and cellular structure scattering (in the NIR), potentially capturing information from both the biochemical and structural (leaf and canopy) aspects of the vegetation. The wavelengths adjacent to this range are not showing strong feature importance, which means the unique reflectance in the 748-775 nm range could be capturing some species-specific anatomical features, possibly related to internal leaf structure affecting scattering of near-infrared light (Ustin et al., 2009).

Other wavelengths were also of secondary importance. In the visible range, red wavelengths of 640-660 nm were relatively important: Chlorophyll a and b have peak absorbance at different wavelengths of red light (660-680 nm vs 640-660 nm respectively), suggesting that variation in Chlorophyll-b may be important for discriminating species (Gitelson et al., 2003). Bands within the green region of 560-575 nm showed some importance which could be due to chlorophyll reflectance differences (Gitelson et al., 2003) but could equally be due to leaf structure, carotenoids and anthocyanins (Sims and Gamon, 2002), or even stress (Carter and Knapp, 2001). Water has strong absorption features in the SWIR region. The relative importance of bands in the 1630-1700 nm region suggests detection of variation in

leaf water content (Ceccato et al., 2001; Gao, 1996) and/or differences in cellulose and lignin composition (Kokaly et al., 2009; Serrano et al., 2002).

We found "crown reflectance spectra" (i.e. spectra influenced by leaf reflectance spectra plus influences of absorption, reflectance and transmission by leaves in a multi-layered tree crown) showed phylogenetic structure. Madritch et al. (2014) showed the capacity for aerial hyperspectral imagery to characterise genotypic identity while Schweiger et al. (2021) showed a correlation between leaf level spectral dissimilarity with phylogenetic distance (Schweiger et al., 2021). Other studies have shown phylogenetic structure of foliar spectral traits (Cavender-Bares et al., 2016; Meireles et al., 2020) including for leaves in tropical forest canopies (McManus et al., 2016). For the first time, we have linked crown reflectance spectra to the phylogenetic signal to help explain species classification from aerial hyperspectral data.

Most bands exhibited some phylogenetic signal but the importance for classification was negatively correlated to the signal. This could be for a number of reasons including spectral overlap between closely related species, the importance of ecological, environmental and stress factors, and convergent adaptive traits that occur broadly across the phylogeny. Looking within lineages instead of across the whole phylogeny may be a way to probe this relationship further (Meireles et al., 2020).

Traits may vary in their degree of phylogenetic conservation depending on a variety of factors including environmental pressures, mutation rates, and the particular evolutionary history of the species in question. Of those traits discussed above, some might be expected to be more conserved than others. Polymers including lignin and cellulose are critical for plant structure, and their relative concentrations are generally highly conserved within lineages (Weng and Chapple, 2010). Additionally, leaf structure, including traits such as leaf thickness or specific leaf area, is conserved to some extent within phylogenetic lineages (Ackerly and Donoghue, 1998). Other traits may be more labile and lack phylogenetic conservatism. For instance, while the ability to retain water might be conserved within specific lineages adapted to particular environments, there can be significant variability in this trait both within and between species based on immediate environmental conditions (Donovan et al., 2011). Other traits are highly plastic and have weak phylogenetic signals. Although chlorophyll is essential for photosynthesis in all plants, the specific amount and ratio between Chla and Chlb concentration can vary greatly even within a single species based on a variety of factors, including age, health, and immediate environmental conditions like light and nutrient availability (Gitelson et al., 2003). Stress responses are highly variable traits that can differ significantly even within a species based on environmental pressures and are likely to be among the least conserved traits phylogenetically. Furthermore, the classifier may be getting clues from soil and other external factors that the plant interacts with, rather than intrinsic traits of the plant. With this is mind,

it may be that the classifier is basing its decisions more on how species are responding to environmental conditions rather than on their intrinsic biophysical properties. Analyses of functional traits variation (e.g. Asner, 2014; Asner et al., 2014; Schmitt et al., 2022) may hold the key to understanding the extent to which hyperspectral sensing is detecting interspecific vs intraspecific variation in biophysical traits (e.g. see Nunes et al., 2017).

Our species-focused approach differs from biodiversity metric approach adopted by many studies, which focuses on mapping taxonomic diversity using spatial variance in the hyperspectral signal (Jucker et al., 2018; Kamoske et al., 2022; Laliberté et al., 2020; Vaglio Laurin et al., 2014). Mapping diversity in this way is challenging because of uneven spectral distances among species (e.g. because related species have more similar spectra), and the large variance in spectral properties of single species when compared across landscapes (Badourdine et al., 2023).

Future work

Improvements in classification. Accurate classification of species from hyperspectral data required learning the spectral features exhibited across crowns of the same species. To close the gap between realised and intrinsic predictability of species classification we need to understand the sources of error. Simple mis-classifications due to the misalignment of data sources or mistakes in the field labelling, are trivial to address. External sources of signal variation (e.g. atmospheric disturbance, sun-sensor geometry) may be addressed with improved data processing based on physical models. Variation of crown spectra within species driven by biological factors such as water stress, or phenology may be addressed with more informed feature selection/engineering and more sophisticated classification approaches.

The more flexible models, including MLP and SVM were better able to fit the local features of the pixels within crowns but this failed to transfer to performance across crowns where the simpler LDA did best. This highlights the challenge in applying cutting-edge machine learning methods to this task. More work needs to be done on understanding how best to constrain the more advanced methods so they can learn to encode the key crown-level features from limited training data.

Addressing transferability. The ultimate aim is to train classifiers that can accurately predict the species of tree crowns when transferred in time (e.g. different seasons) and space (e.g. new forest locations). While we were able to identify a wide range of species at a single date and location from hyperspectral data we have not demonstrated that it is possible to transfer this to other locales. It is known that classification degrades with time between training and prediction scans (Laybros et al., 2019) and if applied to new regions.

Trees exhibit temporal variability in their spectral signatures due to seasonal phenological changes (Chen et al., 2022b; Hesketh and Sánchez-Azofeifa, 2012) and external stressors like pests or drought, complicating year-to-year or season-to-season species identification. Even at a fixed point in time, individuals of the same species can be at different phenological stages (see Chapter 5). Spatially, even trees of the same species can have different spectral signatures based on geographical factors like soil type, local climate and topographic position (see e.g. Fig. C.4). Atmospheric conditions, from clouds, gases constitution, to airborne particulates, further modify the spectral data acquired by airborne sensors (Arroyo-Mora et al., 2021; Schläpfer et al., 2018; Theiler et al., 2019). This problem is exacerbated by variations in atmospheric conditions between airborne sensor and the Earth's surface, which can modify the spectral signatures, making hyperspectral data from different times and locations difficult to compare directly (Theiler et al., 2019). Additionally, spectral responses can differ between sensors, and even the same sensor can vary over time due to calibrations or degradations (Baumgartner et al., 2012), complicating data comparison. Finally, the spectral data acquired can be influenced by the illumination (Arroyo-Mora et al., 2021; Schläpfer et al., 2018; Theiler et al., 2019) and viewing geometry (Duthoit et al., 2008; Lyapustin et al., 2012; Montes and Ureña, 2012; Schläpfer and Richter, 2014; Schläpfer et al., 2015; Theiler et al., 2019), including the angle of sunlight and sensor viewing angle, introducing additional variability across space and time and necessitating more sophisticated data handling and analysis approaches.

More work needs to be done on collecting hyperspectral data (and labelled tree crowns) at different dates with a broader range of locations but also across a wider range of acquisition conditions. With this it may be possible to identify features/encodings that remain stable through time and space so that models may be flexible enough to be applied broadly across an ecosystem. Most features could have some innate biochemical explanation, or have an explanation more indicative of different levels of stress among different species. Indeed, there are likely many interacting effects between bands that we have analysed here. However, it is also possible that machine learning models applied to real data may find importance in specific spectral ranges due to noise or collinearity in the data that may not have an easy biological explanation. Without further analyses, explanations are speculative. Confirming the reason for the peaks in feature importance would likely require controlled studies involving leaf-level spectroscopy, coupled with biochemical assays to identify the specific compounds or structure responsible for these spectral features.

Conclusions. Three key ingredients were responsible for the accuracy, and comprehensiveness of the mapping: high-quality field data; a precise segmentation algorithm; repeat imagery. Being able to accurately locate the crowns of trees and assign a species to them allows for much better forest monitoring at a granular level. We can study biodiversity by tracking the

compositional changes with climate change. With the advent of freely available, high spectral (e.g EnMAP) and spatial resolution (e.g. PlanetLabs) satellite imagery, the approaches here could be extended for mapping species at a pan-tropical scale.

Chapter 5

Characterising leaf phenology of tropical forest trees with repeated drone multispectral and LiDAR surveys

Abstract

Leaf phenology has a strong influence on carbon and water cycling but there remain large uncertainties around when tropical tree species cycle their leaves and what environmental and physiological processes govern the patterns. The development of drone-mounted sensors has vastly improved the spatiotemporal resolution with which biodiverse tropical forest can be monitored. Here we track the leafing phenology of 3000 tree crowns in tropical moist forests of French Guiana by conducting lidar and photographic surveys with UAVs every three weeks over the course of two and a half years. These data allowed us to track changes in Plant Area Density (PAD) and greenness (as Green Leaf Index; GLI) of individual tree crowns over time, and understand how leaf quantity and quality varied through time, particularly in relation to the short dry seasons prevalent in the region. To do this we calculated the periodicity, synchronicity and regularity of phenological signals, and seasonality and timing of leaf flushing and shedding from the GLI patterns for 100 dominant species. Species showed distinct leaf phenology patterns, with considerable variation among species which we hypothesised related to resource acquisition strategy and local environmental heterogeneity. The majority of trees showed a leaf cycle period close to one year with some shorter-term cycles and some cycles close to two years. Half of the species showed a statistically significant degree of "seasonality" for flushing and shedding of leaves. Most tree species flush new leaves around the transition from wet to dry season. Interspecific variation was strongly associated with local topographic position (Topographic Position Index; TPI) preference and tree height. In general, shorter species showed little variation in phenological character but the patterns of taller species diverged based on their local topographic habitat preference (gullies through to ridges). Taller species flushed fresh leaves earlier with respect to the dry season, an effect that was strongest for species associated with gully location (low TPI) but not apparent for species associated with ridges (high TPI). This suggested that emergent species that were unlikely to experience water stress were better able to maximise light interception by presenting efficient leaves throughout the dry season, compared to emergent trees in drier conditions that were more constrained by the risk of hydraulic failure. The strength and synchronicity of the leaf cycling signals were greater for ridge species compared to species that were associated with gullies, suggesting species associated with better-drained land had greater sensitivity to environmental cues. Taller species

tended to have shorter leaf cycle periods, potentially linked to a faster resource acquisition strategy required to emerge beyond the canopy. Intraspecific variation in timing was associated with tree height; the taller the tree, the earlier it was likely to flush. Intraspecific variation in leaf cycle period was greatest at high TPI, potentially linked to a facultative exchange strategy. There was evidence that the amplitude of greenness variation across seasons was phylogenetically constrained, but none of the other indices were closely linked to evolutionary history suggesting leaf phenological traits are labile for opportunistic evolutionary niche filling. Incorporating additional leaf and wood traits would allow for more precise categorisation of leaf phenology and clearer links to resource acquisition strategy. Linking tree-level variation in leaf amount and quality to the observed fluxes (e.g. via the eddy covariance method) will shed light on how seasonal changes of individual tree crowns combine to influence seasonal patterns of forest productivity.

5.1 Introduction

Leaf phenology — the timing of leaf flushing, senescence and abscission — has important implications for the seasonality of ecosystem processes such as primary productivity and litter decomposition (Lieth, 1974). The timing of leaf production and loss determines the temporal dynamics of the leaf area index (LAI, total leaf area per unit ground area), which is a key driver of gross and net primary productivity (Chen et al., 2012; Manoli et al., 2018) and a key determinant of evapotranspiration rates, which in turn influence the availability of soil water resources, local hydrology and regional water cycles (Teuling et al., 2010). Leaf phenology also determines the age structure of leaves within a canopy, which influences photosynthetic rates of leaves (Wu et al., 2016, 2017a,b). Furthermore, leaf phenology affects multi-trophic species interactions. For instance, leaf emergence impacts herbivore populations which rely on the availability of young leaves and fruit (van Schaik et al., 1993), which can have cascading effects on predator populations (Visser and Both, 2005). Synchronicity of leaf emergence among trees of a given species may help to suppress specialist herbivores (Aide, 1993). The timing of leaf senescence and abscission also impacts microbial processes in soil, as fallen leaves provide resources for below-ground communities (Chave et al., 2010; Estiarte and Peñuelas, 2015). Tropical forests have a major influence on global carbon and water cycles (Bonan, 2008; Field et al., 1998; Piao et al., 2019) and variation in leaf phenology within forest communities influences these cycles and potentially tropical forest resilience to climate change (Chen et al., 2020; Cleland et al., 2007). However, the magnitude of matter and energy fluxes are poorly resolved both spatially and temporally, and the underlying mechanisms through which they emerge are imprecisely described (Chen et al., 2020; Restrepo-Coupe et al., 2021, 2017; Saleska

et al., 2003). As a result of these knowledge gaps, dynamic global vegetation models (DGVMs) fail to predict observed productivity dynamics and climate-vegetation feed-backs (Chen et al., 2020; Restrepo-Coupe et al., 2017). Improved understanding of leaf phenology patterns could be vital for the development of new theoretical frameworks to predict resilience to climate change (Abernethy et al., 2018; Chen et al., 2020; Davis et al., 2022; Pau et al., 2011; Pereira et al., 2013; Restrepo-Coupe et al., 2017; Wu et al., 2016).

Patterns of whole-canopy leaf phenology vary tremendously across the tropics, from strongly deciduous in the dry tropics to weakly seasonal evergreen in persistently wet tropics (Reich, 1995). As a first approximation, the whole-canopy phenology of seasonally dry forests is influenced primarily by the timing of water limitation, whereas the phenology of rainforests is more influenced by insolation, with leaves being produced to coincide with peak irradiance (van Schaik et al., 1993; Wagner et al., 2011; Wright and van Schaik, 1994; Yang et al., 2021). In regions with a pronounced dry period, leaf fall peaks during the dry season (de Camargo et al., 2018; Frankie et al., 1974; Kumar et al., 2023; Parsons et al., 2014; Reich and Borchert, 1984; Rivera et al., 2002; Zhang et al., 2014) and new leaves are produced in anticipation of the wet season to come (de Camargo et al., 2018; Murali and Sukumar, 1993; Ryan et al., 2017; van Schaik et al., 1993). In the wetter regions of the tropics, peaks in leaf production have been observed at the end of the wet season in Uganda (Ssali and Sheil, 2023), and at the start of (or during) the dry season in wet Costa Rica, West Cameroon and the Amazon (Anderson et al., 2011; Doughty et al., 2019; Guan et al., 2015; Huete et al., 2006; Lopes et al., 2016; Myneni et al., 2007; Reich and Borchert, 1984; Saleska et al., 2016; Taffo et al., 2019). During drier periods, insolation increases as a result of reduced cloud cover, leading to debate over whether observed patterns are attributable to there being more light or reduced moisture. In the seasonally dry tropics, the degree of deciduousness (complete leaf loss) is greatest in regions with intense or long dry seasons, but is reduced in areas where the soils have better moisture retention properties (Bohlman, 2010; Ouédraogo et al., 2016) or ground water is accessible (Borchert, 1994a; Do et al., 2005). Other studies in the dry tropics link declines in soil moisture availability to increased rates of leaf shedding during dry season (Reich and Borchert, 1982) and post-drought re-hydration to the renewal of leaf production (Borchert, 1994a,b; Reich and Borchert, 1982). In these dry systems, vapour pressure deficit of air can influence inter-annual variability in timing (Do et al., 2005). In contrast, there is evidence that insolation is a key determinant of leaf phenology in the wet tropics, where peaks in leaf production have been observed to coincide with seasonal peaks in irradiance, allowing trees to maximize photosynthesis by increasing the area of highly productive young leaves (Lopes et al., 2016; van Schaik et al., 1993; Wright and van Schaik, 1994). Analyses of remote sensing data suggest that sunlight is the dominant control of leaf phenology across the tropics with soil

water stress and vapour pressure deficit playing secondary roles (Li et al., 2021a). However, large scale vegetation models suggest that vapour pressure deficit and sunlight best explain the seasonality of leaf flush and fall in the Amazon (Chen et al., 2021, 2020). Water loss is an unavoidable consequence of photosynthesis, because stomatal pores that allow CO_2 to enter leaves also allow water vapour to escape, so the optimal approach a tree takes to retaining or shedding leaves during dry periods may be governed by trends in both water availability and radiation, and their influences on photosynthetic processes (Li et al., 2021a; van Schaik et al., 1993; Wu et al., 2021). Based on current evidence, it appears that whole-canopy leaf phenology in the tropics is structured around the timing of maximum insolation unless water limitation prevents this, so for much of the tropics it may be more helpful to think of light and dark seasons (rather than wet and dry seasons) (Li et al., 2021a; Wagner et al., 2017; Wright and Cornejo, 1990). For example, one study demonstrates that different tropical forest types exhibit asynchronous responses to seasonal and El Niño-driven drought, and suggests that mechanisms controlling dry forest leaf phenology are related to water-limitation, whereas rainforests are more light-limited (Pau et al., 2010).

Whole-canopy phenology, as described above, arises from the leaf phenology of individual species, which can vary considerably within a single site as local heterogeneity in light and water regimes leaves space for a range of strategies to emerge (de Camargo et al., 2018; Frankie et al., 1974; Kumar et al., 2023; Kushwaha et al., 2011; Loubry, 1994; Ribeiro et al., 2022; Sabatier, 1983; Seyoum et al., 2012). Evergreen species, which retain foliage throughout the year, predominate in tropical moist forests, but can have strong seasonal variation in the rate of leaf turnover (Coley, 1983; Frankie et al., 1974; Kumar et al., 2023). Deciduous trees, which become increasingly prevalent in the seasonally dry tropics, can be leafless for just a few days to several months (Loubry, 1994; Williams et al., 2008). Species can be conceptualized as falling along a continuum, with leaf phenology responding purely to environmental cues at one extreme, and being controlled by an internal "clock" at the other (Reich, 1995). Leaflessness, for example, can be part of a predictable annual cycle (obligately deciduous) or only in response to specific conditions or stress (facultatively deciduous) (Harenčár et al., 2022; Stevens et al., 2016). For many species, a complex interplay between internal and external controls regulates leaf production and loss (Reich, 1995). There is some evidence that leaf phenology relates to a species' life history strategy (Álvarez-Yépiz et al., 2017; Chabot and Hicks, 1982; Méndez-Alonzo et al., 2013; Ribeiro et al., 2022). For instance, in wet/rain forests, fast-growing, light-demanding, pioneer species with short life spans tend towards continuous and asynchronous leaf production and leaf senescence (Coley, 1983), while late successional shade-tolerant species show episodic growth in which branches or a whole crown will flush new leaves, followed by a period without further development (Reich, 1995). In seasonally dry tropical

forests, deciduousness is associated with short leaf life span and rapid acquisition of resources, whereas evergreen leaf habit is associated with long leaf life span and a more conservative strategy (Chabot and Hicks, 1982; Sobrado, 1991; Wright et al., 2004). The length of the dry season also dictates the option space: many species are synchronously deciduous in forests with long dry seasons (Bullock and Solis-Magallanes, 1990; Lieberman and Lieberman, 1984; Reich and Borchert, 1982, 1984) whereas wetter forests tend to support species that produce foliage at different times of the year (Reich et al., 1991). Under these conditions, there can be considerable intraspecific variation too: a single site can support highly synchronous species co-occurring with completely asynchronous species (Borchert, 1980; Morel et al., 2015; Reich and Borchert, 1984, 1988; Wright and Cornejo, 1990). Although leaf phenology is usually discussed in relation to trade-offs between photosynthesis, water balance and irradiance, it can also relate to pollination strategies (some deciduous trees produce flowers before leafing, making it easier for pollinators to access the flowers (Frankie et al., 1974)) and avoidance of herbivory (Aide, 1993). One study suggests that the evolution of leaf phenology is labile in the tropics (Pau et al., 2011), but more work on phylogenetic constraints is needed (Davis et al., 2022).

Until recently, field campaigns were the sole approach to monitoring tree-level leaf phenology, but these were necessarily limited in scope and duration (see Reich, 1995 and Abernethy et al., 2018 for an overview). Satellite remote sensing has long been used to characterise and understand whole-canopy phenology patterns including the green-up of Amazonian forests during dry periods (Huete et al., 2006; Myneni et al., 2007; Saleska et al., 2016; Wang et al., 2020), but the suitability of these optical methods to detect changes in LAI and disentangle effects of sun-sensor artifacts from subtle phenological signals is contested (Morton et al., 2014; Samanta et al., 2010), and they do not have the resolution to monitor individual trees. Phenocam networks have allowed for near constant monitoring of individual trees across a range of systems but each camera is limited by its field of view and phenocams are therefore unable to capture the true range of patterns that can be observed at a single site (Alberton et al., 2023, 2017). This study uses new drone technologies, combined with careful fieldwork at the Paracou field station in French Guiana, to bridge the knowledge gap. Unoccupied Aerial Vehicles (drones) present a unique opportunity to observe the leaf phenology of individual trees over time (Park et al., 2019). Being cost-effective and capable of frequent flights over the same area, UAVs equipped with multispectral and LiDAR sensors can provide high-resolution data, enabling a more nuanced understanding of tropical forest dynamics. By scanning 60 ha of seasonal moist forest with optical and lidar sensors every three weeks for nearly three years, we could extract phenological signals for 3000 mapped tree crowns and observe how the range of patterns observed related to the environmental seasonality. We processed these signals

to extract phenological metrics for individual trees and species to quantitatively characterise the timing and variation of leaf flushing and shedding across the site. There has been limited long-term time series acquisition and there remains uncertainties around what triggers leaf emergence and shedding across species (Abernethy et al., 2018; Morellato et al., 2016). We address the following research questions:

1. *How does canopy plant area index (PAI; the area of vegetative matter) vary seasonally?* We expected that, as Paracou has a brief dry season and unlikely to experience water limitation, the total canopy leaf area (and therefore PAI) would peak during the driest (and sunniest) time of year.

2. *Does variation in canopy and crown greenness (proxy for leaf quality/age) align with variation in PAI (leaf quantity)?* We expected that the canopy would flush new leaves at the transition into the dry season as it prepared new leaves to make the most of increasing insolation, creating a peak in greenness. At the crown level, we expected PAI and Green Leaf Index (GLI) to oscillate synchronously (with flushing and shedding) but for there to be divergence within the cycles as greenness would decline gradually with leaf age but the total leaf amount would stay fairly constant until shedding (Chavana-Bryant et al., 2017; Wu et al., 2019a).

3. *Do species vary markedly in leaf phenology?* We anticipated that, because the dry season is brief and not severe, periods of water stress are unlikely in typical years, and species can therefore exhibit a wide range of leaf phenology patterns. It is unlikely that a single habit will be dominant (Kikuzawa, 1991), but where full leaflessness does occur, it is unlikely to be for more than a short fraction of the year.

4. *Do differences observed among species relate to their resource acquisition strategy?* We expected that the following groups of strategies could be distinguished by their distinctive leaf phenologies: (1) fast-growing pioneers - little cyclical predictable behaviour, (2) Strongly seasonal brevi-deciduous / seasonal evergreens, (3) Evergreen, asynchronous (Reich, 1995).

5. *Do topographic position and tree height help to explain inter- and intra-specific variation in leaf phenology?* We anticipated that taller trees and trees with high relative topographic position (large TPI; ridge locations) would have a more clearly defined phenological pattern and flush earlier due to their relative hydraulic sensitivity and that differences would be intensified for trees that are both tall and at high TPI (Bittencourt et al., 2022; Liu et al., 2021; Markesteijn et al., 2011).

6. *Do the observed leaf phenological traits of species exhibit a phylogenetic signal?* Based on the predictions of (Pau et al., 2011) study we did not expect phenological traits to be phylogenetically conserved.

Here, we address these questions, by analysing leaf phenology patterns recorded in a detailed spatio-temporal survey of the forest at Paracou Research Station over a 30 month period. Applying the AMAPvox modelling framework to track lidar pulses through the canopy at each time step (Vincent et al., 2017, 2021), we estimated seasonal variation in canopy Plant Area Index. By extracting crown-level information on Green Leaf Index (GLI; a proxy for amount of leaves and also their age/quality) and Plant Area Index (PAI; the amount of vegetative material) for each time step across 3000 tree crowns, we compared how leaf spectral signal and leaf amount co-varied at the individual level through time. The spectral signals were represented as continuously varying phenological *waveforms* that corresponded to the observed oscillatory patterns of leaf development. To visualise the range of phenological patterns we generated a series of *phenograms* which were heatmap representations of the the GLI time series with each individual tree plotted through time, and the individual time series were grouped by species. We analysed the crown-level phenological waveforms of individual trees by developing phenological metrics to quantitatively characterise the variation of leafing patterns across 100 of the dominant species at the site. Specifically, we extracted the periodicity and amplitude of the signals, and the timing of peak flushing and shedding. We compared waveforms across individuals to determine the synchronicity and regularity of a species' leaf patterns and evaluated the degree of seasonality of flushing and shedding for each species. Finally, we compared the phenological metrics to local variation in topography (topographic position index) and tree height, and analysed the metrics for phylogenetic signal (Pagel's λ and Blomberg's K) to try to explain the observed variation in leaf phenologies. To our knowledge, this is the first time that variation of leaf phenological patterns at the individual and species level has been investigated in relation to environmental and phylogenetic drivers in the wet tropical context.

5.2 Materials and Methods

5.2.1 Study site

The research was conducted in forests at Paracou Field Station, French Guiana (5°16'N 52 °55'W) (see Fig. 3.1). The lowland tropical moist forests grow mostly on shallow ferralitic soils underlain by a variably transformed loamy saprolite (Gourlet-Fleury et al., 2004). The mean annual rainfall is approximately 3200 mm with a three month dry season from mid-August

to mid-November. September, the driest month will typically have 50 mm of rainfall (Bonal et al., 2008; Wagner et al., 2011) which makes the climate just within tropical monsoon regime of the Köppen climate classification system (tropical rainforests have a minimum monthly precipitation of 60 mm) (Köppen, 1884). Full details including a map of the site are given in Section 3.2.1.

To generate a larger dataset than would have been obtained from manual crown delineation alone, we combined the manual crowns with crowns generated with the automatic methods described in Chapter 4. For any area of forest where a manual crown and an automatic crown overlapped, only the manual crown was retained. To further minimise contamination of the phenological signals with noise from poorly delineated or incorrectly identified automatic crowns, we only included automatic crowns if they scored well on both the "delineation" and "species identification" confidence scores assigned to them (as described in Chapter 4). An automatic crown polygon required a strong match ($IoU > 0.75$) across three independent dates and good average delineation confidence at each date to be considered satisfactory. Then, to be retained for the species-specific phenological metric extractions, at least 50% of the pixels contained within the crown had to have been classified as the same species. This is a stricter threshold than it might otherwise seem as crowns often contained some shaded (NA) pixels, pixels that contained lianas, and pixels around the edge of the crown mixed with neighbouring individuals, arising from potential misalignment issues between the lower spatial resolution hyperspectral scan and the more precise RGB surveys. This method of expanding sample sizes was not particularly valuable for the most common species in the flux tower region (that already had a large sample size from the manual crown dataset) or the rarest species across the whole of the Paracou site (that were unlikely to get an addition of automatic crowns due to low "species identification" confidence scores). However, it was helpful for expanding the sample sizes of those species that had good representation across the entirety of the hyperspectral scan (see Chapter 4) but few manual crowns specifically in the flux tower footprint scanning region. After this selective expansion of the dataset, the 100 most common species were selected for analysis of their patterns as these had sample sizes sufficient to reasonably estimate the phenological metrics described below.

5.2.2 Remote sensing data sources and processing

Both lidar and RGB scans were taken approximately every three weeks to compare how the spectral properties of the top of the canopy changed with the PAI measured through the canopy (see Table 5.1). For the lidar scans we included a "dense" area at the core of the flux tower footprint with an extremely high point density and a "large" area covering approximately 60%

Table 5.1 The remote sensing data sources used to investigate leaf phenology. Resolution is given as ground resolution for the RGB orthomosaic and as the average point density for the lidar scans. Altitude is given as height above canopy.

Scan dates	Modality	Resolution	Altitude	Wavelengths	Sensor
18-Oct-2020—23-Sep-2023	RGB	5 cm	70 m	400-750 nm (3 bands)	1" CMOS (P4M and Mavic 2)
19-Oct-2020—30-Nov-2022	LiDAR (dense)	430 points/m^2	75 m	905 nm	RIEGL miniVUX-1UAV
19-Oct-2020—03-Jan-2023	LiDAR (large)	100 points/m^2	105 m	905 nm	RIEGL miniVUX-1UAV

of the total flux area with a lower point density (see Fig. D.1). This was to compare scanning parameter requirements but is not elaborated on here (but see Vincent et al., 2023).

RGB

Mosaicking: The RGB orthomosaics were compiled from the raw geotagged UAV photographs using structure from motion (SfM) photogrammetry in AgiSoft Metashape. The software aligns overlapping images to produce a sparse point cloud, refines it into a dense point cloud, and subsequently constructs a 3D mesh. This mesh, integrated with original photo textures, facilitates the creation of a Digital Elevation Model (DEM). The DEM, combined with the aligned images, allows for the generation of an orthomosaic, a georeferenced image free from perspective distortions. Supplying images across several dates in single blocks to the first steps of the SfM processing improves spatio-temporal coherency (Feurer and Vinatier, 2018a). Following this approach, instead of processing each date separately, five date blocks were supplied for the alignment and initial sparse point cloud formation establishing a common geometry between dates. The dates were then separated for the dense matching steps and final orthomosaic generation.

Co-registration: Accurate co-registration of data from RGB and hyperspectral imagery was important to ensure spatial alignment. We used the LiDAR-derived Canopy Height Model (CHM) as the baseline layer, with all other data being registered to it. This choice was due to the CHM's stability and precision in representing the physical landscape, providing a solid reference for co-registration. Eight control points were manually assigned across the different datasets, using identifiable features within the LiDAR CHM, such as the flux tower, roads and dominant trees, as primary reference points and an affine transform was applied based on these. This co-registration process ensured that the crowns represented across the datasets corresponded to the same geographical location, serving as the foundation for subsequent analysis steps, including tree crown delineation and species classification.

Greenness metric extraction: The spectral properties of leaves are known to change through time as they age (Chavana-Bryant et al., 2019, 2017; Kitajima et al., 1997; Wu et al., 2019a). Several vegetation indices were compared before deciding which to use for assessing

temporal trends. The Green Leaf Index (GLI; Louhaichi et al., 2001) was selected as it gave a clear temporal signal that corresponded well to the observed leaf stages at the crown level (Arkin et al., 2023; Ciocırlan et al., 2022; Vorovencii et al., 2023; see Fig. 5.4):

$$GLI = \frac{(Green - Red) + (Green - Blue)}{(2 * Green) + Red + Blue} \tag{5.1}$$

where Red, Green and Blue are the colour coordinate values of the respective bands at each pixel of the processed orthomosaic (0-225). Due to local perturbations and varying conditions across the orthomosaics (from atmospheric differences and differences in illumination from cloud cover and shading between trees) that affected the bands in inconsistent ways, it was necessary to perform a local normalisation when extracting the crown level metrics. This involved dividing the average index within the crown by the index measured in a 25 m buffer around the crown to give a locally normalised crown average value. In effect this gave a value of GLI for a crown relative to that of the local background vegetation (see Fig. D.3). To account for the irregular sampling through time, we interpolated the GLI data so that each day was assigned a value. This was done by fitting splines using the *zoo* R package (Zeileis and Grothendieck, 2005). This allowed us to analyse and compare between the signals of trees and calculate the phenology metrics described below.

Point clouds to plant area

Bayesmap (Jalobeanu and Gonçalves, 2014) and the Iterative Closest Point (ICP) algorithm in CloudCompare (Besl and McKay, 1992; Rusinkiewicz and Levoy, 2001) were used to align the point clouds through time. To calculate the PAD through space at each point in time from the point clouds it was necessary to model the extinction of light through the canopy. This was done by combining the point cloud information with the flight trajectory information and tracking each laser pulse through the canopy in the AMAPvox software package. Details can be found in Vincent et al. (2017, 2021). PAD and PAI computed from lidar returns have been shown to be sensitive to the lidar system characteristics, particularly so to the transmitted power and the receiver's sensitivity (Vincent et al., 2023). In the present case a conversion factor was applied to the raw PAD profiles derived from the low power UAV-LS. This conversion factor was derived by aligning UAV-LS profile to a coincident PAD profile obtained using TLS over a one hectare plot. A voxel size of 1 m was selected as it gave a sub-crown spatial resolution while maintaining the required sampling density of lidar points within each voxel. We also tested a 2 m voxel size and the signals were consistent with the 1 m method. To compare the PAD measurements with the GLI measurements, the top 3 meters of PAD was summed before extracting a value within each tree crown polygon. This ensured that we were

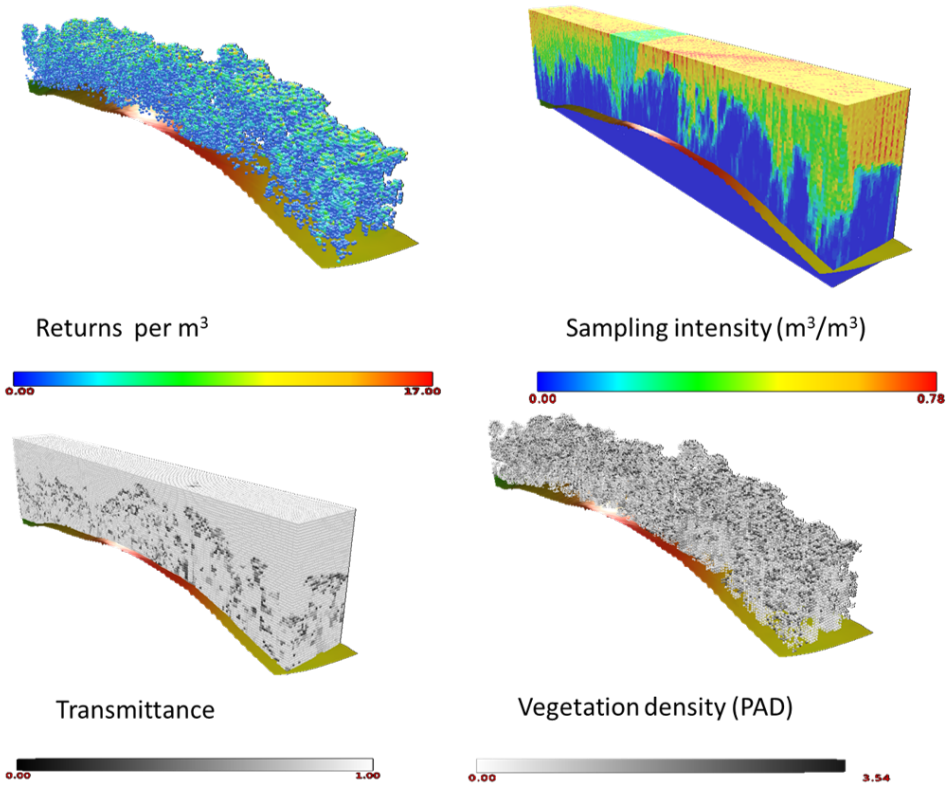


Fig. 5.1 A simplified visualisation on the stages of PAD calculation using AMAPvox. Extinction is modelled based on the tracing of lidar pulses through a voxilised space using the UAV flight trajectory data combined with the point cloud. From the way the light pulses are partially intercepted, the PAD in each of the voxels is inferred. For full details see Vincent et al. (2017, 2021). Figure credit: Gregoire Vincent.

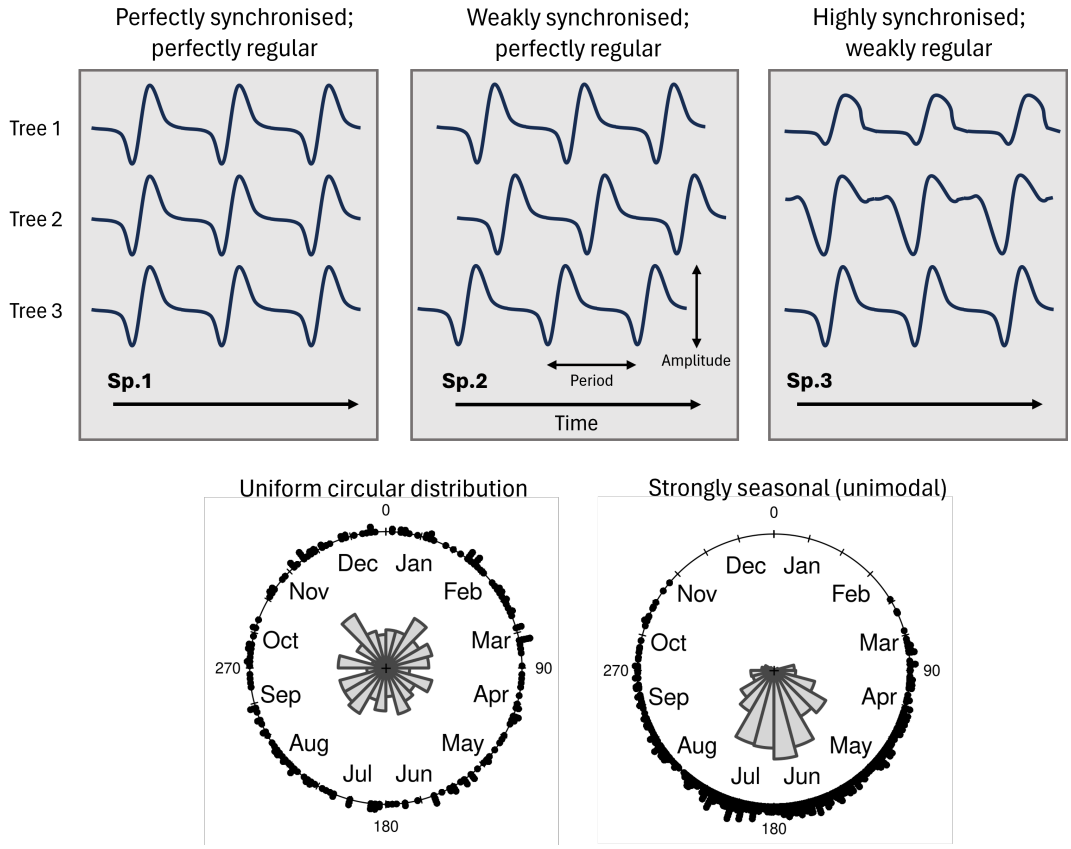


Fig. 5.2 Illustration of the key phenological metrics used to characterise trees and species. As a tree cycles through its leaves the GLI will oscillate with some kind of characteristic shape and frequency (a waveform). A peak in GLI is seen with a peak output of fresh leaves, a decline is seen with leaf aging/senescence and a trough is seen with shedding (minimum leaf balance). Trees of a single species can have a very *regular* signal (i.e. each tree's GLI waveform is very similar) or irregular signals (i.e. the waveforms appear very different). Trees' signals can be highly synchronous (i.e. aligned clearly in time) or asynchronous (i.e. no clear alignment in time). The distribution of flushing or shedding events can be uniformly distributed through the year or clustered around a seasonal peak.

focusing our analysis on just the crown level vegetation (that correspond with canopy surface GLI measurements) rather than including variation from the understorey.

5.2.3 Quantifying the dimensions of phenology (key phenological metrics)

By overlaying each tree crown outline on the time series of GLI and PAD it was possible to extract the leaf phenological time series. To visualise the range of phenological patterns we generated a series of *phenograms* which were heatmap representations of the GLI time series. For each individual the GLI was plotted through time, and the individual time series

were grouped by species and stacked. To make sense of the phenological time series of each tree/species it was necessary to extract some key phenological metrics from the signals (see Fig. 5.2). In particular, we were interested in characterising:

- The timing of flush/shedding (of individuals and average/spread of species) relative to the dry season
- Period of phenological cycles (of individuals and average/spread of species)
- Amplitude of the phenological signals (of individuals and average/spread of species)
- The relative seasonality of flush/shedding (within species) i.e. do events happen uniformly through the year or do they tend to be clustered around a specific point of the year?
- Synchronicity of the signals (within species) i.e. how well aligned in time are the waveforms?
- Regularity of the signals (within species) i.e. how similar were the waveforms?

When dealing with phenological timings it is important to treat the data as circular (i.e. periodic, repeating) otherwise results may be biased or misleading (Staggemeier et al., 2020). For this it was necessary to find a point in the year to centre the data (zero point). As previous studies have suggested precipitation predicts phenology well, we used rainfall as the metric to centre on. Based on historic daily meteorological data from 2004 to 2023 a smoothed precipitation profile over the year was constructed. From this, a central driest day was calculated (27-Sept) to act as a zero point for flushing and shedding events over the course of the study. There was significant change in overall annual rainfall prior to and during the study: 2020 was slightly drier than usual (2986 mm) whereas 2021 (4238 mm) and 2022 (4337 mm) were particularly wet. However, 27-Sept was a representative centering of the dry season over the study period (see Fig. D.2).

Timing of minimum and maximum crown greenness (individuals to species): For each tree, the dates of peak and minimum GLI were identified. The time to the temporally closest central dry day (27-Sept) was extracted for each point (e.g. 26-Sept-2022 would be -1 and not +364). To find the representative dates for all trees and for each species, we converted the differences in days to degrees, leveraging circular statistics. This conversion allowed for the computation of average angles, which correspond to the mean timings of peak and minimum greenness across all trees. A (circular) mean of these timings over all crowns for each species was calculated along with their respective standard errors. Subsequently, the averaged angles were transformed back into days, revealing the average peak and minimum greenness timings across the dataset.

Periodicity (individuals to species): To estimate the periods over which individual trees exhibit greenness patterns the temporal autocorrelation across dates was calculated. This evaluated how well the greenness of a tree over a given interval correlates with its past/future values. After computing the temporal autocorrelation for a tree's greenness, we identified the length of time over which the autocorrelation values are maximised (enforcing that maximum value, that is at least 90 days from the start to exclude the influence of short-term fluctuation in the analysis). The maximal autocorrelation time then represents the periodicity of a tree's greenness pattern. For instance, if this point occurs at 365 days, it suggests an annual pattern in the tree's greenness. The periods of individual trees were then aggregated to give average periods for each species.

Amplitude (individuals to species): For each tree, the difference between the maximum and minimum observed GLI values was calculated, giving an amplitude value that signifies the magnitude of GLI fluctuation (relative to the background vegetation in the 25 m window). This gave an indication of the magnitude of spectral change in the crown and the degree of difference between peak flushing and shedding. For example, we expected amplitude to be low for weakly seasonal evergreen species and large for crowns that underwent full leaflessness and flushed fully within 1-2 weeks. The individual amplitudes were averaged at the species level.

Synchronicity (within species): To measure how similar the leaf stage of trees of the same species was through the year we calculated their synchronicity. First, the cross-correlation across the time series of each pairwise combination of trees was computed. This cross-correlation matrix quantifies how well the greenness patterns of each tree pair correlate with each other (without allowing any time lags). The average of these pairwise correlations give a single synchronicity value for that species.

Regularity (within species): This method assessed how similar the greenness waveforms / phenological patterns of individual trees of the same species were to one another regardless of when in the year the patterns of each individual were occurring. Flexibility (given through time lags) was allowed to best align the pairwise patterns of individual trees. Specifically, it used the average pairwise cross-correlation across crowns of a species. This differs from synchronicity, as a flexible time lag (of up to one year) is introduced to the GLI signal of one of the pair of trees and the lag that leads to the best match between the patterns is calculated. Once matched, the correlation between the signals is calculated. This matching meant the correlation value for regularity will always be greater than that of the synchronicity value. The regularity value essentially evaluates whether trees of the same species exhibit similar leaf cycling patterns without requiring that they are seasonally aligned. For every pair of trees within a species, the highest correlation and the associated lag are stored in matrices. From these matrices, average synchronicity and lag values are calculated for each species.

Seasonality of flushing and shedding (species): From the angles of peak and minimum greenness (as described above) a radial length for each species was calculated (along with the average angle). This radial length represents how strongly the peak/minimum dates were clustered around the mean angle/date. This length gave a measure of seasonality of flushing/shedding across species that could be tested with the (circular) Rayleigh test of uniformity which assesses the significance of the mean resultant length. The null hypothesis was that the flushing/shedding is a uniform circular distribution and the alternative hypothesis that they form a unimodal distribution with some mean direction and mean resultant length (see Fig. 5.2). This allowed us to test whether there was significant annual seasonality in flushing and shedding for each species.

Metric ordination: To understand whether phenological characters could be separated based on the extracted metrics, we used Phylogenetic Principal Component Analysis (pPCA) which is a statistical method that combines the principles of Principal Component Analysis (PCA) and phylogenetic comparative methods (Revell, 2009, 2012). It aims to identify and understand patterns of variation and covariation in multivariate trait data among related species, while accounting for the shared evolutionary history that can confound traditional statistical analyses. We plotted the species on the axes of the first two resultant principal components and coloured points by tree height to try to elicit a pattern that could be related back to resource acquisition strategy.

5.2.4 Testing the influence of tree height and topographic position

To begin to understand whether heterogeneity in hydrological conditions influences phenological patterns, we compared the observed phenological metrics of the trees to their Topographic Position Index (TPI; Guisan et al., 1999). TPI compares the elevation of a point on the Earth's surface to the average elevation of a surrounding area. This comparison can help to determine whether a point is, for example, in a valley, on a ridge, or on a slope or flat ground. Valley bottoms are wetter than ridges due to runoff and drainage (Ågren et al., 2014). TPI is calculated by comparing the elevation of a specific point to the mean elevation of a surrounding area. This area is defined by a window size that can vary depending on the scale of analysis. A ridge location would have a positive value whereas a gully location would have a negative value. While TPI correlates with relative water availability across a landscape, other factors including soil properties and variable upstream catchment areas can lead to divergences between TPI based estimates of soil moisture and direct soil moisture content measurements (Jarecke et al., 2021). Using a lidar derived DTM we calculated the topographic position index with a 91 m window at 50 cm resolution across the site. This scale was chosen as it was similar to the size of the forest plots over which soil conditions are known to vary considerably. From this

we extracted the median TPI value within each tree crown. The mean TPI was calculated for each species to determine a habitat preference and the standard deviation of TPI values was calculated to indicate the degree of habitat specialisation.

Tree height is related to tree hydraulics and light availability (Bittencourt et al., 2022) and so it is likely that tree stature influences observed phenological patterns (Itoh et al., 2003). To test this effect, we also compared tree heights to the observed phenological metrics. Using a 50 cm DSM and DTM we extracted the tree height, as the 98th percentile height within each tree crown. Selecting the 98th percentile reduced the chance of defining height based on an outlying pixel. The mean tree height was calculated for each species to determine a characteristic height or *stature*. The standard deviation of height values was also calculated to determine the range of light conditions a species can tolerate. We allowed for potential interaction between TPI and tree height.

Interspecific variation

To assess the relationship between topographic position preference and characteristic tree height and the species level phenological characteristics we used Phylogenetic Generalized Least Squares (PGLS) analysis. PGLS accounts for the non-independence of species data due to shared evolutionary history. This method incorporates a covariance matrix derived from the phylogenetic tree of the studied species, adjusting the regression for their phylogenetic relationships. The degree of phylogenetic correlation in trait evolution was assessed using the lambda (λ) parameter. We used the time calibrated phylogeny of Baraloto et al. (2012) which included 95 of the 100 species to derive the covariance matrix.

We wanted to understand whether the average tree height, the spread of tree heights and the average TPI and the spread of TPI had an influence on the (1) timing of flush, (2) period of leaf cycle, (3) synchronicity of signal, (4) regularity of signal, (4) seasonality of flush and (5) seasonality of shedding. We hypothesised that (H_1) some interacting combination of species height (i.e. relative location of crown within the canopy) and topographic position index (i.e. association with gully through to ridge conditions) would explain the variation seen between species (H_0: no association between phenological metrics, and species height and TPI preference).

Intraspecific variation

To assess the effect of topographic position and tree height at the tree level we used mixed effects model with species as a random effect. We hypothesized that some combination of tree height and relative topographic position (TPI) would explain the intraspecific variation in leaf flush timing, leaf shed timing, period of leaf cycle and amplitude (H_0: no association between phenological metrics, and tree height and TPI).

5.2.5 Phylogenetic signal

To investigate whether closely related species exhibited similar leaf phenology patterns we tested the phenological metrics for *phylogenetic signal* which is the degree to which related species resemble each other with respect to a specific trait. For this we used the time calibrated phylogeny of Baraloto et al. (2012) which included 95 of the 100 species. The timing of flushing and shedding relative to dry season are circular metrics and if it is not treated as such the phylogenetic signal would be underestimated (Staggemeier et al., 2020). We transformed the timing data using principal coordinate analysis of pairwise angular distances (as described in Staggemeier et al., 2020) before testing for phylogenetic signal using Pagel's λ (Pagel, 1999) and Blomberg's K (Blomberg et al., 2003), which differ in their sensitivity to certain aspects of the data: λ is a scaling parameter for the correlations between species, relative to the correlation expected under a Brownian motion model of evolution; K is a scaled ratio of the variance among species over the contrasts variance (which measures how much more closely related species resemble each other than would be expected under a Brownian motion model of evolution). Blomberg's K can be sensitive to incompletely resolved phylogenies and suboptimal branch-length information. Using Blomberg's K with polytomic chronograms (incompletely resolved phylogenies) can give inflated estimates of phylogenetic signal and moderate levels of type I and II biases. Furthermore, pseudo-chronograms (phylogenies with suboptimal branch-length information) can lead to high rates of type I biases. In contrast, Pagel's λ is more robust to incompletely resolved phylogenies and suboptimal branch-length information (Molina-Venegas and Rodriguez, 2017). This means it may be a more appropriate alternative over Blomberg's K to measure and test phylogenetic signal in ecologically relevant traits when phylogenetic information is likely to be incomplete (Molina-Venegas and Rodriguez, 2017). For completeness, we present both Pagel's λ and Blomberg's K but noted their relative sensitivity when analysing the results of each.

5.3 Results

5.3.1 Plot-level variation in plant area index

There was an approximate 10% change in PAI through time at all four sampled plots (see Fig. 5.3). While the average PAI across the plots ranged from 8 m^2/m^2 to over 9 m^2/m^2 the seasonal pattern of change across the plots was consistent. The PAI started to increase prior to the dry season, continued to increase during the dry season reaching its maximum value towards the end of the dry season before plateauing and declining again during the wetter months.

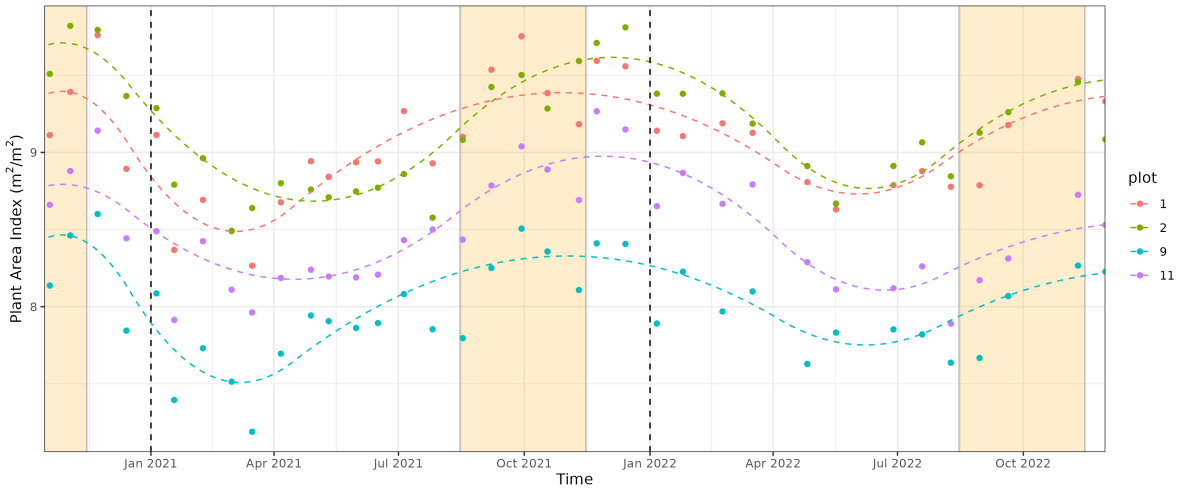


Fig. 5.3 Changes in PAI measured at four plots. PAI was inferred from extinction that was calculated by tracing the lidar pulses (that penetrated the canopy) through a voxelised space and recording each partial interception (using AMAPvox). The yellow box represents the dry season timing. Smoothing splines were fitted to illustrate the overall patterns.

5.3.2 Crown level phenology patterns

Clear leaf phenological patterns were observed at the crown level. Fig. 5.4 gives an example of the phenophase of a *Parkia nitida* (Fabaceae) crown. From 2020-10-23 leaf senescence is visible until in 2020-12-14 a single branch sheds its leaves. By 2021-01-18 that branch has flushed new leaves while the rest of the crown has shed. By 2021-02-08 a verdant crown with fresh leaves is visible. Fig. 5.4b shows how the plant area density (PAD) profile of the tree changes during this period. The crown (from about 23 m upwards) shows a substantial oscillation of leaf amount over this period. Fig. 5.4c shows how this translates into the normalised GLI and PAD metrics through time with the full time series included to illustrate the cyclical nature of its flushing and shedding. A strong peak in GLI is observed as new leaves are flushed followed by a drop as leaves mature and a gradual decline as leaves undergo senescence. PAD on the other-hand peaks less strongly with the emergence of fresh leaves and holds more steady until there is a rapid drop as leaves are shed.

5.3.3 Looking across crowns and species

The extrapolated tree crown map described in Chapter 4 allowed us to look across 3000 species labelled crowns across the landscape and explore the variability between individuals and species. A great variety of behaviours was observed (see Fig. 5.5). Where periodicity was observable, it was usually on a cycle of 9-24 months. Some species put out fresh leaves in advance of dry season (e.g. *Pouteria eugeniifolia*, *Manilkara bidentata*), some during (e.g. *Sextonia rubra*,

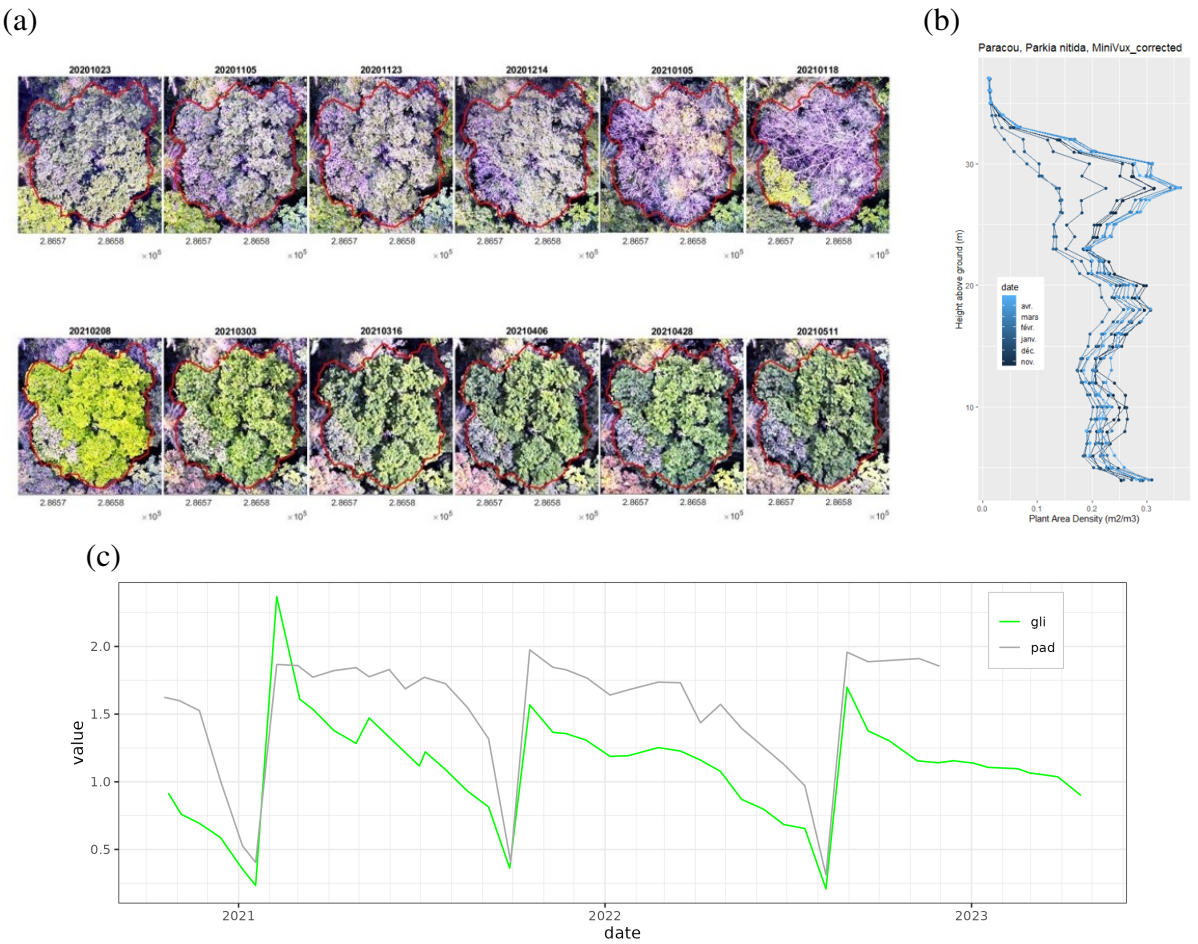


Fig. 5.4 Seasonal changes in a *Parkia nitida* crown: **(a)** shows the pattern of shedding and flushing of the crown over a year in the cropped photographs. **(b)** shows the vertical profile of PAD change in the same period with the shedding in the crown clearly visible. **(c)** shows the normalised Green Leaf Index (GLI) and Plant Area Density measured across the crown over the full length of the available time series.

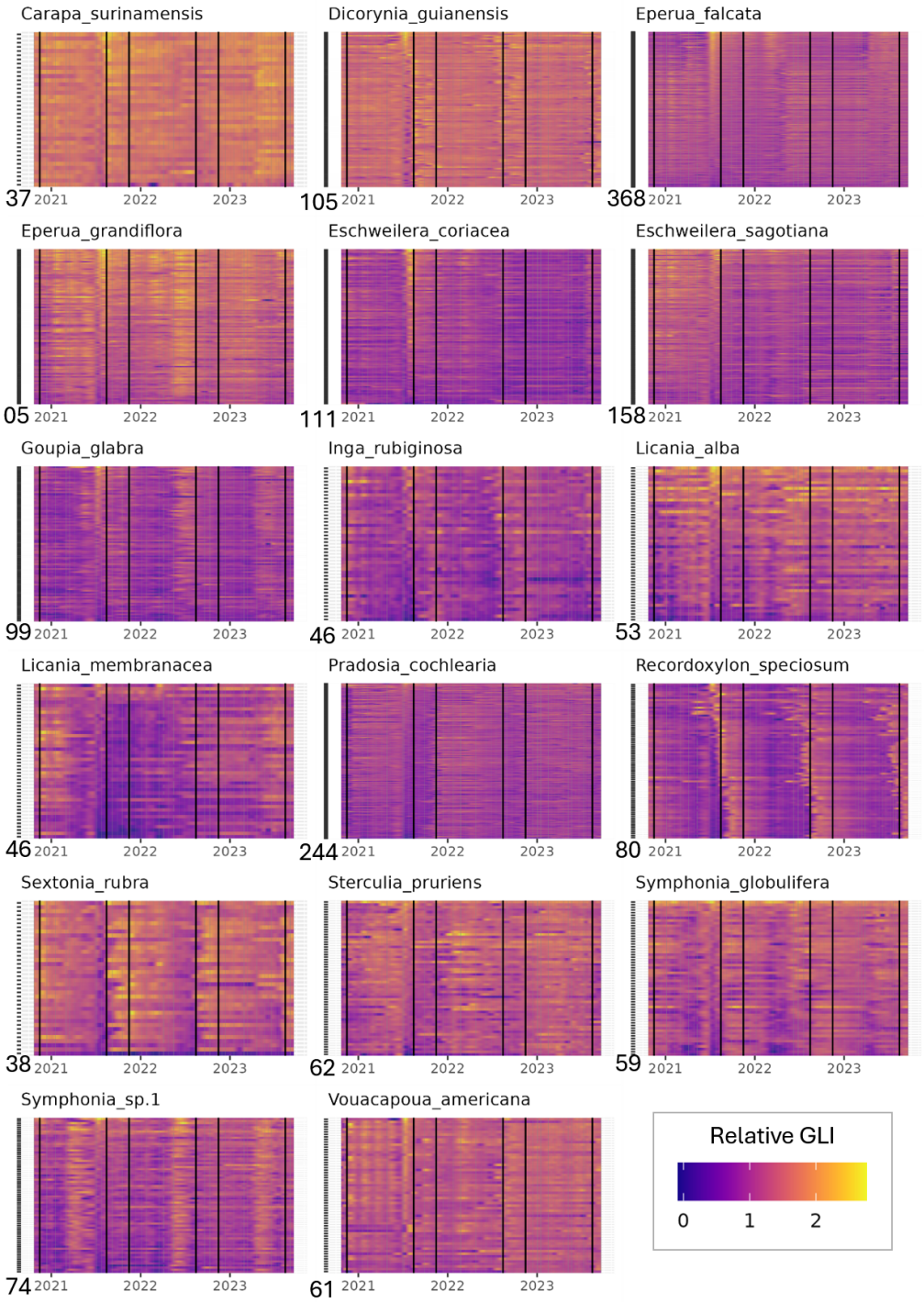


Fig. 5.5 Phenograms for seventeen common species. The heatmaps show the temporal variation in relative Green Leaf Index (GLI) for each species. The y-axis (with ticks) shows each individual of the species and the x-axis shows time. The number of individuals of each species observed is given at the bottom of the y-axis of each sub-plot. To highlight the variation in the patterns, the individuals were ordered along the y-axis based on the relative GLI value between 2021-06-15 and 2021-08-30 with the individuals with the largest summed values positioned at the top. The variation is normalised against the background GLI so that 1 represents a value equal to the GLI of the surrounding canopy and 2 represents a GLI value double that of the surrounding canopy. Bright colours show fresh leaves whereas dark colours show aging/senescence/shedding. The black vertical lines show the beginning and end of the dry season.

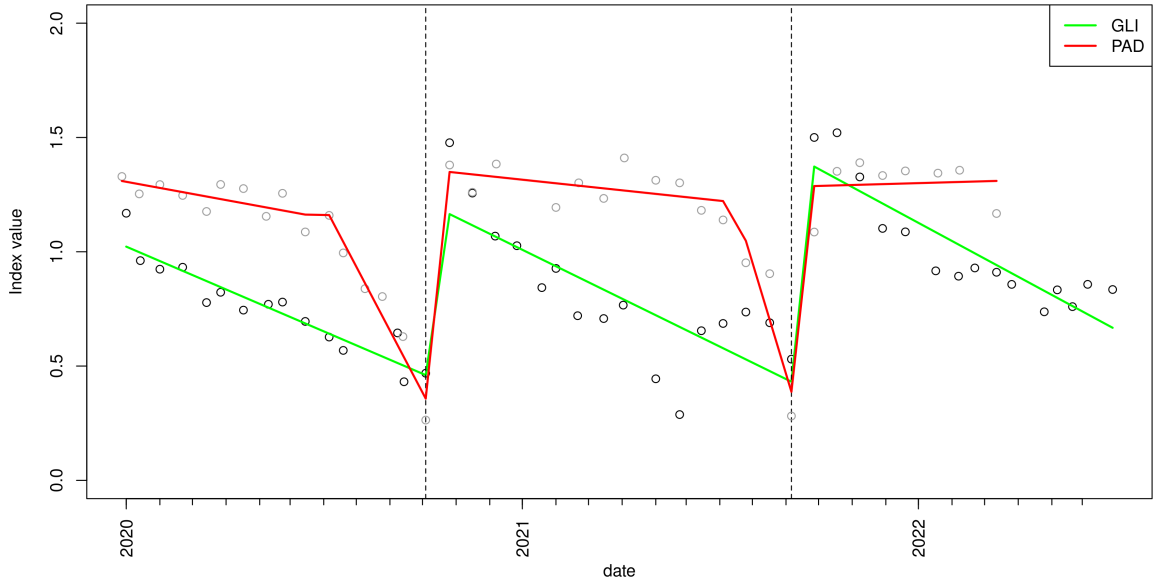


Fig. 5.6 Normalised PAD and GLI profiles for a typical *Recordoxylon speciosum* (Fabaceae) crown. The lines are derived from breakpoint analysis whereby separate linear functions could be fitted in different portions of the time series based on improvements to fit estimated with the Bayesian information criterion (using the *strucchange* R package). They show the gradual decline in leaf quality compared to the leaf area that remains relatively steady until shedding. The full leaf off phase is very brief (0-3 weeks).

Parkia nitida, *Albizia pedicellaris*) and some just after (e.g. *Pradosia cochlearia*). Some species were highly synchronised (*Albizia pedicellaris*, *Microphilis obscura*), others had a range of dates over which individuals shed/flushed (*Recordoxylon speciosum*) and some were highly asynchronous (e.g. *Parkia pendula*). Some species were possibly grouped into two cohorts on inter-annual cycles (e.g. *Chrysophyllum pomiferum*). Of the three most common families, Fabacaeae and Sapotaceae tended to show more obvious patterns whereas species of Chrysobalancaceae seemed to be more cryptic in their patterning (see Figs. D.4 and D.5). Many species did not exhibit a clear leaf-off phase but still exhibited some periodicity in leaf renewal.

Those species that exhibited clear annual cycles showed a fairly predictable relationship between GLI and PAD (see Fig. 5.6). The greenness of leaves decays gradually as the leaves age. The amount of leaves remains relatively stable until a brief leaf shedding event leads to a sharp decline. Greenness can be thought of as a proxy for photosynthetic capacity (or leaf quality). This divergence between leaf amount and leaf quality within the leaf cycle gives some opportunity for comparing their relative importance with respect to forest primary productivity (Wu et al., 2017b).

5.3.4 Key phenological metrics

To quantitatively characterise the observed phenological patterns (and the inter- and intra-specific variation) we processed the signals from the crowns into key phenological metrics that described the strength, timing and agreement between signals. The pairwise correlation between the variables at the species level is shown in Fig. D.10.

The phylogenetic PCA revealed some evidence of trends in phenological strategies, but no obvious clusters. Towards the bottom of Fig. 5.7 shorter species had highly variable periodicity and little to no regular phenological structure to their leaf patterns (possible pioneer species). Towards the top of the plot lie taller trees with highly structured leaf phenology patterns, typically with strong seasonality in flushing/shedding and synchronicity (possible deciduous strategy). Towards the right of the plot lie taller trees with long periods and weakly structured leaf phenology patterns (possible evergreen strategy).

Average peak crown greenness (weighing each crown equally) occurred 70.8 $\pm$ 2 days before the driest point of the year and minimum crown greenness 72.2 $\pm$ 2.5 days before the driest point of the year (20th of September). By comparing the radial histograms of Fig. 5.8 it was possible to understand the relative distribution through the year of flushing and shedding and why average peak crown greenness and peak crown shedding occur so close to one another. Green-up happens suddenly just before the start of the dry season in July as many crowns reach peak greenness coincidentally. The number of crowns flushing remains high for a period but decreases gradually as we move into the dry season. On the other hand, the number of crowns shedding or reaching their minimum greenness builds gradually before the dry season, peaking at the start of the dry season (August) before drastically dropping off.

By plotting the mean peak in greenness for each species it was possible to see the interspecific variation in green-up timing (see Fig. 5.9). The majority of species greened up as rainfall started to decline at the end of the wet season and as the dry season was commencing. 63% of species had their peak in flushing in the four-month window between 120 and 0 days ahead of the driest point of the year.

The majority of trees showed a periodicity of close to one year (see Fig. 5.10). The strong peak had significant tails either side. There were secondary peaks that were less than 9 months and a secondary peak close to 2-years. Some species showed very close agreement among individuals whereas other species showed a great deal of intraspecific variability (see Fig. D.7).

Just over half of species showed seasonal flushing (54%) and seasonal shedding (50%; although not necessarily the same species) (see Fig. 5.11). Synchronicity, regularity, amplitude of oscillation and seasonality of shedding/flushing were all correlated (see Fig. D.10) but key differences between species are clear where there was divergence. For example, *Chrysophyllum pomiferum* (Sapotaceae) had a strong (large amplitude) regular signal but the leaf patterns

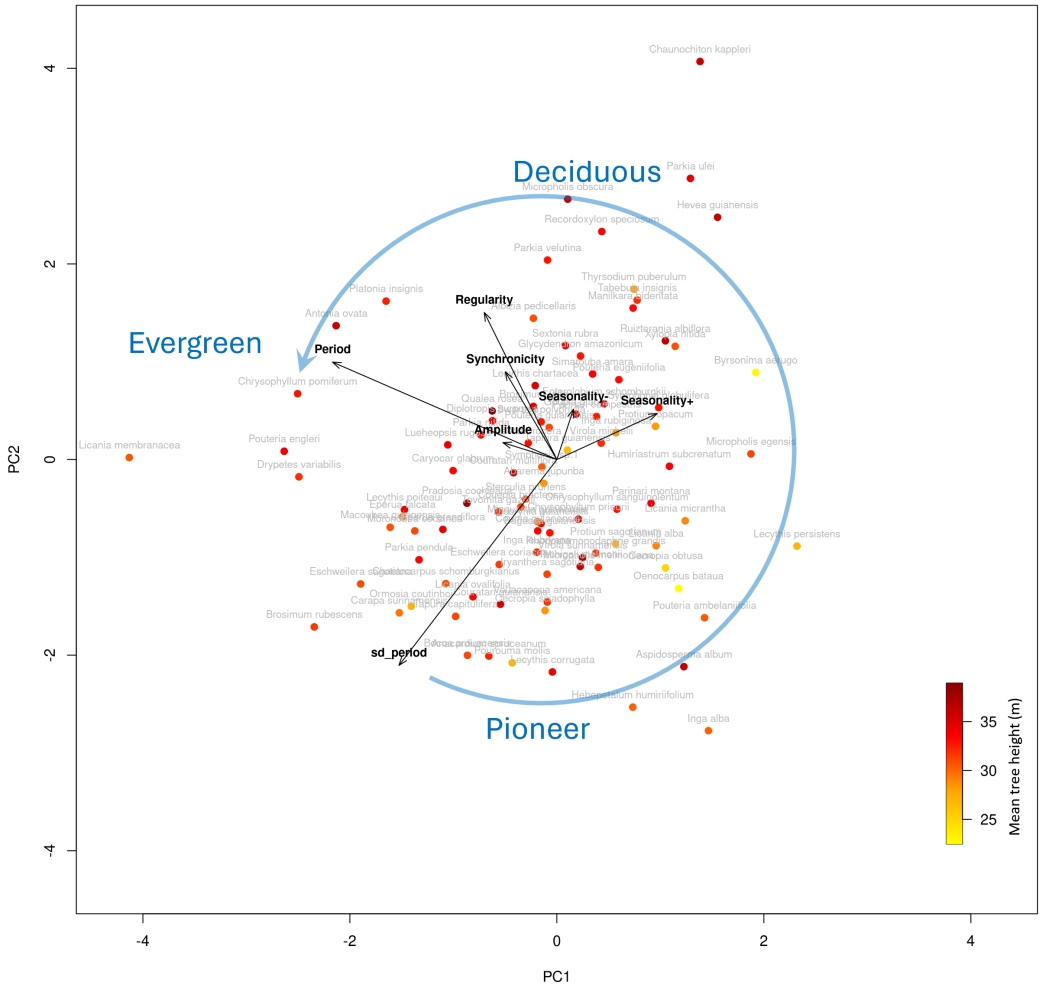


Fig. 5.7 Phylogenetic PCA (Revell, 2009) of the key phenological variables. *Synchronicity* is a measure of how well the signals of trees within a species match at any given point in time. *Regularity* is a measure of how similar the signal is between trees within a species (regardless of how well aligned they are in time). *Amplitude* is a measure of the magnitude of oscillation of the observed signal. *Seasonality+* is a measure of the tendency for individuals of a species to flush at the same time of the year. *Seasonality-* is a measure of the tendency for individuals of a species to shed leaves at the same time of the year. Flush/shedding timing is not included due to its circular nature. The blue arrow shows the hypothesized slow-fast resource acquisition strategy axis. Pioneer species continually grow and shed leaves with little regularity as they opportunistically race to fill gaps (Coley, 1983). The more "deciduous" (clear leaf of phases) trees employ a more structured phenological strategy of regular leaf/growth shedding to make the most of the seasonal changes in conditions. Shade-tolerant, slow-growing species exhibit longer periods with less obvious seasonality of leaf patterns (Chabot and Hicks, 1982; Sobrado, 1991; Wright et al., 2004). Species points are coloured by mean tree height. Short-lived pioneer species are unlikely to reach the same height as the deciduous or evergreen canopy trees.

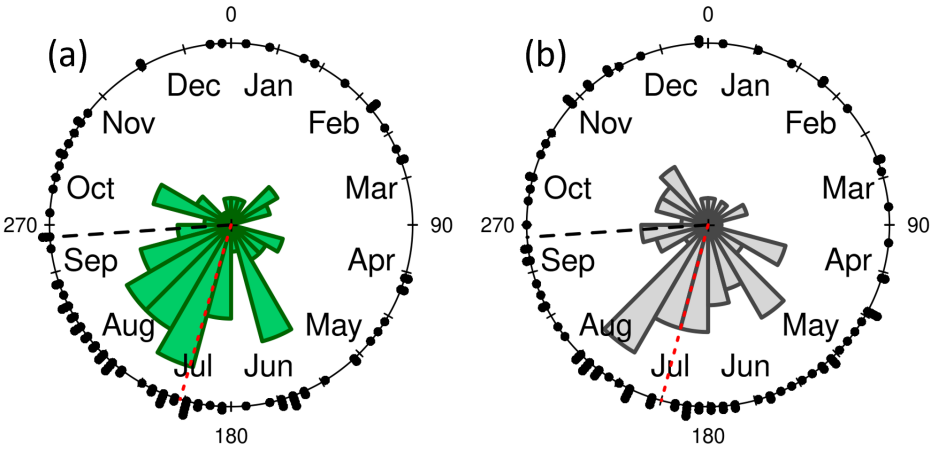


Fig. 5.8 Crown level seasonality of **(a)** flushing and **(b)** shedding as shown on a radial histogram and the black dots show the timings of each individual. The area of each segment is proportional to the frequency (number of observed crowns). The driest day is shown with a black dashed line and the mean day of peak/minimum of greening is shown with a red dotted line.

of individuals of this species were highly asynchronous. This species showed a two-year periodicity and the asynchrony derives from the fact it has two cohorts offset by approximately a year.

5.3.5 Phenological character with topographic position and tree height

To test the interaction of local topographic position (a proxy for water availability) and tree height we used mixed effects models at the individual level and phylogenetic least squares regression at the species level.

Interspecific variation

The average tree height and average TPI (an estimate of landscape habitat association; low = preference for gullies; high = preference for ridges) of a species, and the interaction between the two, tended to have a strong influence on the species level phenological metrics (see Table 5.2). The strength of the interaction term often led to a marked variation in the effect across the range of TPI and species height which required careful interpretation (see Fig. 5.12).

Short species tended to have their leaf flushing peak later (more positive) relative to the driest point of the season compared (with little influence of TPI association). As species topographic position association moves from (high TPI) ridges to (low TPI) gullies, peak flushing tends to occur earlier relative to the driest point (more negative). There is a significant interaction between species height and TPI association: towards gullies (low TPI), the timing of flushing is much more responsive to differences in species height than towards ridges where differences in species height has little effect of timing. Tree height had a similar effect on the

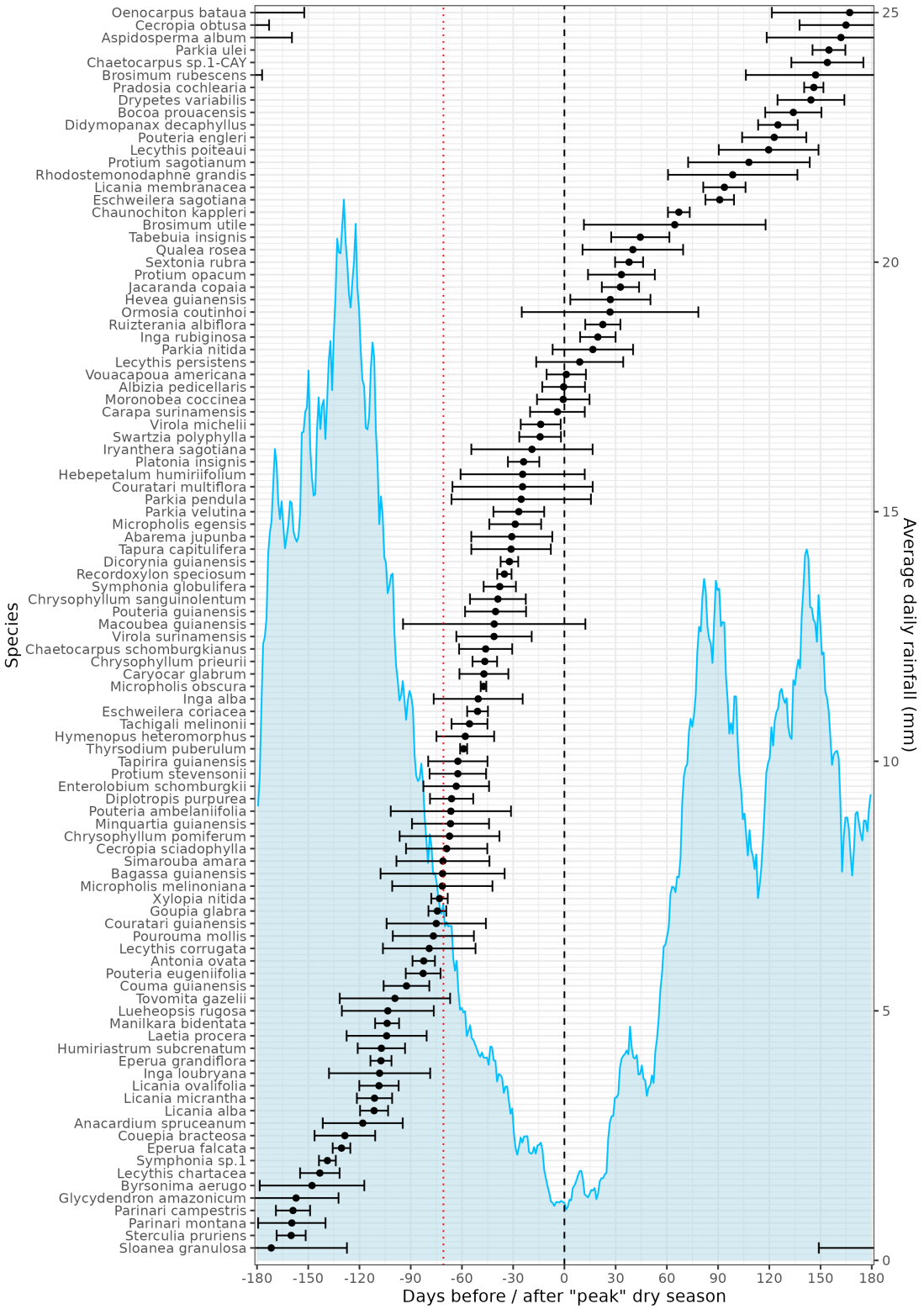


Fig. 5.9 The average timing of peak greenness (normalised GLI) for each species relative to the driest point of the year (20th of September) overlaid on the average daily rainfall at the site. Negative values are before the driest point and positive values are after. The dotted red line show the (circular) mean timing of peak greenness across all crowns.

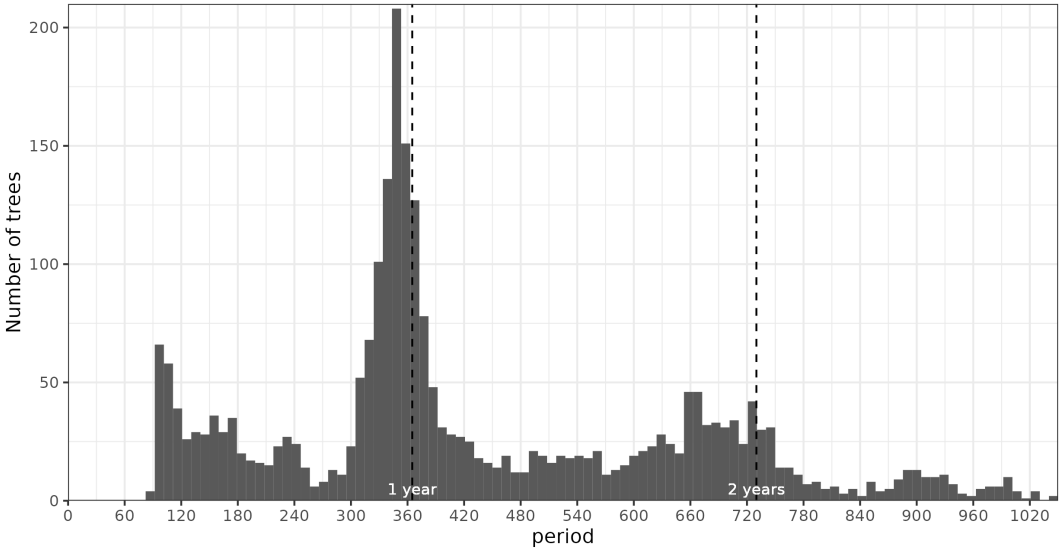


Fig. 5.10 Histogram of the period duration of the trees in the study. The period of each tree was assessed using the temporal autocorrelation function. Each tree's GLI signal was compared against a temporally lagged version of itself and the lag (greater than 90 days) that gave the maximum correlation was taken to be the leaf cycle period for that tree.

Table 5.2 Summary table of PGLS Regression Analysis testing the effect of species height and topographic position index (TPI) association on the key phenological metrics. The effect of mean species height, mean TPI and their interaction in each on the phenological metrics was tested.

	Mean height	Mean TPI	Height:TPI	R^2
Period variability	$-^{***}$	$+^{***}$	$+^{***}$	0.810
Peak timing	$-^{***}$	$+^{***}$	$+^{***}$	0.599
Min timing	$-^{***}$	$-$	$-^{***}$	0.596
Period	$-^{***}$	$+^{***}$	$+^{***}$	0.487
Amplitude	$-$	$+^{***}$	$+^{***}$	0.445
Seasonality+	$+^{***}$	$-$	$-^{***}$	0.393
Seasonality-	$-$	$-$	$-$	0.030
Synchronicity	$-$	$+^{***}$	$+$	0.190
Regularity	$-^{*}$	$+^{*}$	$+$	0.103

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

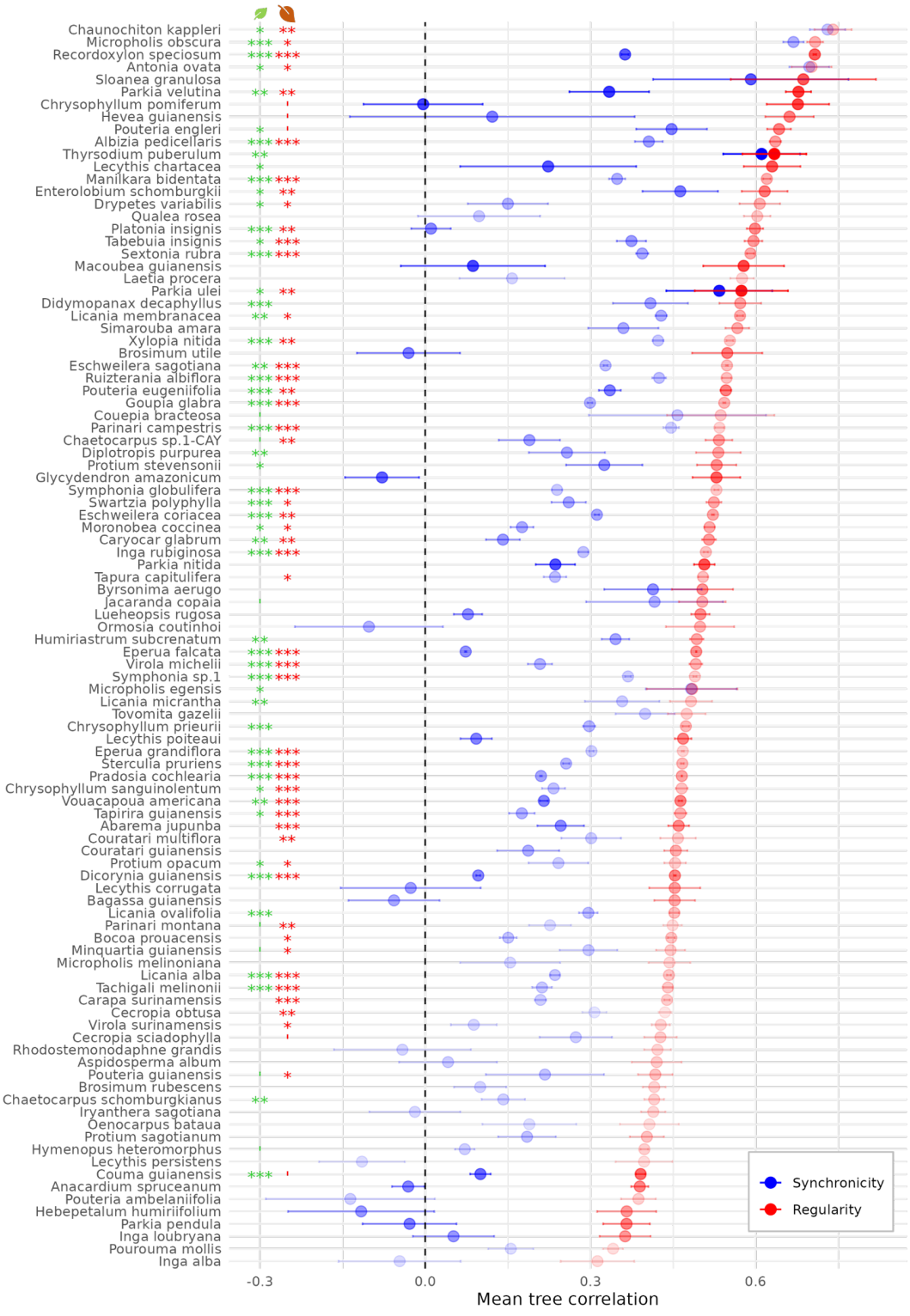


Fig. 5.11 Synchronicity and regularity of species where alpha (opacity) is scaled between 0-1 and represents the average amplitude of the GLI signals of individuals of a species (strong oscillation is shown as a stronger colour). The asterisks represent where there is statistically significant seasonality in flushing (green) and shedding (red) for a species. Significance codes for seasonality: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1' ' 1.

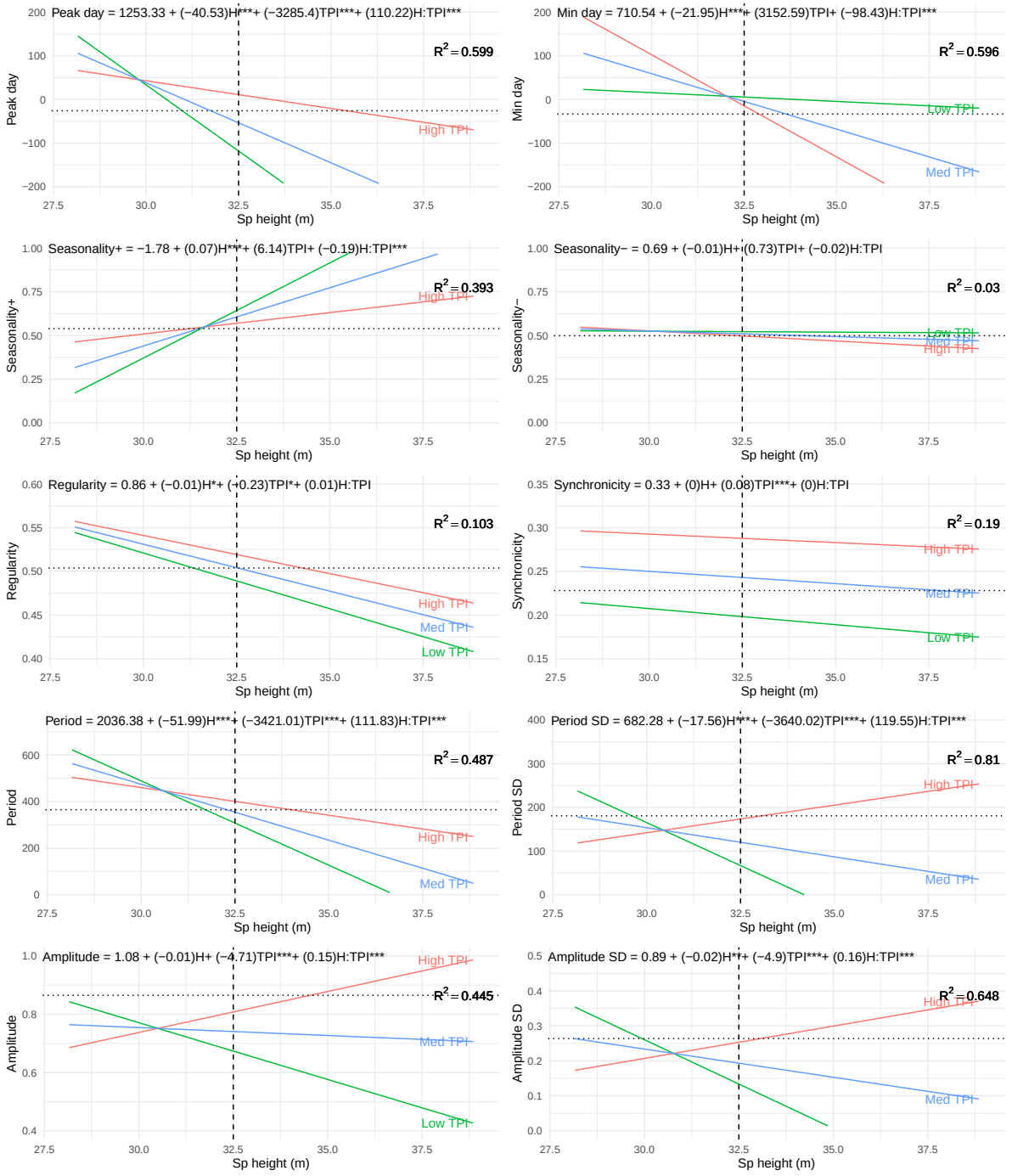


Fig. 5.12 The (inter-specific) influence of characteristic species tree height and TPI association on the phenological metrics based on Phylogenetic Generalized Least Squares (PGLS) regression. For each metric, three fixed TPI lines are plotted: low (mean-sd), medium (mean) and high (mean+sd). The equations of the best fit line are given with the significance codes next to those variables seen to have a significant influence. The significance symbols (p-values) are based on the scaled predictors whereas the plots (and quoted coefficients) are in terms of the native values. The R^2 values of the regression are given. *Peak/min day* is the timing of the peak emergence/shedding of fresh leaves relative to the driest moment of the year (negative values are prior and positive ones are after). The mean metric value is given as a horizontal dotted line.

timing of minimum greenness with taller species shedding leaves earlier however the effect of TPI association was inverted. The timing of shedding was much more responsive to differences in tree height for ridge associated species than gully associated species.

Taller species tend to have a shorter leaf cycle period and species with a high TPI association tend to have longer leaf cycle periods. The interaction between the two is again important: at low TPI association, period is more responsive to species height whereas at high TPI association the response in period is less pronounced with less deviation from the usual annual cycle. The variability of leaf phenology cycle period length had the greatest degree of its variance explained by these variables ($R^2 = 0.81$). Taller species tended to have a more stable leaf phenology cycle period length, but this effect is modified by TPI preference. As TPI increases, the leaf phenology cycle period tends to become more variable (which would appear inconsistent with its effect seen on synchronicity and regularity). For low TPI associated species, a small increase in mean height will cause a substantial drop in variability and for medium TPI species the drop in variability is less pronounced. For high TPI species, this trend is reversed and variability begins to increase with tree height.

The amplitude of phenological signal of a species generally increases with its TPI preference but the interaction with tree height is important. For tall species, an increase in TPI preference causes an increase in amplitude of signal but for medium trees this is less pronounced and for shorter species there is no apparent association. The seasonality of flushing of a species tends to increase with species height and the effect is strongest for low TPI associated species and weak for high TPI associated species. On the other hand, seasonality of shedding had no clear association with species height or TPI association.

The variation in species synchronicity and regularity were least well explained by TPI association or tree height. Synchronicity of signals was greater for trees with a high TPI association. Regularity was also higher for trees with a high TPI association but lower for taller species. This would appear inconsistent with the results for leaf cycle period variability.

Intra-species variation

Some intra-species variation in the phenological metrics was associated with TPI and tree height (see Table 5.3). Leaf cycle period tended to increase for trees with higher TPI. Flushing tended to occur earlier for taller trees; for each meter a tree grows it flushes roughly one day earlier in the season.

5.3.6 Phylogenetic structure of the phenological metrics

We tested each key phenological metric for phylogenetic signal using Pagel's λ and Blomberg's K tests. In general, the phenological metrics did not show a strong phylogenetic signal with the exception of amplitude. This meant the spectral difference of a crown of a species (between

Table 5.3 Summary table of tree level mixed effects models testing the influence of tree height and tree topographic position index on intra-species variation in phenological characteristics.

	Height	TPI	Height:TPI
Period	+	$+^{*}$	−
Amplitude	+	+	−
Peak timing	$-^{*}$	−	−
Min timing	+	−	+.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

shedding and fresh leaves) showed some similarity between more closely related species than more distantly related species. For context, the mean tree height of a species had a much stronger phylogenetic signal (λ = 0.692, p-value = 0.0007; K = 0.2168, p-value = 0.0028) than any of the phenological metrics.

5.4 Discussion

By combining UAV imagery with field-mapped crowns and machine learning, we were able to characterize detailed leaf phenological patterns across a moist tropical forest with a brief dry season in French Guiana. Other studies have looked at dry/seasonally dry patterns but this is the first to unpick what happens in a less water limited environment.

How much does canopy PAI vary seasonally? PAI varied through time by about 10%. Albeit small, this variation in PAI is likely to affect the carbon balance at the forest canopy scale. Indeed the reduction in PAI occurs in patches (via the defoliation of individual crowns). Therefore light interception efficiency at the whole canopy level is likely to decrease more significantly than if PAI was reduced homogeneously across all the crowns. In addition, light use efficiency will also be reduced as more light will be intercepted by non-photosynthetic material and because the lower canopy which is briefly exposed to higher-than-usual incoming light may not adapt readily. We expected that, as Paracou has a brief dry season and is unlikely to experience water limitation, the PAI of the forest would peak during the driest (and sunniest) time of year. We saw an increase in PAI that started prior to the dry season, increased through the dry season and reached its maximum towards the end of the dry season. This was consistent with the idea that non-water limited tropical forests put out new leaves to maximise light interception during the relatively cloud free periods (van Schaik et al., 1993; Wagner et al.,

Table 5.4 Phylogenetic signal in the key phenological variables with the associated p-values.

Variable	Pagel's λ	P	Blomberg's K	P
Synchronicity	0.0033	1	0.0901	0.1752
Regularity	0.0019	1	0.0230	0.8167
Lag	0	1	0.0747	0.3321
Period	0	1	0.4709	0.3527
Amplitude	0.2495	0.0252 *	0.1722	0.0596 .
Peak date†	0.1004	0.4197	0.0058	0.9391
Min date†	0.1782	0.1849	0.0057	0.9300
Flush seasonality	0	1	0.0154	0.3471
Shed Seasonality	0	1	0.0228	0.1388

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

†These were estimated using the circular approach described in Staggemeier et al. (2020) and so some caution should be shown when comparing directly with the other estimates.

2011; Wright and van Schaik, 1994; Yang et al., 2021). This direct measurement of canopy PAI is important as, while consensus is growing around the so-called Amazon green-up (in most places) during dry periods (Huete et al., 2006; Lopes et al., 2016; Myneni et al., 2007; Saleska et al., 2016; Samanta et al., 2010; Wang et al., 2020), there has been some dissent (Atkinson et al., 2011; Morton et al., 2014; Samanta et al., 2010; Silva et al., 2013), and a lack of *in situ* quantititative validation of variation in leaf amount. Optical satellite remote sensing infers LAI from spectral signals of the surface of the canopy. This is known to have issues, particularly with respect to saturation of signal (beyond a certain point additional leaves do not contribute to an increase in the optical signal) and artifacts related to imperfect geometric and radiometric corrections (including the bidirectional reflectance distribution function or BRDF) (Atkinson et al., 2011; Lyapustin et al., 2012). By directly measuring the extinction of laser pulses through the upper canopy we showed that there was a 10% seasonal swing in PAI. The variation in LAI would be greater (as as the relatively stable woody contribution would not be included) but, while methods are emerging to separate UAV lidar point clouds into leaf and wood components (Bai et al., 2023; Vicari et al., 2019), that was not within the scope of the current study. A simpler approach may be to assess the PAD difference between a fully leaved crown and a leafless crown but this is likely to be limited due to differences between species.

Does variation in canopy and crown greenness align with variation in PAI? Canopy green-up happened over a matter of weeks, just before the start of the dry season, as many crowns reached peak greenness coincidentally. The peak flushing (calculated as the circular average across all crowns) happened 71 days before the driest point of the year which was

followed by the number of crowns flushing decreasing gradually as the dry season progressed, until reaching a relative minimum shortly after the driest point of the year. This is consistent with our observations of PAI increasing through the dry season. We expected that the canopy would flush new leaves at the transition into the dry season as it prepared new leaves to make the most of increasing insolation (van Schaik et al., 1993; Wagner et al., 2017; Wright and van Schaik, 1994; Yang et al., 2021). At the crown level, where there was cyclical flushing/shedding, we showed that cyclical PAI and greenness followed closely aligned temporal cycles but within each cycle there was divergence as greenness declined gradually whereas leaf amount stayed relatively steady until a sudden drop off with shedding. This was in line with what we anticipated, as leaf spectral properties are known to change; in the visible domain, young leaves tend to be reddish / yellow towards light green as they approach their peak photosynthetic capacity, mature leaves are a darker green, and old/senescent leaves become brown/red as their Chlorophyll breaks down (Chavana-Bryant et al., 2019, 2017; Karageorgou and Manetas, 2006). There has been debate over whether overall leaf quantity (LAI) is the primary driver for observed photosynthetic seasonality, or whether age-related photosynthetic capacity is of primary importance (leaf *quality*) (Wu et al., 2016, 2017a,b). Separating the (spectral) signal of leaf aging from the (volumetric/areal) signal of leaf amount (Fig. 5.6) gives us insight into variation in tropical forest photosynthesis and productivity. The fact that we could discern between the two state variables holds promise for future, comprehensive large scale analyses that could corroborate or contest the findings of smaller scale phenocam based studies that have suggested "quality" is the more important driver (Wu et al., 2017b). Our study was conducted in the flux tower footprint of Paracou Research Station, with the scanning region carefully delineated/located based on atmospheric modelling, ensuring that all trees that contribute significantly to the fluxes received by the sensors on the tower are covered. The next step in the analysis will be to draw correlations between leaf age and leaf amount for the crowns in the flux tower footprint and the GPP as measured on the flux tower. Eventually, this will be parameterised within a mechanistic modelling framework such as TROLL (Chave, 1999; Maréchaux and Chave, 2017) which may in turn allow for improved representations of tropical forest phenology in dynamic global vegetation models (DGVMs) (Chen et al., 2020; Restrepo-Coupe et al., 2017). This ongoing work is discussed in Chapter 6.

Do species vary in leaf phenology? We quantitatively characterised the observed phenological patterns (and the inter- and intra- specific variation) extracting a set of key phenological metrics from the GLI signals that described the strength, timing and degree of agreement between signals. We anticipated that, because the dry season is brief and not severe, periods of water stress are unlikely in typical years, and species can therefore exhibit a wide range of leaf phenology patterns (Loubry, 1994). We thought it unlikely that any single habit would

be dominant. A broad range of phenological patterns were indeed exhibited, and they were somewhat structured around the opportunity for light presented in the dry season. For instance, the majority of species focused their leaf renewal (shedding and flushing) in the transition to the dry season but there was a lot of variation; there was no point in the year that was not occupied by a species' preference for either shedding or flushing. Some species were highly synchronous in their phenological patterns, flushing and shedding in near perfect unison, while others showed no sense of alignment in their patterns in agreement with some previous observations (Morel et al., 2015). The majority of species showed a periodic leaf renewal pattern that was approximately annual but some species showed a more short term cycling and others exhibited a periodicity of up to two years (e.g *Platonia insignis* and *Chrysophyllum pomiferum*). Species that had a two-year periodicity scored low in terms of synchronicity and seasonality but this was because their individuals were split between two cohorts offset by a year. Within these cohorts there appeared to be good alignment but our approach was unable to identify this. Around half of the species showed a statistically significant degree of seasonality (i.e. likely to be focused around a specific point in the year rather than uniformly distributed through the year) in their flushing and their shedding (although not necessarily the same ones for each). Those species with very regular signals (i.e. similar waveforms), were likely to be highly synchronised and have a strong amplitude in their signal although there were several exceptions (e.g. *Chrysophyllum pomiferum* as mentioned above). Additional methods that could cluster species with longer cycles into cohorts could be informative for describing species strategy.

Do any differences observed among species relate to their resource acquisition strategy? While no clear groupings of strategies emerged, it seemed that species phenological patterns fell along a continuum that could be related back to resource acquisition strategy. We had expected to see three groups: (1) Fast-growing pioneer with little cyclical predictable behaviour, (2) Strongly seasonal brevi-deciduous / seasonal evergreens, (3) Evergreen, asynchronous (Reich, 1995). The phylogenetic PCA was plotted with species coloured by its mean tree height. It seemed that pioneer species, which were likely to be shorter due to shorter life spans, flushed and shed with little regularity or structure (and high variability of periodicity) as they opportunistically raced to fill gaps. The more "deciduous"/seasonal trees employed a structured phenological strategy of regular leaf/growth shedding to make the most of the seasonal changes in conditions. Shade-tolerant, slow-growing species exhibit longer phenological periods with less obvious seasonality of leaf patterns. A comparison to the successional status of the trees could help to substantiate this.

Key to thinking about leaf habits are the leaf and plant economic spectra (Li et al., 2022; Reich, 2014; Reich et al., 1997; Wang et al., 2023; Wright et al., 2004; Zhao et al., 2017).

It is generally assumed that deciduous (and to a greater extent pioneer) species employ a more acquisitive strategy with short-lived leaves that are more photosynthetically active and incur lower construction and maintenance costs than the longer lived leaves of shade-tolerant, conservative evergreen species (Eamus, 1999; Reich, 1995; Ribeiro et al., 2022; Sobrado, 1991). In forests that experience a pronounced dry period, deciduous species must maximise resource capture during their limited growing seasons and display higher gas exchange capacity, assimilation rate, and water use efficiency compared to their evergreen counterparts (Ribeiro et al., 2022). In these conditions, leaf habits are intricately linked to water management and drought tolerance (with alignment to other traits) (de Souza et al., 2020; Méndez-Alonzo et al., 2013; Sastry and Barua, 2017). Deciduous trees can shed their leaves to prevent water loss whereas evergreen trees must be equipped to survive water stress with leaves intact which leads to divergence in approaches (Kaproth et al., 2023; Sakschewski et al., 2021). For example, evergreen species become more embolism resistant as rainfall decreases but the embolism resistance in deciduous species does not respond to rainfall (Oliveira et al., 2021). This influences distributions, as leaf turgor loss point shapes local and regional distributions of evergreen but not deciduous tropical trees (Kunert et al., 2021), and types of drought resistance, as deciduous trees are more sensitive to both air and soil drought, whereas evergreen trees are only sensitive to soil drought (de Souza et al., 2020). Thermo-tolerance, which is greater in evergreen species than deciduous species, may enforce another axis upon which adaptive strategy is focused (Sastry and Barua, 2017). Within deciduous species, variation arises which is linked to the density of their wood; hardwood trees have lower stem water storage capacity which leads to leaf patterns that are more reactive to water availability than softwood deciduous species (Lima et al., 2021). However, in systems that do not experience regular water stress such as wet forests the source of variation in leaf phenology, and its link to resource acquisition strategy, is less well described/defined. From what we have observed, there is a mechanism for shedding/flushing leaves that, in general, aligns with maximisation of light interception. Many of the brevi-deciduous or strongly seasonal evergreen trees appear to use shedding/flush to renew their leaves for the dry/light season to exploit/capitalise on fresh, efficient photosynthetic machinery ready for their chance to intercept light and grow (without any clear/sustained periods of dormancy) (to out-compete their neighbours). Is this mechanism phenomenologically/physiologically similar to the mechanism employed by deciduous trees in drier regions that shed to enter a dormant/protective period? I.e., is the shedding mechanism taking cues from moisture availability but, in this case, as a proxy to thereby optimise for (the strongly covarying) availability of light? Or is the shedding mechanism in light-optimised forest phenomenologically distinct, taking a more direct light associated cue? More work needs to be done to compare suites of (leaf and wood) traits and local habitat preferences (soil, topology)

to leaf phenological metrics to disentangle and categorise the morass of leaf phenology patterns observed in non-water limited, seasonally light areas (Ouédraogo et al., 2016).

Do topographic position and tree height help to explain inter- and intra-specific variation in leaf phenology? Interspecific variation seemed to be explained (partially but significantly) by local topographic association and characteristic species height. Tall trees can make the most of light availability but are typically more hydrologically sensitive as they are exposed to warmer and drier air which increases evaporative demand and they need to transport water further upwards against the force of gravity (Midgley, 2003; Ryan and Yoder, 1997). Lowlying (gully) sites are generally wetter and less likely to experience drought conditions.

In general, shorter species showed little clear variation in phenological character (perhaps due to the necessity for shade tolerance and minimal risk from hydraulic stress) but the patterns of taller species diverged based on their local topographic habitat preference (gullies through to ridges). Taller species flushed fresh leaves earlier with respect to the dry season, an effect that was strongest for species associated with gully location (low TPI) but not apparent for species associated with ridges (high TPI). This suggested that emergent species that were unlikely to experience water stress were more able to maximise light interception by presenting efficient leaves throughout the dry season than emergent trees in drier conditions that were more constrained by the risk of hydraulic failure. The strength and synchronicity of the leaf cycling signals were greater for ridge species compared to species that were associated with gullies, suggesting species associated with ridges had greater sensitivity to environmental cues. Taller species tended to have shorter leaf cycle periods potentially linked to a faster resource acquisition strategy required to emerge beyond the canopy. This relationship is unlikely to hold across all tree heights as the fastest growing species are short lived (pioneer) but this could not be captured in our linear analysis. Additionally, leaf lifespan is highly plastic and can be variable across a tree's life (Laurans et al., 2012) and emergent trees are not necessarily fast growers. Examining phenological patterns in relation to growth rates (both in diameter and height) would help to further elucidate this relationship. Intraspecific variation in timing was associated with tree height; the taller the tree, the earlier it was likely to flush. Intraspecific variation in leaf cycle period was greatest at high TPI, potentially linked to a facultative exchange strategy.

Do the observed leaf phenological metrics of species exhibit a phylogenetic signal? In general, we found that the characteristics of leaf phenology did not exhibit phylogenetic signal with the exception of the amplitude of the waveforms which corresponded to the magnitude of spectral change in the crown between peak flushing and shedding. We anticipated that the leaf phenology characteristics would not have clear phylogenetic signal in agreement with Pau et al. (2011)'s predictions for the tropics. The amplitude relates to how close to complete leaflessness

a species is when it has minimal leaf cover (and potentially bark spectral properties) but also the "greenness" of its young leaves. It is possible that there is phylogenetic signal present in the axis that goes from deciduous habit to strongly seasonal evergreen to weakly/aseasonal evergreen but this could be confounded with the phylogenetic signal present in spectral properties of young leaves. Young leaves are vulnerable to herbivory and leaf defensive traits, especially among young leaves, tend to be on the more conservative end of leaf traits (Endara et al., 2017; Kursar et al., 2009; Uckele et al., 2021). If these traits (e.g. presence of secondary metabolites / allelochemicals) translate into spectral differentiation this may exaggerate the signal. A comparison against the lidar variation will be performed to help partition these effects.

There are several reasons why phenological metrics might have a low phylogenetic signal in a moist tropical system. Convergent evolution due to environmental adaptation is highly unlikely as there is such a range of realised strategies. Phenotypic plasticity may explain some but interspecific variability in itself covered the phenological and environmental variable space. The capacity to enact rapid (in evolutionary terms) changes may be preferable in these dynamic, highly populated, energy rich systems leading to highly labile phenological traits. See for example Parkia that varies considerably in its timing and synchronicity/regularity (Morel et al., 2015). As we have discussed, leaf phenology is tightly linked to water and light availability so evolutionary pressure may lead to the ability of clades to evolve quickly into new leafing patterns to exploit dynamically opening niches.

Conclusions. Leaf phenology is known to have a strong influence on carbon and water cycling but due to the diversity of patterns it has previously been difficult to comprehensively characterise in the wet tropics (Zou et al., 2023). The work presented here presents methods to assess and characterise the broad range of behaviours. The data could be used to parameterise variations in leaf area index and photosynthetic capacity in vegetation models used to understand tropical forest productivity under climate change (Chen et al., 2020). By comparing the phenological patterns we observed to flux tower measurements we will further elucidate the links between leaf amount, leaf age / quality and forest primary productivity (Wu et al., 2016, 2017a,b). The methods presented here can be employed at other sites across the biome to broaden our understanding of leaf phenology variation.

Chapter 6

Conclusions: Ecological and technological change

Abstract

In this thesis, I have applied a range of remote sensing technologies and deep learning (computer vision) approaches to understand the dynamics of tropical forests from the individual tree to the landscape scale. I have shown that computer vision approaches can generate insights on fine scale changes in complex tropical forest landscapes: convolutional neural networks can be trained to map tree crowns across large forest extents and predict the spread of forest loss from freely available remote sensing and geospatial data. The individual tree mapping helped to demonstrate that growth declines with tree height and that taller trees have higher mortality than intermediate sized trees. Segmentation of multitemporal imagery was shown to further improve the accuracy of the mapping process, and the integration of hyperspectral data allowed tree species accounting for 70% of the total tree crown area to be mapped accurately. The species mapping was a key step in the process of characterising diversity of leaf phenological patterns at a moist tropical forest site. I found that species height and topographic position preference explained much of the variability in leaf phenology: specificially, I showed that taller trees found in moister locations flushed fresh leaves well in advance of the dry season, possibly to maximise interception of insolation in locations where water was not limiting. Together, these results demonstrate that deep learning methods can be effective for making ecological inference from remote sensing data, and emphasise the value of integrating these approaches into studies of tropical forest dynamics. In this Chapter, I place the findings in a broader context, and highlight the opportunities for continued development. I discuss the rapid development of deep learning techniques and how they can be effectively transferred into ecological remote sensing.

6.1 Recent advances in computer vision

The field of deep learning is moving at an incredible pace with new approaches continually emerging. We did not use transformer networks in the research described in this thesis but they are becoming increasingly popular and it is worth considering what role they will play in the remote sensing of tropical forests and the development of work described in this thesis. Transformer networks signalled a particularly significant advancement, providing advantages

over recurrent neural network architectures such as LSTMs when processing sequential data, particularly in natural language processing (Vaswani et al., 2017). Transformers process entire sequences in parallel, unlike sequential processing by LSTMs, enhancing efficiency in handling large datasets. Their self-attention mechanism captures relationships between sequence elements, regardless of distance, aiding in understanding entire sequences. Highly scalable, transformers perform well in large-scale datasets and complex tasks, showing flexibility across domains including natural language processing (Vaswani et al., 2017), computer vision (Dosovitskiy et al., 2020), and protein structure prediction (Jumper et al., 2021). In contrast, LSTMs, while designed for long-term dependencies, face challenges with long sequences due to the vanishing gradient problem (Noh, 2021). Transformers avoid this issue with better gradient flow, making them suitable for deeper model training. Their architecture is also simpler than LSTMs, lacking complex gating mechanisms. The Vision Transformer (ViT) applies this approach to computer vision, breaking down images into patches for processing (Dosovitskiy et al., 2020). While ViTs excel in capturing global context in images and scale well with larger datasets and compute resources, they have limitations, such as high computational costs and reliance on large quantities of training data. Convolutional Neural Networks (CNNs), with their inductive biases, remain efficient for smaller datasets. The field is witnessing a trend of combining both architectures, leveraging CNNs for feature extraction and ViTs for capturing global dependencies (Mauricio et al., 2023; Zhang et al., 2021). Recent developments in CNNs, like InternImage-H's deformable convolution (Wang et al., 2023), highlight their continued relevance. This model recently claimed the top spot on two of the COCO instance segmentation benchmarks[1]. In domains such as forest ecology, where data are often limited, CNNs might maintain an advantage. The future of image processing lies in leveraging the strengths of both ViTs and CNNs, depending on task requirements and available data.

Integrating state-of-the-art approaches, such as transformer networks, into the field of ecological remote sensing presents several significant challenges. Approaches that have been established for a number of years, such as U-Net (Ronneberger et al., 2015) and Mask R-CNN (He et al., 2017), often have a more straightforward integration process into existing workflows, supported by extensive and clear documentation, which is not true of the newest techniques. Additionally, ecology typically deals with datasets that are smaller and noisier compared to well-curated benchmarks like COCO, posing challenges for models that require large, clean datasets for optimal performance. Furthermore, ecologists are often interested in features of remote sensing data that are not typically of interest to other users (such as those that can help to distinguish between tree species), requiring tailored approaches for effective data processing and analysis. To an ever growing degree, collaborations between

[1] Since November 2022 and last accessed 15-11-2023: https://paperswithcode.com/task/instance-segmentation

computer scientists and ecologists are needed to apply the most effective approaches to pressing ecological questions.

6.2 Forecasting future deforestation

CNNs showed an exciting capacity to automatically extract landscape features relating to different drivers of forest loss and infer whether loss was likely to continue from a source. The work highlights the potential for deep learning approaches to make valuable contributions to preventing forest loss in line with the Glasgow pledge at COP26. Ever increasing volumes of remote sensing data are becoming freely available and approaches such as those described in Chapter 2 (Ball et al., 2022) have the capacity to turn them into valuable insights. Previous attempts at forecasting deforestation have been limited due to the need to engineer features upon which to predict, which is often unreliable for poorly mapped, dynamic landscape features. From satellite data and other complementary spatial datasets, our model appeared able to discriminate between natural forest loss that was unlikely to spread from anthropogenic deforestation frontiers that are likely to progress over time (e.g. unauthorised gold mines or newly built roads). However, further work is needed to probe how the "black box" arrives at its decisions (Carter et al., 2019; Li et al., 2021b). Including additional remote sensing layers as predictors could provide extra features for the network to learn from that would be otherwise "hidden" in the multispectral data. Night lights, which provide a dynamic signal for human presence and economic activity (Levin et al., 2020), could be particularly informative. Higher resolution imagery (e.g. Planet NICFI, Sentinel-2) may make that feasible to detect more localised disturbance (e.g. selective logging). The Copernicus Global Land Cover collection (Buchhorn et al., 2020) could provide labels for more types of land cover transitions which could help to differentiate between the different kinds of forest loss, although it is only available at 100 m resolution. Automatic classification of drivers has been demonstrated at coarse resolutions (10 km) with simple decision tree models applied to derived RS products (Curtis et al., 2018). Given the apparent ability shown here to discern between natural and human cause loss, it is likely that a network could be trained to classify and map drivers directly from satellite imagery at high resolution. However, a comprehensive training set of labelled deforestation drivers at this resolution does not exist and would require a considerable investment to produce.

Our machine learning approach assumes that historic trends will continue into the near future. It does not account for shifts in political regimes that could have substantial impacts on deforestation (Pereira et al., 2020). A changing climate, volatile international economic conditions and advances in technology further challenge the assumption of stationarity. However, as the forecasting system can be dynamically updated, the approach is able to learn

emerging patterns. Evaluating relative risk across pixels makes it possible to prioritise areas for intervention based on a given strategy or budget. The forecasts could also direct interventions to emerging frontiers of loss in otherwise isolated and intact landscapes, such as those caused by illegal mining activities. The approach could potentially be scalable to forests globally. Governmental and non-governmental organisations could use the tools to understand how the landscapes in which they operate are likely to change in the near future and refine their protocols for managing and responding to deforestation risk.

The ability of 3D CNNs to view a pixel's spatial and temporal context simultaneously has been shown to improve classification accuracy of land-use classification from remote sensing data, compared to 2D CNNs (Ji et al., 2018; Xu et al., 2018). One would expect this to be important for accurate forecasting. However, the 3D approach had fewer individual data points available than the 2D approach, due to the way data were stacked. The 2D CNN appeared more robust in cloudier and heterogeneous locations. We tried to use an LSTM RNN approach to deal with the temporal progression in the data but we found that it was not possible to get these models to exceed the accuracy of the 2D and 3D CNN approaches. This was in part due to its inability to be paralellised and make full use of the computing resources available to us. However, the recent innovations in transformer networks (see Section 6.1) could have provided greater accuracy and efficiency in handling the sequential aspect of the data. Using self-attention, they avoid the difficulties in parallelising RNNs, and so can be trained more efficiently. They will become more effective as the length of the available time series grows and could help to extend the forecasting horizon.

6.3 Locating and delineating individual trees

We showed that Mask R-CNN was effective at delineating tree crowns in tropical forest landscapes and that accuracy increased with tree height, meaning that tall trees which store most carbon are most reliably delineated (Ball et al., 2023). We found that the method could be quickly trained to perform well on new areas of forest using around ten ~1-hectare images within which all visible tree crowns had been manually delineated. Our approach (*detectree2*) has since been shown to outperform the previous state-of-the art algorithms for tree detection (*DeepForest*) with the additional benefit that it can generate accurate estimates of tree crown area from its precise delineation (Gan et al., 2023). *Detectree2*, available on GitHub as an open-source Python packages, has also been incorporated into tree inventory pipelines (Nieding, 2023; Troles et al., 2023). We went on to demonstrate that combining information from scans at different dates can considerably improve delineation accuracy (Chapter 4).

Using single-date delineations, the well-differentiated tree crowns at Sepilok West and Danum in Malaysian Sabah were the easiest to delineate while the even height and tight packing of upper canopy trees in Paracou resulted in lower accuracy delineation. Furthermore, at Paracou it is common to observe crowns mixing and growing into each other, which makes visual separation of crowns challenging. The best performing models were those that were exposed to training data from all the sites and then "honed" with a limited number of training iterations on the site to be predicted on, suggesting that our trained models could be transferred to a new site with little manual data.

Further improvements in accuracy could be achieved by integrating additional data layers beyond RGB imagery from one time. Firstly, multispectral imagery with NIR bands could help to distinguish crowns that are difficult to separate in the visible spectrum, but a disadvantage of introducing this band is that large pre-trained models based on RGB data could no longer be transferred. Secondly, traditional approaches to tree crown delineation often use a variation of the watershed algorithm applied to a Digital Surface Models constructed from lidar point clouds or photogrammetry; Ene et al., 2012); a surface model is a standard output as part of the mosaicking of UAV imagery using photogrammetry, and it would be possible to stack an additional band representing the canopy surface onto the RGB data, to give the network additional clues to help it separate crowns. Thirdly, in our Python package we have implemented a feature that allows multiclass prediction (as opposed to just the single 'tree class). We have started with liana infestation detection in which we identify crowns with and without liana presence (see Fig. 6.1) and also try to grade the severity of infestations. This work is currently being drafted as a paper. *Detectree2* could also be implemented for species detection in low diversity systems and tree health mapping and is available to researchers to implement on their own questions and data. Finally, as I have shown in Chapter 4, delineating crown in RGB imagery collected on different dates, and then evaluating consistency of delineation, proved highly effective at providing high-quality information for use in species detection algorithms.

Tracking growth and mortality

Although in theory repeat surveys can be used to assess the growth of trees, we were cautious in our interpretation of data (Chapter 3). We could be confident that across all sites taller trees had higher mortality rates and lower growth rates, which is consistent with the finding of field-based studies (Iida et al., 2014). However, the apparent higher growth and mortality rates in French Guiana as compared to the sites in Malaysia was potentially a result of biases introduced by difference in scan parameters (flight height, pulse density, time of year) among sites, and we resisted the temptation to make direct comparisions across sites (Vincent et al., 2023). Recent improvements, in lidar inter-calibration (between flights and sensor parameters; Vincent et al., 2023) may give the consistency required to make comparative

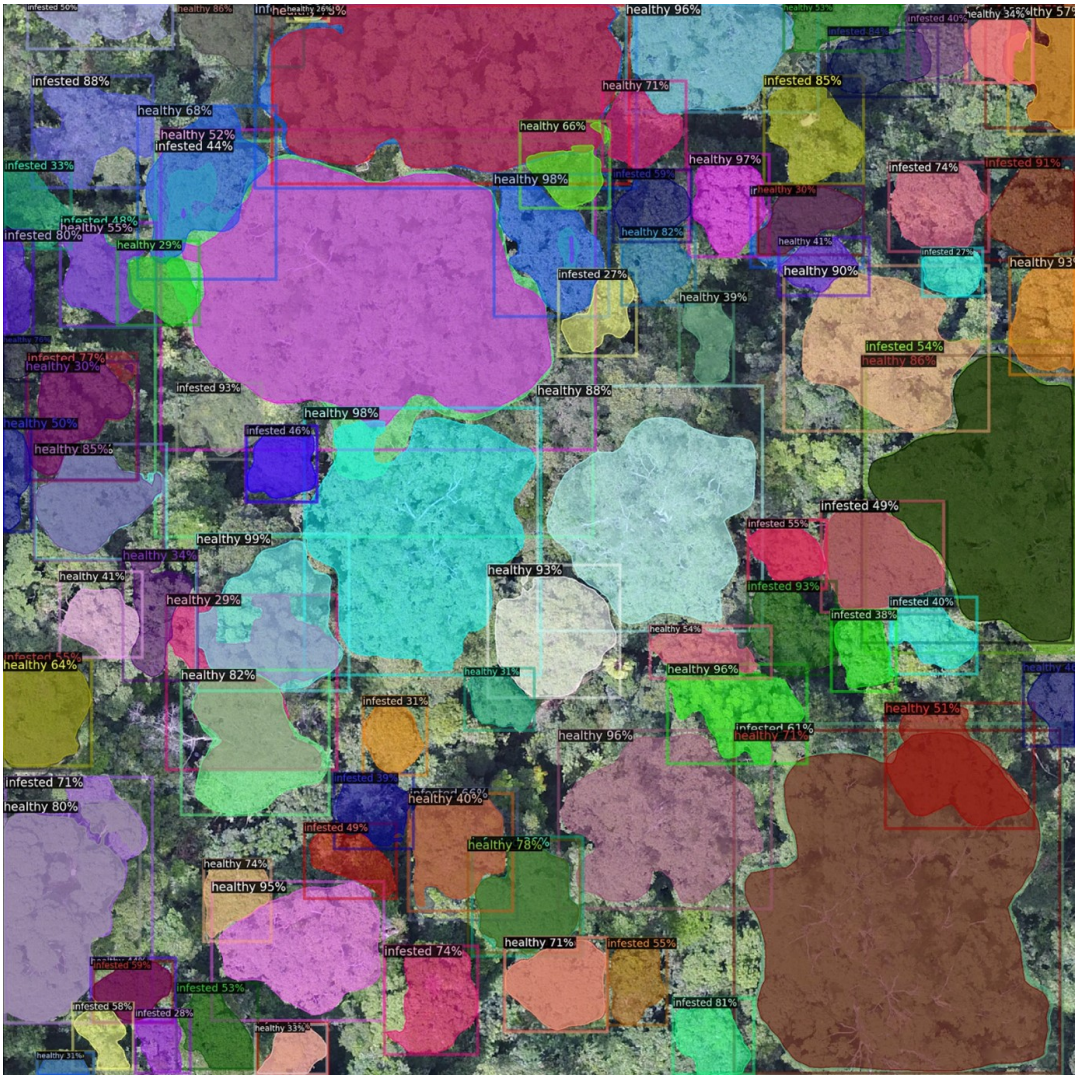


Fig. 6.1 Multiclass tree crown delineation (instance segmentation) with *detectree2* classifying crowns as either infested or free from lianas.

assessments of forest growth, but this was beyond the scope of the current study. Inventory data shows that mean stem diameter growth for trees at Paracou was 1.2 mm/yr (Wagner et al., 2010) compared to 0.9 mm/yr in Sepilok East, 1.1 mm/yr in Sepilok West and 0.5 mm/yr in Danum (Ordway et al., 2022; Piponiot et al., 2022) which was not consistent with the difference in observed height growth rates. We note that these stem diameter growth rates are not directly comparable to height growth measured in this thesis. Another caveat is that we defined mortality as a drop in height of more than a statistically determined threshold, but did not verify that a tree had actually died using field data.

6.4 Mapping tree species across diverse tropical forest landscapes

In Chapter 4, our classification of species using hyperspectral imagery extended the number of species that could be mapped and the completeness of coverage within a diverse tropical forest landscape. Approximately, 70% of tree crown area at landscape scale was accurately mapped with 64 species predicted from the hyperspectral data with F_1-score > 0.7. This mapping provided an expanded sample of trees on which the phenology could be analysed in Chapter 5.

Why was the upper red edge important for discriminating species?

Information about leaf chemical and physical characteristics are embedded within spectral data in complex ways (Féret and Asner, 2011; Jacquemoud and Baret, 1990) because multiple traits may influence the same spectral regions (Curran, 1989). We found bands in the 748 to 775 nm (upper red edge) range were by far the most important bands for discriminating species, which was at odds with the finding of Laybros et al. (2019) but align with a previous assessment of band importance when assessing canopy taxonomic diversity at the same site Badourdine et al. (2023). The "red edge" is typically identified as lying within the 700-750 nm range, and contains information used to predict as chlorophyll concentration, leaf size, water content, and general plant health, as well as to classify vegetation (Boochs et al., 1990; Filella and Penuelas, 1994; Gitelson et al., 2003; Hennessy et al., 2020; Horler et al., 1983; Thomas and Gausman, 1977). Our research indicates that the spectral region just beyond the red edge, as classically defined, exhibits particular sensitivity to species variations in the tropical forest we surveyed. This "upper red edge" lies at the intersection between two regions, one in which red light is absorbed by chlorophyll and the other in which near-infrared (NIR) is scattered by cellular structures, potentially providing insights into both biochemical and structural facets of vegetation. Wavelengths adjacent to this range do not exhibit significant feature importance, suggesting that the unique reflectance found in the 748-775 nm range might

capture distinct anatomical characteristics of species, likely pertaining to internal leaf structures that influence the scattering of NIR light (Ustin et al., 2009). It is not obvious why these bands came through as overwhelmingly important. A comparable analysis of data obtained by leaf spectroscopy, accompanied by an analysis of anatomical traits, may help to shed light on this problem. Perhaps the most perplexing result from Chapter 4 relates to phylogenetic signals. Most wavebands exhibited a significant phylogenetic signal, but the bands with the strongest signal were the ones that had least importance when identifying species. Does this mean that labile traits are more helpful for differentiating between species? Looking within lineages instead of across the whole phylogeny may be a way to probe this relationship further (Meireles et al., 2020).

6.4.1 Further improvements in species classification?

We found that linear discrimination analysis, among the more basic classification approaches, performed as well, or better than advanced approaches such as a multilayer perceptron. More work is needed to understand how best to constrain the more advanced methods so they can learn to encode the key crown-level features from limited training data. Scarcity of labelled data relative to unlabelled data across a landscape lends itself to semi-supervised approaches. I was involved in a paper using a graph-regularized neural network (GRNN) algorithm for tree species classification. The proposed algorithm encompasses superpixel-based segmentation for graph construction, a pixel-wise neural network classifier, and the label propagation technique to generate an accurate and realistic (emulating tree crowns) classification map on a sparsely annotated data set (Bandyopadhyay et al., 2022).

Challenges for consistency

A major concern about species classification from hyperspectral imagery is that models generated at one time and place are not generalisable to other times or regions. A central problem is that the spectral signatures of individual trees change over time due to factors including leaf phenology and external factors such as pests or drought (Chen et al., 2022b; Hesketh and Sánchez-Azofeifa, 2012). Moreover, even within the same species, trees can display different spectral signatures based on location-specific factors such as soil type, local climate, structure and topography (Knyazikhin et al., 2013). The spectral data captured by airborne sensors are also influenced by atmospheric conditions, including cloud cover, gas composition, and airborne particles (Arroyo-Mora et al., 2021; Schläpfer et al., 2018; Theiler et al., 2019). These atmospheric variations, occurring between the sensor and the Earth's surface, can alter the spectral signatures, thereby complicating the direct comparison of hyperspectral data from different times and places (Theiler et al., 2019). Additionally, the spectral responses may vary between sensors, and even the same sensor can show differences over time due to

calibration shifts or degradation (Baumgartner et al., 2012). Another influencing factor is the illumination conditions (Arroyo-Mora et al., 2021; Schläpfer et al., 2018; Theiler et al., 2019) and viewing geometry, which include the angle of sunlight and the sensor viewing angle (Duthoit et al., 2008; Lyapustin et al., 2012; Montes and Ureña, 2012; Schläpfer and Richter, 2014; Schläpfer et al., 2015; Theiler et al., 2019). These aspects add further variability and necessitate more complex data processing and analysis methods. Laybros et al. (2019) highlights these issues demonstrating a 10% drop in tree species classification accuracy when models are trained and tested on data separated by just a single day (as compared to those trained and tested on the same scan). As of November 2023, a new hyperspectral scan is available for Paracou. Because of the high costs associated with hyperspectral sensing in remote tropical regions, very few multi-date datasets of this sort are available. By comparing the new scan to the one from 2016, we will be better able identify the features of the data that are stable through time allowing for more consistent species mapping.

6.4.2 The value of carefully collected field data

The field dataset from the Paracou field site in French Guiana was developed, curated and ground validated over a number of years. The careful mapping of what was observed from above to the reality on the ground was labour intensive but the resultant database of 3600 crowns provided a robust foundation upon which the research in Chapters 3-5 was built. Without substantial, high quality ground datasets, remote sensing cannot provide insight into biological processes (Cavender-Bares et al., 2022; Davies et al., 2021). Despite this, there are few studies that provide benchmark data upon which tropical tree species mapping approaches can be tested (Laliberté et al., 2020). The dataset in French Guiana is being expanded: 2500 additional crowns are being mapped in Nouragues, and these sites also have repeat-surveyed hyperspectral data. A consortium approach is needed to bring together more researchers to collect similar data across multiple tropical sites, to advance the generality of hyperspectral modelling of species.

Scaling up biodiversity mapping using space-borne hyperspectral imagery

The EnMAP (Environmental Mapping and Analysis Program) satellite, a German-led mission launched in 2022, is equipped with a hyperspectral sensor with 230 spectral channels from 420 nm to 2450 nm, spanning visible, near-infrared, and shortwave infrared regions (Guanter et al., 2015). The 30m resolution imagery permits detailed analysis of Earth's surface features, although it falls short of having the spatial resolution needed for individual tree crown analysis. However, EnMAP operates on a 27-day repeat cycle, which can be reduced to four days for a broader swath, utilizing $\pm 5°$ and $\pm 30°$ off-nadir tilts, respectively allowing for regular repeat surveys. The high spectral resolution of EnMAP should allow it to detect subtle changes in

vegetation but the extent to which this will translate into species or biodiversity mapping in the tropics is yet to be seen. Hyperspectral data requires difficult, often imperfect, atmospheric correction when retrieved from an aeroplane which can have a big impact on species classification. These issues are likely to be exaggerated further for a space-based sensor taking reading over high-rainfall regions. The data from EnMAP has been freely available to users since November 2, 2022 but there are not yet any publications using the data to study tropical vegetation.

The detailed analyses at specific sites such as Paracou may prove helpful in calibrating sensors such as EnMAP. Echoing Chave et al. (2019)'s suggestion for biomass remote sensing, a collection of comprehensively ground surveyed 'supersites' could help to bridge the gap between the rich sources of data from space and insights related to biodiversity and tropical phenology. Barro Colorado Island (BCI) in Panama is an obvious candidate given its rich history of research and detailed multi-temporal mapping, but the sites must be representative of the tropics in their entirety. The Global South in general, and Sub-Saharan Africa in particular, are chronically underrepresented when it comes to detailed, multi-temporal vegetation surveys that can be linked to other monitoring infrastructure such as flux towers (Hortal et al., 2015; Pastorello et al., 2020; Stephenson and Stengel, 2020; Tydecks et al., 2018). As of October 2023, detailed phenological monitoring, methodologically aligned with the monitoring at Paracou, has been established at Bouamir Research Station, located in the Cameroonian portion of the Congo Basin. This will provide a valuable comparison with the patterns observed in the northern Amazon and will be a test of the transferability of the protocols that were developed in the relatively affluent, European administered department of French Guiana. It goes without saying that greater financial investment in under-represented regions is a prerequisite for an effective global monitoring network, but more thought needs to be given to how utile ecological monitoring can be established in these places in a reliable, consistent and sustainable fashion. A first step will be to establish and share freely affordable and reproducible processes for mapping and tracking the dynamics of individual trees. In this vein, I hope this thesis (and its accompanying datasets and code) can contribute to an improved understanding of the globally significant ecological processes of the tropics.

6.5 Linking leaf phenology patterns with the carbon cycle in moist tropical forests

The goal of the PhenObs project, to which my PhD research has contributed, is to evaluate the links between leaf phenology and the carbon cycle of tropical forests. Tropical forests are known to play a crucial role in the global carbon (contributing approximately half the stock and

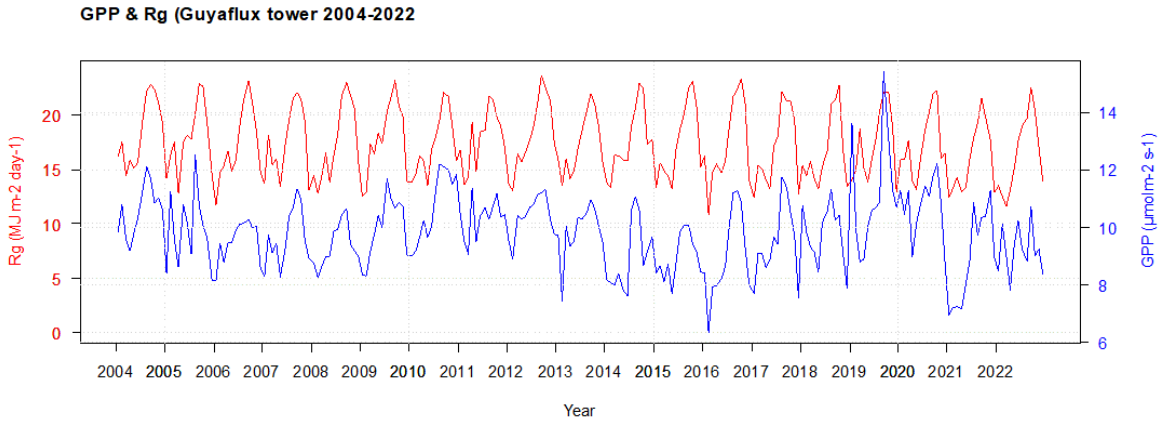


Fig. 6.2 Solar radiation and Gross Primary Productivity 2004-2022 at Paracou Research Station.

sink of the world's forests), water and energy budgets. However, the magnitude of matter and energy fluxes are poorly resolved both spatially and temporally, and the driving mechanisms remain unclear. For instance, net carbon fluxes between forest and atmosphere depend on emissions arising from deforestation and degradation activities, but they also relate to forest net primary productivity. The latter appears to fluctuate over seasonal and inter-annual time scales, for reasons that are yet to be fully elucidated. Overall carbon assimilation might well be larger than the emissions caused by land use changes, due, in part, to CO_2 fertilisation (Chen et al., 2022a). Consequently, estimates of the tropical forest carbon sink have the largest uncertainty of all forests worldwide. The findings of Chapter 5 will be used to inform the ongoing work on carbon cycling in the PhenObs project. Here, I outline how this work will be developed in the coming months and years.

6.5.1 The impacts of leaf phenology on primary productivity

As mentioned in Chapter 4 and Chapter 5, all the fieldwork in French Guiana was conducted within the footprint of a flux tower that monitors gas exchanges between the forest ecosystem and the atmosphere. In Chapter 5, we demonstrated that while crown greenness and PAI were synchronised there was divergence within a cycle as PAI stayed relatively level until shedding while GLI showed a more gradual decline, showing the aging of leaves. By comparing these patterns to meteorological data and gross primary productivity estimated using the eddy covariance method from the flux tower sensors (Fig. 6.2), it should be possible to evaluate the primary driver of seasonality of gross ecosystem productivity and net ecosystem exchange at the site (Fig. 6.3) and further understand the role of leaf phenology relative to other contributing factors.

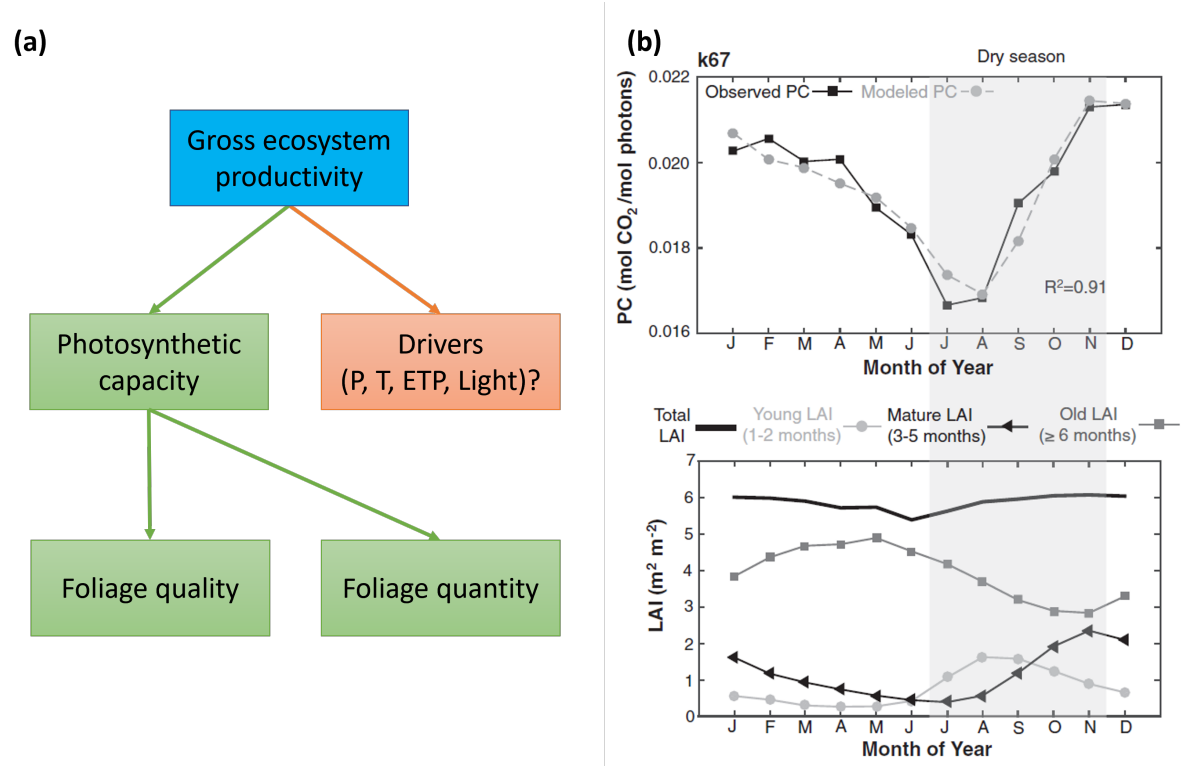


Fig. 6.3 Partitioning the controls on ecosystem productivity. (a) A schematic of the potential controls on gross ecosystem productivity; (b) Photosynthetic capacity (PC) and LAI change by leaf ages at Amazonian site k67 near Santarém, Brazil. PC is amount of photosynthesis per unit of incoming light measured at flux tower leaf age is defined from phenocam observations and light interception measurements made by paired LAI 2200 measurements. Source: Wu et al. (2016).

.

6.5.2 Improving the parameterisation of ecosystem models

Dynamic global climate-vegetation models (DGVMs) are currently the main tools used to understand and predict Earth's climate dynamics. They have played a critical role in informing and influencing major international agreements to mitigate climate change. As a result of the knowledge gaps identified at the start of this section, DGVMs consistently fail to mimic observed productivity dynamics or climate-vegetation feedbacks (Restrepo-Coupe et al., 2021), and therefore cannot be relied upon to accurately predict the fate and role of tropical forests under changing climate conditions. Lovenduski and Bonan (2017) demonstrated that ocean models were more congruent than terrestrial models in projections of carbon uptake, and that reducing uncertainty in terrestrial models requires improving processes understanding, monitoring and modelling.

A promising way to improve the accuracy of DGVMs is to build a new generation of individual tree simulators that are built on a detailed understanding of tree biology, including responses to water supply and leaf phenology (see e.g. Longo et al., 2019). TROLL is among the advanced process-based models for tropical forests, that predicts carbon and water fluxes at daily timescale and includes a module that takes leaf phenology into account (Maréchaux and Chave, 2017). Future work will focus on using TROLL to model ecosystem processes, building on a detailed knowledge of plant traits gathered from the field site over a number of years as well as the leaf cycling patterns presented in Chapter 5. It is hoped that this refinement to the parameterisation will lead to ecosystem models that better reflect observed GPP variation in tropical forest and thus give more a more reliable foundation for predicting future states of tropical forests.

6.5.3 Scaling up phenological observations

My thesis has primarily focused on a few hectares of rainforest in French Guiana that are among the most carefully and intensely studied of any forest on the planet. What opportunities are there to work over larger scales and gain a more complete picture of the drivers of leaf phenology? As explained in the introduction of Chapter 5, the vast majority of remote sensing studies have used passive optical imagery, for example from Sentinel 2. The Global Ecosystem Dynamics Investigation (GEDI) mission provides an opportunity to assess phenology using an active sensor in space (Cushman et al., 2023). GEDI is a NASA-led project aimed at providing high-resolution laser ranging observations of the Earth's forests and topography. Launched in December 2018, GEDI is specifically designed to measure the vertical structure of the Earth's surface, with a particular focus on forest canopy height, canopy vertical structure, and surface elevation. GEDI has an unusual spatial structure; a single sample (laser shot) has a footprint of

25 m and these are collected along eight narrow footprint tracks. Consequently, resampling the exact same location is extremely rare. Preliminary analyses in the Forest Ecology and Conservation Group in Cambridge show that GEDI can successfully track bulk patterns of PAD. This may corroborate the observed trends of overall canopy variation presented in Chapter 5 but lacks the granularity necessary to characterise the variety of patterns observed at the species and individual level. Linking our UAV-lidar measurements to GEDI and other optical, spaceborne sensors may help to generate insights on the subtle phenological patterns across the tropics.

6.6 Concluding remarks

Tropical forests are under intense anthropogenic pressures that are causing them to change at a rate that is greater than at any other point in human history. Historically, these pressures have been driven by technological advancements. As we entered an era in which machines possess interpretation and reasoning capabilities akin to those of humans, and there exists exponential growth in the amount of Earth observation data, this thesis has explored how deep learning and remote sensing technologies can be leveraged to comprehend and safeguard these vital ecosystems. I have shown that deep neural networks coupled with fieldwork can help to map these forests in unprecedented detail and allow us to observe and decode their cryptic rhythms. The leaf patterns of tropical forest trees are varied but exhibit adapted structure that capitalises on the cyclically varying availability of resources. By continuing to refine our understanding of these patterns, we will enhance our ability to predict the resilience of these ecosystems to the pressures of global change and understand the degree to which they can buffer and protect us from this change.

References

Abdi, O., Uusitalo, J., & Kivinen, V.-P. (2022). Logging trail segmentation via a novel U-Net convolutional neural network and High-Density laser scanning data. *Remote Sensing*, *14*(2), 349.

Abernethy, K., Bush, E. R., Forget, P.-M., Mendoza, I., & Morellato, L. P. C. (2018). Current issues in tropical phenology: A synthesis. *Biotropica*, *50*(3), 477–482.

Achard, F., Beuchle, R., Mayaux, P., Stibig, H.-J., Bodart, C., Brink, A., Carboni, S., Desclée, B., Donnay, F., Eva, H. D., Lupi, A., Raši, R., Seliger, R., & Simonetti, D. (2014). Determination of tropical deforestation rates and related carbon losses from 1990 to 2010. *Glob. Chang. Biol.*, *20*(8), 2540–2554.

Ackerly, D. D., & Donoghue, M. J. (1998). Leaf size, sapling allometry, and corner's rules: Phylogeny and correlated evolution in maples (acer). *Am. Nat.*, *152*(6), 767–791.

Agarwal, R. (2019). Demystifying object detection and instance segmentation for data scientists [Accessed: 2023-11-14].

AgiSoft. (2021). Agisoft metashape (version 1.7.3).

Agrawal, A. A. (2007). Macroevolution of plant defense strategies. *Trends Ecol. Evol.*, *22*(2), 103–109.

Ågren, A. M., Lidberg, W., Strömgren, M., Ogilvie, J., & Arp, P. A. (2014). Evaluating digital terrain indices for soil wetness mapping – a Swedish case study. *Hydrol. Earth Syst. Sci.*, *18*(9), 3623–3634.

Ahlswede, S., Schulz, C., Gava, C., Helber, P., Bischke, B., Förster, M., Arias, F., Hees, J., Demir, B., & Kleinschmit, B. (2023). *TREESATAI BENCHMARK ARCHIVE*: A multi-sensor, multi-label dataset for tree species classification in remote sensing. *Earth Syst. Sci. Data*, *15*(2), 681–695.

Ahmed, S. E., Ewers, R. M., & Smith, M. J. (2014). Large scale spatio-temporal patterns of road development in the Amazon rainforest. *Environ. Conserv.*, *41*(3), 253–264.

Aide, T. M. (1993). Patterns of leaf development and herbivory in a tropical understory community. *Ecology*, *74*(2), 455–466.

Ajani, E. N., Mgbenka, R. N., & Okeke, M. N. (2013). Use of indigenous knowledge as a strategy for climate change adaptation among farmers in sub-Saharan Africa: Implications for policy. *AJAEES*, 23–40.

Alberton, B., Martin, T. C. M., Da Rocha, H. R., Richardson, A. D., Moura, M. S. B., Torres, R. S., & Morellato, L. P. C. (2023). Relationship between tropical leaf phenology and ecosystem productivity using phenocameras. *Front. Environ. Sci. Eng. China*, *11*.

Alberton, B., Torres, R. d. S., Cancian, L. F., Borges, B. D., Almeida, J., Mariano, G. C., Santos, J. d., & Morellato, L. P. C. (2017). Introducing digital cameras to monitor plant phenology in the tropics: Applications for conservation. *Perspectives in Ecology and Conservation*, *15*(2), 82–90.

Almeida, A., Gonçalves, F., Silva, G., Mendonça, A., Gonzaga, M., Silva, J., Souza, R., Leite, I., Neves, K., Boeno, M., & Sousa, B. (2021). Individual tree detection and qualitative inventory of a eucalyptus sp. stand using uav photogrammetry data. *Remote Sensing*, *13*(18). https://doi.org/10.3390/rs13183655

Alroy, J. (2017). Effects of habitat disturbance on tropical forest biodiversity. *Proc. Natl. Acad. Sci. U. S. A.*, *114*(23), 6056–6061.

Álvarez-Yépiz, J. C., Búrquez, A., Martinez-Yrizar, A., Teece, M., Yépez, E. A., & Dovciak, M. (2017). Resource partitioning by evergreen and deciduous species in a tropical dry forest. *Oecologia*, *183*(2), 607–618.

Anderson, L. O., Aragão, L. E. O. C., Shimabukuro, Y. E., Almeida, S., & Huete, A. (2011). Fraction images for monitoring intra-annual phenology of different vegetation physiognomies in Amazonia. *Int. J. Remote Sens.*, *32*(2), 387–408.

Araujo, R. F., Grubinger, S., Celes, C. H. S., Negrón-Juárez, R. I., Garcia, M., Dandois, J. P., & Muller-Landau, H. C. (2021). Strong temporal variation in treefall and branchfall rates in a tropical forest is related to extreme rainfall: Results from 5 years of monthly drone data for a 50 ha plot. *Biogeosciences*, *18*(24), 6517–6531.

Arkin, J., Coops, N. C., Daniels, L. D., & Plowright, A. (2023). A novel post-fire method to estimate individual tree crown scorch height and volume using simple RPAS-derived data. *Fire Ecol*, *19*(1), 17.

Arroyo-Mora, J. P., Kalacska, M., Løke, T., Schläpfer, D., Coops, N. C., Lucanus, O., & Leblanc, G. (2021). Assessing the impact of illumination on UAV pushbroom hyperspectral imagery collected under various cloud cover conditions. *Remote Sens. Environ.*, *258*, 112396.

Artaxo, P., Hansson, H. C., Machado, L. A. T., & Rizzo, L. V. (2022a). Tropical forests are crucial in regulating the climate on Earth. *PLOS Climate*, *1*(8), e0000054.

Artaxo, P., Hansson, H.-C., Andreae, M. O., Bäck, J., Alves, E. G., Barbosa, H. M. J., Bender, F., Bourtsoukidis, E., Carbone, S., Chi, J., Decesari, S., Després, V. R., Ditas, F., Ezhova, E., Fuzzi, S., Hasselquist, N. J., Heintzenberg, J., Holanda, B. A., Guenther, A., … Kesselmeier, J. (2022b). Tropical and boreal forest – atmosphere interactions: A review. *Tellus. Series B: Chemical and Physical Meteorology.*, *74*(1), 24–163.

Asner, G. P. (2014). A chemical-evolutionary basis for remote sensing of tropical forest diversity. In David A. Coomes, David F. R. P. Burslem, William D. Simonson (Ed.), *Forests and global change* (pp. 343–358). Cambridge University Press.

Asner, G. P., & Martin, R. E. (2011). Canopy phylogenetic, chemical and spectral assembly in a lowland Amazonian forest. *New Phytol.*, *189*(4), 999–1012.

Asner, G. P., Martin, R. E., Tupayachi, R., Anderson, C. B., Sinca, F., Carranza-Jiménez, L., & Martinez, P. (2014). Amazonian functional diversity from forest canopy chemical assembly. *Proc. Natl. Acad. Sci. U. S. A.*, *111*(15), 5604–5609.

Asner, G. P., Nepstad, D., Cardinot, G., & Ray, D. (2004). Drought stress and carbon uptake in an Amazon forest measured with spaceborne imaging spectroscopy. *Proceedings of the National Academy of Sciences*, *101*(16), 6039–6044.

Asner, G. P., & Tupayachi, R. (2017). Accelerated losses of protected forests from gold mining in the Peruvian Amazon. *Environ. Res. Lett.*, *12*(9), 094004.

Atkinson, P. M., Dash, J., & Jeganathan, C. (2011). Amazon vegetation greenness as measured by satellite sensors over the last decade. *Geophys. Res. Lett.*, *38*(19).

Aubry-Kientz, M., Dutrieux, R., Ferraz, A., Saatchi, S., Hamraz, H., Williams, J., Coomes, D., Piboule, A., & Vincent, G. (2019). A comparative assessment of the performance of individual tree crowns delineation algorithms from ALS data in tropical forests. *Remote Sensing*, *11*(9), 1086.

Aubry-Kientz, M., Laybros, A., Weinstein, B., Ball, J. G. C., Jackson, T., Coomes, D., & Vincent, G. (2021). Multisensor data fusion for improved segmentation of individual tree crowns in dense tropical forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 3927–3936.

Audibert, J.-Y., & Catoni, O. (2011). Robust linear least squares regression. *The Annals of Statistics*, *39*(5), 2766–2794.

Avitabile, V., Herold, M., Heuvelink, G. B. M., Lewis, S. L., Phillips, O. L., Asner, G. P., Armston, J., Ashton, P. S., Banin, L., Bayol, N., Berry, N. J., Boeckx, P., de Jong, B. H. J., DeVries, B., Girardin, C. A. J., Kearsley, E., Lindsell, J. A., Lopez-Gonzalez, G., Lucas, R., … Willcock, S. (2016). An integrated pan-tropical biomass map using multiple reference datasets. *Glob. Chang. Biol.*, *22*(4), 1406–1420.

Baccini, A., Walker, W., Carvalho, L., Farina, M., Sulla-Menashe, D., & Houghton, R. A. (2017). Tropical forests are a net carbon source based on aboveground measurements of gain and loss. *Science*, *358*(6360), 230–234.

Badourdine, C., Féret, J.-B., Pélissier, R., & Vincent, G. (2023). Exploring the link between spectral variance and upper canopy taxonomic diversity in a tropical forest: Influence of spectral processing and feature selection. *Remote Sensing in Ecology and Conservation*, *9*(2), 235–250.

Bai, Y., Durand, J.-B., Forbes, F., & Vincent, G. (2023). Semantic segmentation of sparse irregular point clouds for leaf/wood discrimination arXiv 2305.16963.

Baldeck, C. A., & Asner, G. P. (2014). Improving remote species identification through efficient training data collection. *Remote Sensing*, *6*(4), 2682–2698.

Baldeck, C. A., Asner, G. P., Martin, R. E., Anderson, C. B., Knapp, D. E., Kellner, J. R., & Wright, S. J. (2015). Operational tree species mapping in a diverse tropical forest with airborne imaging spectroscopy. *PLoS One*, *10*(7), e0118403.

Ball, J. G. C. (2022). Detectree2: Python package for automatic tree crown delineation based on the detectron2 implementation of mask R-CNN.

Ball, J. G. C., Hickman, S. H. M., Jackson, T. D., Koay, X. J., Hirst, J., Jay, W., Archer, M., Aubry-Kientz, M., Vincent, G., & Coomes, D. A. (2023). Accurate delineation of individual tree crowns in tropical forests from aerial RGB imagery using Mask R-CNN. *Remote Sensing in Ecology and Conservation*, *9*(5), 641–655.

Ball, J. G. C., Petrova, K., Coomes, D. A., & Flaxman, S. (2022). Using deep convolutional neural networks to forecast spatial patterns of Amazonian deforestation. *Methods in Ecology and Evolution*, *13*(11), 2622–2634.

Ballas, N., Yao, L., Pal, C., & Courville, A. (2015). Delving deeper into convolutional networks for learning video representations.

Bandyopadhyay, D., Mukherjee, S., Ball, J. G. C., Vincent, G., Coomes, D. A., & Schönlieb, C.-B. (2022). Tree species classification from hyperspectral data using graph-regularized neural networks arXiv 2208.08675.

Baraloto, C., Hardy, O. J., Paine, C. E. T., Dexter, K. G., Cruaud, C., Dunning, L. T., Gonzalez, M.-A., Molino, J.-F., Sabatier, D., Savolainen, V., & Chave, J. (2012). Using functional traits and phylogenetic trees to examine the assembly of tropical tree communities. *J. Ecol.*, *100*(3), 690–701.

Barber, C. P., Cochrane, M. A., Souza, C. M., & Laurance, W. F. (2014). Roads, deforestation, and the mitigating effect of protected areas in the Amazon. *Biol. Conserv.*, *177*, 203–209.

Bass, M. S., Finer, M., Jenkins, C. N., Kreft, H., Cisneros-Heredia, D. F., McCracken, S. F., Pitman, N. C. A., English, P. H., Swing, K., Villa, G., Di Fiore, A., Voigt, C. C., & Kunz, T. H. (2010). Global conservation significance of Ecuador's Yasuni National Park. *PLoS One*, *5*(1), e8767.

Baumgartner, A., Gege, P., Köhler, C., Lenhard, K., & Schwarzmaier, T. (2012). Characterisation methods for the hyperspectral sensor HySpex at DLR's calibration home base, In *Sensors, systems, and Next-Generation satellites XVI*.

Beech, E., Rivers, M., Oldfield, S., & Smith, P. P. (2017). GlobalTreeSearch: The first complete global database of tree species and country distributions. *J. Sustainable For.*, *36*(5), 454–489.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Roupsard, O., ... Papale, D. (2010). Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate. *Science*, *329*(5993), 834–838.

Béland, M., & Kobayashi, H. (2024). Drivers of deciduous forest near-infrared reflectance: A 3D radiative transfer modeling exercise based on ground lidar. *Remote Sens. Environ.*, *302*, 113951.

Beloiu, M., Heinzmann, L., Rehush, N., Gessler, A., & Griess, V. C. (2023). Individual Tree-Crown detection and species identification in heterogeneous forests using aerial RGB imagery and deep learning. *Remote Sensing*, *15*(5), 1463.

Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D., & Bengio, Y. (2010). Theano: A CPU and GPU math expression compiler, In *Proceedings of the python for scientific computing conference (SciPy)*.

Besl, P. J., & McKay, N. D. (1992). A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, *14*(2), 239–256.

Beuchle, R., Achard, F., Bourgoin, C., Vancutsem, C., Eva, H., & Follador, M. (2021). *Deforestation and forest degradation in the Amazon* (tech. rep.). European Union. Luxembourg (Luxembourg), Publications Office of the European Union.

Biggs, C. R., Yeager, L. A., Bolser, D. G., Bonsell, C., Dichiera, A. M., Hou, Z., Keyser, S. R., Khursigara, A. J., Lu, K., Muth, A. F., Negrete, B., Jr, & Erisman, B. E. (2020). Does functional redundancy affect ecological stability and resilience? A review and meta-analysis. *Ecosphere*, *11*(7).

Bittencourt, P. R. d. L., Bartholomew, D. C., Banin, L. F., Bin Suis, M. A. F., Nilus, R., Burslem, D. F. R. P., & Rowland, L. (2022). Divergence of hydraulic traits among tropical forest

trees across topographic and vertical environment gradients in Borneo. *New Phytol.*, *235*(6), 2183–2198.

Blomberg, S. P., Garland, T., Jr, & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, *57*(4), 717–745.

Bohlman, S. (2019). Inter- and intra-specific variation in tropical canopy phenology: Insights from field data and images from unmanned aerial vehicles (UAVs). *School of Geographythe Environment, University of Oxford.*

Bohlman, S. A. (2010). Landscape patterns and environmental controls of deciduousness in forests of central Panama. *Glob. Ecol. Biogeogr.*, *19*(3), 376–385.

Bohlman, S. A. (2015). Species diversity of canopy versus understory trees in a neotropical forest: Implications for forest structure, function and monitoring. *Ecosystems*, *18*(4), 658–670.

Bonal, D., Bosc, A., Ponton, S., Goret, J.-Y., Burban, B., Gross, P., Bonnefond, J.-M., Elbers, J., Longdoz, B., Epron, D., Guehl, J.-M., & Granier, A. (2008). Impact of severe dry season on net ecosystem exchange in the neotropical rainforest of French Guiana. *Glob. Chang. Biol.*, *14*(8), 1917–1933.

Bonan, G. B. (2008). Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science*, *320*(5882), 1444–1449.

Boochs, F., Kupfer, G., Dockter, K., & Kühbauch, W. (1990). Shape of the red edge as vitality indicator for plants. *Int. J. Remote Sens.*, *11*(10), 1741–1753.

Borchert, R. (1980). Phenology and ecophysiology of tropical trees: Erythrina poeppigiana o. f. cook. *Ecology*, *61*(5), 1065–1074.

Borchert, R. (1994a). Induction of rehydration and bud break by irrigation or rain in decidous trees of a tropical dry forest in Costa Rica. *Trees*, *8*(4), 198–204.

Borchert, R. (1994b). Water status and development of tropical trees during seasonal drought. *Trees*, *8*(3), 115–125.

Boulton, C. A., Lenton, T. M., & Boers, N. (2022). Pronounced loss of Amazon rainforest resilience since the early 2000s. *Nat. Clim. Chang.*, *12*(3), 271–278.

Boyd, J., Epanchin-Niell, R., & Siikamäki, J. (2015). Conservation planning: A review of return on investment analysis. *Rev. Environ. Econ. Pol.*, *9*(1), 23–42.

Braakmann-Folgmann, A., Roscher, R., Wenzel, S., Uebbing, B., & Kusche, J. (2017). Sea level anomaly prediction using recurrent neural networks.

Braga, J. R. G., Peripato, V., Dalagnol, R., P Ferreira, M., Tarabalka, Y., OC Aragão, L. E., F de Campos Velho, H., Shiguemori, E. H., & Wagner, F. H. (2020). Tree crown delineation algorithm based on a convolutional neural network. *Remote Sensing*, *12*(8), 1288.

Bredin, Y. K., Peres, C. A., & Haugaasen, T. (2020). Forest type affects the capacity of Amazonian tree species to store carbon as woody biomass. *For. Ecol. Manage.*, *473*, 118297.

Brienen, R. J. W., Phillips, O. L., Feldpausch, T. R., Gloor, E., Baker, T. R., Lloyd, J., Lopez-Gonzalez, G., Monteagudo-Mendoza, A., Malhi, Y., Lewis, S. L., Vásquez Martinez, R., Alexiades, M., Álvarez Dávila, E., Alvarez-Loayza, P., Andrade, A., Aragão, L. E. O. C., Araujo-Murakami, A., Arets, E. J. M. M., Arroyo, L., … Zagt, R. J. (2015). Long-term decline of the Amazon carbon sink. *Nature*, *519*(7543), 344–348.

Brockerhoff, E. G., Barbaro, L., Castagneyrol, B., Forrester, D. I., Gardiner, B., González-Olabarria, J. R., Lyver, P. O., Meurisse, N., Oxbrough, A., Taki, H., Thompson, I. D., van der Plas, F., & Jactel, H. (2017). Forest biodiversity, ecosystem functioning and the provision of ecosystem services. *Biodivers. Conserv.*, *26*(13), 3005–3035.

Brodie, J. F., Mohd-Azlan, J., Chen, C., Wearn, O. R., Deith, M. C. M., Ball, J. G. C., Slade, E. M., Burslem, D. F. R. P., Teoh, S. W., Williams, P. J., Nguyen, A., Moore, J. H., Goetz, S. J., Burns, P., Jantz, P., Hakkenberg, C. R., Kaszta, Z. M., Cushman, S., Coomes, D., … Luskin, M. S. (2023). Landscape-scale benefits of protected areas for tropical biodiversity. *Nature*, *620*(7975), 807–812.

Brodrick, P. G., Davies, A. B., & Asner, G. P. (2019). Uncovering ecological patterns with convolutional neural networks. *Trends Ecol. Evol.*, *34*(8), 734–745.

Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., & Smets, B. (2020). Copernicus global land cover Layers—Collection 2. *Remote Sensing*, *12*(6), 1044.

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw.*, *106*, 249–259.

Bullock, S. H., & Solis-Magallanes, J. A. (1990). Phenology of canopy trees of a tropical deciduous forest in mexico. *Biotropica*, *22*(1), 22–35.

Busch, J., & Ferretti-Gallon, K. (2017). What drives deforestation and what stops it? a Meta-Analysis. *Review of Environmental Economics and Policy*, *11*(1), 3–23.

Calderon, L. d. A., Silva-Jardim, I., Zuliani, J. P., Silva, A. d. A. e., Ciancaglini, P., Silva, L. H. P. d., & Stábeli, R. G. (2009). Amazonian biodiversity: A view of drug development for Leishmaniasis and malaria. *J. Braz. Chem. Soc.*, *20*(6), 1011–1023.

Cao, Y., Ball, J. G. C., Coomes, D. A., Steinmeier, L., Knapp, N., Wilkes, P., Disney, M., Calders, K., Burt, A., Lin, Y., & Jackson, T. D. (2023). Benchmarking airborne laser scanning tree segmentation algorithms in broadleaf forests shows high accuracy only for canopy trees. *Int. J. Appl. Earth Obs. Geoinf.*, *123*, 103490.

Carter, G. A., & Knapp, A. K. (2001). Leaf optical properties in higher plants: Linking spectral characteristics to stress and chlorophyll concentration. *Am. J. Bot.*, *88*(4), 677–684.

Carter, S., Armstrong, Z., Schubert, L., Johnson, I., & Olah, C. (2019). Activation atlas. *Distill*, *4*(3).

Cavender-Bares, J., Meireles, J. E., Couture, J. J., Kaproth, M. A., Kingdon, C. C., Singh, A., Serbin, S. P., Center, Alyson, Zuniga, E., Pilz, G., & Townsend, P. A. (2016). Associations of leaf spectra with genetic and phylogenetic variation in oaks: Prospects for remote detection of biodiversity. *Remote Sensing*, *8*(3), 221.

Cavender-Bares, J., Schneider, F. D., Santos, M. J., Armstrong, A., Carnaval, A., Dahlin, K. M., Fatoyinbo, L., Hurtt, G. C., Schimel, D., Townsend, P. A., Ustin, S. L., Wang, Z., & Wilson, A. M. (2022). Integrating remote sensing with ecology and evolution to advance biodiversity conservation. *Nat Ecol Evol*, *6*(5), 506–519.

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proc. Natl. Acad. Sci. U. S. A.*, *114*(30), E6089–E6096.

Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S., & Grégoire, J.-M. (2001). Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sens. Environ.*, *77*(1), 22–33.

Chabot, B. F., & Hicks, D. J. (1982). The ecology of leaf life spans. *Annu. Rev. Ecol. Syst.*, *13*, 229–259.

Chai, Y., Martins, G., Nobre, C., von Randow, C., Chen, T., & Dolman, H. (2021). Constraining Amazonian land surface temperature sensitivity to precipitation and the probability of forest dieback. *npj Climate and Atmospheric Science*, *4*, 6. https://doi.org/10.1038/s41612-021-00162-1

Chambers, D., Périé, C., Casajus, N., & de Blois, S. (2013). Challenges in modelling the abundance of 105 tree species in eastern North America using climate, edaphic, and topographic variables. *For. Ecol. Manage.*, *291*, 20–29.

Chan, A. H. Y., Barnes, C., Swinfield, T., & Coomes, D. A. (2021). Monitoring ash dieback (Hymenoscyphus fraxineus) in British forests using hyperspectral remote sensing. *Remote Sens. Ecol. Conserv.*, *7*(2), 306–320.

Chance, C. M., Coops, N. C., Crosby, K., & Aven, N. (2016). Spectral wavelength selection and detection of two invasive plant species in an urban area. *Can. J. Remote Sens.*, *42*(1), 27–40.

Chavana-Bryant, C., Malhi, Y., Anastasiou, A., Enquist, B. J., Cosio, E. G., Keenan, T. F., & Gerard, F. F. (2019). Leaf age effects on the spectral predictability of leaf traits in Amazonian canopy trees. *Sci. Total Environ.*, *666*, 1301–1315.

Chavana-Bryant, C., Malhi, Y., Wu, J., Asner, G. P., Anastasiou, A., Enquist, B. J., Cosio Caravasi, E. G., Doughty, C. E., Saleska, S. R., Martin, R. E., & Gerard, F. F. (2017).

Leaf aging of Amazonian canopy trees as revealed by spectral and physiochemical measurements. *New Phytol.*, *214*(3), 1049–1063.

Chave, J., Navarrete, D., Almeida, S., Álvarez, E., Aragão, L. E. O. C., Bonal, D., Châtelet, P., Silva-Espejo, J. E., Goret, J.-Y., von Hildebrand, P., Jiménez, E., Patiño, S., Peñuela, M. C., Phillips, O. L., Stevenson, P., & Malhi, Y. (2010). Regional and seasonal patterns of litterfall in tropical South America. *Biogeosciences*, *7*, 43–55.

Chave, J. (1999). Study of structural, successional and spatial patterns in tropical rain forests using TROLL, a spatially explicit forest model. *Ecol. Modell.*, *124*(2), 233–254.

Chave, J. (2014). Floristic shifts versus critical transitions in Amazonian forest systems. In *Forests and global change* (pp. 131–160). Cambridge University Press.

Chave, J., Davies, S. J., Phillips, O. L., Lewis, S. L., Sist, P., Schepaschenko, D., Armston, J., Baker, T. R., Coomes, D., Disney, M., Duncanson, L., Hérault, B., Labrière, N., Meyer, V., Réjou-Méchain, M., Scipal, K., & Saatchi, S. (2019). Ground data are essential for biomass remote sensing missions. *Surv. Geophys.*, *40*(4), 863–880.

Chen, C., Riley, W. J., Prentice, I. C., & Keenan, T. F. (2022a). CO_2 fertilization of terrestrial photosynthesis inferred from site to global scales. *Proc. Natl. Acad. Sci. U. S. A.*, *119*(10), e2115627119.

Chen, J. M., Mo, G., Pisek, J., Liu, J., Deng, F., Ishizawa, M., & Chan, D. (2012). Effects of foliage clumping on the estimation of global terrestrial gross primary productivity. *Global Biogeochemical Cycles*, *26*(1).

Chen, L., Zhang, Y., Nunes, M. H., Stoddart, J., Khoury, S., Chan, A. H. Y., & Coomes, D. A. (2022b). Predicting leaf traits of temperate broadleaf deciduous trees from hyperspectral reflectance: Can a general model be applied across a growing season? *Remote Sens. Environ.*, *269*, 112767.

Chen, X., Ciais, P., Maignan, F., Zhang, Y., Bastos, A., Liu, L., Bacour, C., Fan, L., Gentine, P., Goll, D., Green, J., Kim, H., Li, L., Liu, Y., Peng, S., Tang, H., Viovy, N., Wigneron, J.-P., Wu, J., … Zhang, H. (2021). Vapor pressure deficit and sunlight explain seasonality of leaf phenology and photosynthesis across Amazonian evergreen broadleaved forest. *Global Biogeochem. Cycles*, *35*(6).

Chen, X., Maignan, F., Viovy, N., Bastos, A., Goll, D., Wu, J., Liu, L., Yue, C., Peng, S., Yuan, W., Conceição, A. C., O'Sullivan, M., & Ciais, P. (2020). Novel representation of leaf phenology improves simulation of Amazonian evergreen forest photosynthesis in a land surface model. *J. Adv. Model. Earth Syst.*, *12*(1).

Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transac-*

tions on Geoscience and Remote Sensing, *54*(10), 6232–6251. https://doi.org/10.1109/TGRS.2016.2584107

Chirinos, C., & Ruiz, M. (2003). Informe final del estudio: Desarrollo e implementación de lineamientos de control de la extracción ilegal para un manejo forestal sostenible en el perú. *ITTC, Ciudad de Panamá, Panama*.

Ciocırlan, M. I. C., Curtu, A. L., & Radu, G. R. (2022). Predicting leaf phenology in forest tree species using UAVs and satellite images: A case study for European Beech (Fagus sylvatica L.) *Remote Sensing*, *14*(24), 6198.

Clark, M. L., & Roberts, D. A. (2012). Species-Level differences in hyperspectral metrics among tropical rainforest trees as determined by a Tree-Based classifier. *Remote Sensing*, *4*(6), 1820–1855.

Clark, M. L., Roberts, D. A., & Clark, D. B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sens. Environ.*, *96*(3), 375–398.

Cleland, E. E., Chuine, I., Menzel, A., Mooney, H. A., & Schwartz, M. D. (2007). Shifting plant phenology in response to global change. *Trends Ecol. Evol.*, *22*(7), 357–365.

Climate Focus. (2023). *Off track and falling behind: Tracking progress on 2030 forest goals* (tech. rep.). Forest Declaration Assessment.

Clinton, N., Holt, A., Scarborough, J., Yan, L., Gong, P. Et al. (2010). Accuracy assessment measures for object-based image segmentation goodness. *Photogramm. Eng. Remote Sens*, *76*(3), 289–299.

Coley, P. D. (1983). Herbivory and defensive characteristics of tree species in a lowland tropical forest. *Ecol. Monogr.*, *53*(2), 209–234.

Coomes, D. A., Dalponte, M., Jucker, T., Asner, G. P., Banin, L. F., Burslem, D. F., Lewis, S. L., Nilus, R., Phillips, O. L., Phua, M.-H. Et al. (2017). Area-based vs tree-centric approaches to mapping forest carbon in Southeast Asian forests from airborne laser scanning data. *Remote Sensing of Environment*, *194*, 77–88.

Coomes, D. A., Duncan, R. P., Allen, R. B., & Truscott, J. (2003). Disturbances prevent stem size-density distributions in natural forests from following scaling relationships. *Ecology letters*, *6*(11), 980–989.

Cooper, D. L. M., Lewis, S. L., Sullivan, M. J. P., Prado, P. I., Ter Steege, H., Barbier, N., Slik, F., Sonké, B., Ewango, C. E. N., Adu-Bredu, S., Affum-Baffoe, K., de Aguiar, D. P. P., Ahuite Reategui, M. A., Aiba, S.-I., Albuquerque, B. W., de Almeida Matos, F. D., Alonso, A., Amani, C. A., do Amaral, D. D., … Zent, S. (2024). Consistent patterns of common species across tropical tree communities. *Nature*, *625*(7996), 728–734.

Corbane, C., Syrris, V., Sabo, F., Politis, P., Melchiorri, M., Pesaresi, M., Soille, P., & Kemper, T. (2021). Convolutional neural networks for global human settlements mapping from sentinel-2 satellite imagery. *Neural Comput. Appl.*, *33*(12), 6697–6720.

Corlett, R. T. (2011). Impacts of warming on tropical lowland rainforests. *Trends Ecol. Evol.*, *26*(11), 606–613.

Cottrell, C. (2022). Avoiding a new era in biopiracy: Including indigenous and local knowledge in nature-based solutions to climate change. *Environ. Sci. Policy*, *135*, 162–168.

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., & Totterdell, I. J. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, *408*(6809), 184–187.

Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sens. Environ.*, *30*(3), 271–278.

Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A., & Hansen, M. C. (2018). Classifying drivers of global forest loss. *Science*, *361*(6407), 1108–1111.

Cushman, K. C., Armston, J., Dubayah, R., Duncanson, L., Hancock, S., Janik, D., Král, K., Krůček, M., Minor, D. M., Tang, H., & Kellner, J. R. (2023). Impact of leaf phenology on estimates of aboveground biomass density in a deciduous broadleaf forest from simulated GEDI lidar. *Environ. Res. Lett.*, *18*(6), 065009.

Cushman, K., Detto, M., Garcia, M., & Muller-Landau, H. C. (2022). Soils and topography control natural disturbance rates and thereby forest structure in a lowland tropical landscape. *Ecology Letters*.

Cushman, S. A., Macdonald, E. A., Landguth, E. L., Malhi, Y., & Macdonald, D. W. (2017). Multiple-scale prediction of forest loss risk across Borneo. *Landsc. Ecol.*, *32*(8), 1581–1598.

Dai, Z., Liu, H., Le, Q. V., & Tan, M. (2021). CoAtNet: Marrying convolution and attention for all data sizes arXiv 2106.04803.

Dalponte, M., Bruzzone, L., & Gianelle, D. (2012). Tree species classification in the southern alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sens. Environ.*, *123*, 258–270.

Dalponte, M., & Coomes, D. A. (2016). Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods Ecol. Evol.*, *7*(10), 1236–1245.

Dalponte, M., Ørka, H. O., Ene, L. T., Gobakken, T., & Næsset, E. (2014). Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. *Remote Sens. Environ.*, *140*, 306–317.

Davies, S. J., Abiem, I., Abu Salim, K., Aguilar, S., Allen, D., Alonso, A., Anderson-Teixeira, K., Andrade, A., Arellano, G., Ashton, P. S., Baker, P. J., Baker, M. E., Baltzer, J. L.,

Basset, Y., Bissiengou, P., Bohlman, S., Bourg, N. A., Brockelman, W. Y., Bunyavejchewin, S., … Zuleta, D. (2021). ForestGEO: Understanding forest diversity and dynamics through a global observatory network. *Biol. Conserv.*, *253*, 108907.

Davis, C. C., Lyra, G. M., Park, D. S., Asprino, R., Maruyama, R., Torquato, D., Cook, B. I., & Ellison, A. M. (2022). New directions in tropical phenology. *Trends Ecol. Evol.*, *37*(8), 683–693.

de Camargo, M. G. G., de Carvalho, G. H., Alberton, B. d. C., Reys, P., & Morellato, L. P. C. (2018). Leafing patterns and leaf exchange strategies of a cerrado woody community. *Biotropica*, *50*(3), 442–454.

de Filho, F. J. B. O., & Metzger, J. P. (2006). Thresholds in landscape structure for three common deforestation patterns in the Brazilian Amazon. *Landsc. Ecol.*, *21*(7), 1061–1073.

de Souza, B. C., Carvalho, E. C. D., Oliveira, R. S., de Araujo, F. S., de Lima, A. L. A., & Rodal, M. J. N. (2020). Drought response strategies of deciduous and evergreen woody species in a seasonally dry neotropical forest. *Oecologia*, *194*(1-2), 221–236.

Dirzo, R., & Raven, P. H. (2003). Global state of biodiversity and loss. *Annu. Rev. Environ. Resour.*, *28*(1), 137–167.

Do, F. C., Goudiaby, V. A., Gimenez, O., Diagne, A. L., Diouf, M., Rocheteau, A., & Akpo, L. E. (2005). Environmental influence on canopy phenology in the dry tropics. *For. Ecol. Manage.*, *215*(1), 319–328.

Donovan, L. A., Maherali, H., Caruso, C. M., Huber, H., & de Kroon, H. (2011). The evolution of the worldwide leaf economics spectrum. *Trends Ecol. Evol.*, *26*(2), 88–95.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale arXiv 2010.11929.

Doughty, C. E., Keany, J. M., Wiebe, B. C., Rey-Sanchez, C., Carter, K. R., Middleby, K. B., Cheesman, A. W., Goulden, M. L., da Rocha, H. R., Miller, S. D., Malhi, Y., Fauset, S., Gloor, E., Slot, M., Oliveras Menor, I., Crous, K. Y., Goldsmith, G. R., & Fisher, J. B. (2023). Tropical forests are approaching critical temperature thresholds. *Nature*, *621*(7977), 105–111.

Doughty, R., Köhler, P., Frankenberg, C., Magney, T. S., Xiao, X., Qin, Y., Wu, X., & Moore, B., 3rd. (2019). TROPOMI reveals dry-season increase of solar-induced chlorophyll fluorescence in the Amazon forest. *Proc. Natl. Acad. Sci. U. S. A.*, *116*(44), 22393–22398.

Dubayah, R. O., Sheldon, S. L., Clark, D. B., Hofton, M. A., Blair, J. B., Hurtt, G. C., & Chazdon, R. L. (2010). Estimation of tropical forest height and biomass dynamics

using lidar remote sensing at La Selva, Costa Rica. *Journal of Geophysical Research: Biogeosciences*, *115*(G2).

Duque, A., Muller-Landau, H. C., Valencia, R., Cardenas, D., Davies, S., de Oliveira, A., Pérez, Á. J., Romero-Saltos, H., & Vicentini, A. (2017). Insights into regional patterns of Amazonian forest structure, diversity, and dominance from three large terra-firme forest dynamics plots. *Biodivers. Conserv.*, *26*(3), 669–686.

Duthoit, S., Demarez, V., Gastellu-Etchegorry, J.-P., Martin, E., & Roujean, J.-L. (2008). Assessing the effects of the clumping phenomenon on BRDF of a maize crop based on 3D numerical scenes using DART model. *Agric. For. Meteorol.*, *148*(8-9), 1341–1352.

Eamus, D. (1999). Ecophysiological traits of deciduous and evergreen woody species in the seasonally dry tropics. *Trends Ecol. Evol.*, *14*(1), 11–16.

Ebengo, D. M., de Boissieu, F., Vincent, G., Weber, C., & Féret, J.-B. (2021). Simulating imaging spectroscopy in tropical forest with 3D radiative transfer modeling. *Remote Sensing*, *13*(11), 2120.

Ellis, E. C., Kaplan, J. O., Fuller, D. Q., Vavrus, S., Klein Goldewijk, K., & Verburg, P. H. (2013). Used planet: A global history. *Proc. Natl. Acad. Sci. U. S. A.*, *110*(20), 7978–7985.

Endara, M.-J., Coley, P. D., Ghabash, G., Nicholls, J. A., Dexter, K. G., Donoso, D. A., Stone, G. N., Pennington, R. T., & Kursar, T. A. (2017). Coevolutionary arms race versus host defense chase in a tropical herbivore-plant system. *Proc. Natl. Acad. Sci. U. S. A.*, *114*(36), E7499–E7505.

Ene, L., Næsset, E., & Gobakken, T. (2012). Single tree detection in heterogeneous boreal forests using airborne laser scanning and area-based stem number estimates. *Int. J. Remote Sens.*, *33*(16), 5171–5193.

Esquivel-Muelbert, A., Baker, T. R., Dexter, K. G., Lewis, S. L., Brienen, R. J. W., Feldpausch, T. R., Lloyd, J., Monteagudo-Mendoza, A., Arroyo, L., Álvarez-Dávila, E., Higuchi, N., Marimon, B. S., Marimon-Junior, B. H., Silveira, M., Vilanova, E., Gloor, E., Malhi, Y., Chave, J., Barlow, J., … Phillips, O. L. (2019). Compositional response of Amazon forests to climate change. *Glob. Chang. Biol.*, *25*(1), 39–56.

Estiarte, M., & Peñuelas, J. (2015). Alteration of the phenology of leaf senescence and fall in winter deciduous species by climate change: Effects on nutrient proficiency. *Glob. Chang. Biol.*, *21*(3), 1005–1017.

FAO. (2012). *Global ecological zones for FAO forest reporting: 2010 update* (tech. rep. No. 179). Food & Agriculture Organization of the United Nations.

FAO. (2020). *Global forest resources assessment 2020: Main report*. Food & Agriculture Organization of the United Nations.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.*, *186*, 64–87.

Fassnacht, F. E., White, J. C., Wulder, M. A., & Næsset, E. (2023). Remote sensing in forestry: Current challenges, considerations and directions. *Forestry*.

Fauset, S., Johnson, M. O., Gloor, M., Baker, T. R., Monteagudo M, A., Brienen, R. J. W., Feldpausch, T. R., Lopez-Gonzalez, G., Malhi, Y., ter Steege, H., Pitman, N. C. A., Baraloto, C., Engel, J., Pétronelli, P., Andrade, A., Camargo, J. L. C., Laurance, S. G. W., Laurance, W. F., Chave, J., … Phillips, O. L. (2015). Hyperdominance in Amazonian forest carbon cycling. *Nat. Commun.*, *6*, 6857.

Fei-Fei Li, S. S., Juan Carlos Niebles. (2017). STANFORD VISION AND LEARNING LAB CS231n convolutional neural networks for visual recognition.

Feng, Y., Zeng, Z., Searchinger, T. D., Ziegler, A. D., Wu, J., Wang, D., He, X., Elsen, P. R., Ciais, P., Xu, R., Guo, Z., Peng, L., Tao, Y., Spracklen, D. V., Holden, J., Liu, X., Zheng, Y., Xu, P., Chen, J., … Zheng, C. (2022). Doubling of annual forest carbon loss over the tropics during the early twenty-first century. *Nature Sustainability*, *5*(5), 444–451.

Féret, J.-B., & Asner, G. P. (2011). Spectroscopic classification of tropical forest species using radiative transfer modeling. *Remote Sens. Environ.*, *115*(9), 2415–2422.

Féret, J.-B., & Asner, G. P. (2013). Tree species discrimination in tropical forests using airborne imaging spectroscopy. *IEEE Trans. Geosci. Remote Sens.*, *51*(1), 73–84.

Feurer, D., & Vinatier, F. (2018a). Joining multi-epoch archival aerial images in a single SfM block allows 3-D change detection with almost exclusively image information. *ISPRS J. Photogramm. Remote Sens.*, *146*, 495–506.

Feurer, D., & Vinatier, F. (2018b). The Time-SIFT method : Detecting 3-D changes from archival photogrammetric analysis with almost exclusively image information arXiv 1807.09700.

Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*, *281*(5374), 237–240.

Filella, I., & Penuelas, J. (1994). The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *Int. J. Remote Sens.*, *15*(7), 1459–1470.

Finer, M., Babbitt, B., Novoa, S., Ferrarese, F., Pappalardo, S. E., De Marchi, M., Saucedo, M., & Kumar, A. (2015). Future of oil and gas development in the western Amazon. *Environ. Res. Lett.*, *10*(2), 024003.

Finer, M., Jenkins, C. N., Pimm, S. L., Keane, B., & Ross, C. (2008). Oil and gas projects in the Western Amazon: Threats to wilderness, biodiversity, and indigenous peoples. *PLoS One*, *3*(8), e2932.

Finer, M., Jenkins, C. N., Sky, M. A. B., & Pine, J. (2014). Logging concessions enable illegal logging crisis in the Peruvian Amazon. *Sci. Rep.*, *4*, 4719.

Fisher, R. A., Koven, C. D., Anderegg, W. R. L., Christoffersen, B. O., Dietze, M. C., Farrior, C. E., Holm, J. A., Hurtt, G. C., Knox, R. G., Lawrence, P. J., Lichstein, J. W., Longo, M., Matheny, A. M., Medvigy, D., Muller-Landau, H. C., Powell, T. L., Serbin, S. P., Sato, H., Shuman, J. K., … Moorcroft, P. R. (2018). Vegetation demographics in earth system models: A review of progress and priorities. *Global Change Biology*, *24*(1), https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.13910, 35–54. https://doi.org/10.1111/gcb.13910

ForestPlots.net, Blundo, C., Carilla, J., Grau, R., Malizia, A., Malizia, L., Osinaga-Acosta, O., Bird, M., Bradford, M., Catchpole, D., Ford, A., Graham, A., Hilbert, D., Kemp, J., Laurance, S., Laurance, W., Ishida, F. Y., Marshall, A., Waite, C., … Tran, H. D. (2021). Taking the pulse of earth's tropical forests using networks of highly distributed plots. *Biological Conservation*, *260*, 108849. https://doi.org/10.1016/j.biocon.2020.108849

Forrester, D. I., & Bauhus, J. (2016). A review of processes behind Diversity—Productivity relationships in forests. *Current Forestry Reports*, *2*(1), 45–61.

França, F. M., Benkwitt, C. E., Peralta, G., Robinson, J. P. W., Graham, N. A. J., Tylianakis, J. M., Berenguer, E., Lees, A. C., Ferreira, J., Louzada, J., & Barlow, J. (2020). Climatic and local stressor interactions threaten tropical forests and coral reefs. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *375*(1794), 20190116.

Frankie, G. W., Baker, H. G., & Opler, P. A. (1974). Comparative phenological studies of trees in tropical wet and dry forests in the lowlands of Costa Rica. *J. Ecol.*, *62*(3), 881–919.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Alkama, R., … Zheng, B. (2022). Global carbon budget 2022. *Earth Syst. Sci. Data*, *14*(11), 4811–4900.

Gallup, S. M., Baker, I. T., Gallup, J. L., Restrepo-Coupe, N., Haynes, K. D., Geyer, N. M., & Denning, A. S. (2021). Accurate simulation of both sensitivity and variability for Amazonian photosynthesis: Is it too much to ask? *Journal of Advances in Modeling Earth Systems*, *13*(8), e2021MS002555. https://doi.org/10.1029/2021MS002555

Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., Ruiz-Jaen, M. C., Fröberg, M., Stendahl, J., Philipson, C. D., Mikusiński, G., Andersson, E., Westerlund,

B., Andrén, H., Moberg, F., Moen, J., & Bengtsson, J. (2013). Higher levels of multiple ecosystem services are found in forests with more tree species. *Nat. Commun.*, *4*, 1340.

Gan, Y., Wang, Q., & Iio, A. (2023). Tree crown detection and delineation in a temperate deciduous forest from UAV RGB imagery using deep learning approaches: Effects of spatial resolution and species characteristics. *Remote Sensing*, *15*(3).

Gao, B.-C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.*, *58*(3), 257–266.

Garzon-Lopez, C. X., & Lasso, E. (2020). Species classification in a tropical alpine ecosystem using UAV-Borne RGB and hyperspectral imagery. *Drones*, *4*(4), 69.

Gatti, L. V., Basso, L. S., Miller, J. B., Gloor, M., Gatti Domingues, L., Cassol, H. L. G., Tejada, G., Aragão, L. E. O. C., Nobre, C., Peters, W., Marani, L., Arai, E., Sanches, A. H., Corrêa, S. M., Anderson, L., Von Randow, C., Correia, C. S. C., Crispim, S. P., & Neves, R. A. L. (2021). Amazonia as a carbon source linked to deforestation and climate change. *Nature*, *595*(7867), 388–393.

Gatti, R. C., Reich, P. B., Gamarra, J. G. P., Crowther, T., Hui, C., Morera, A., Bastin, J.-F., de-Miguel, S., Nabuurs, G.-J., Svenning, J.-C., Serra-Diaz, J. M., Merow, C., Enquist, B., Kamenetsky, M., Lee, J., Zhu, J., Fang, J., Jacobs, D. F., Pijanowski, B., … Liang, J. (2022). The number of tree species on Earth. *Proceedings of the National Academy of Sciences*, *119*(6), e2115329119.

Gaubert, B., Stephens, B. B., Basu, S., Chevallier, F., Deng, F., Kort, E. A., Patra, P. K., Peters, W., Rödenbeck, C., Saeki, T., Schimel, D., Van der Laan-Luijkx, I., Wofsy, S., & Yin, Y. (2019). Global atmospheric CO_2 inverse models converging on neutral tropical land exchange, but disagreeing on fossil fuel and atmospheric growth rate. *Biogeosciences*, *16*(1), 117–134.

Geist, H. J., & Lambin, E. F. (2002). Proximate causes and underlying driving forces of tropical deforestation. *Bioscience*, *52*(2), 143–150.

Ghiyamat, A., & Shafri, H. Z. M. (2010). A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment. *Int. J. Remote Sens.*, *31*(7), 1837–1856.

Giam, X. (2017). Global biodiversity loss from tropical deforestation. *Proc. Natl. Acad. Sci. U. S. A.*, *114*(23), 5775–5777.

Gibbs, H. K., Ruesch, A. S., Achard, F., Clayton, M. K., Holmgren, P., Ramankutty, N., & Foley, J. A. (2010). Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proc. Natl. Acad. Sci. U. S. A.*, *107*(38), 16732–16737.

Giljum, S., Maus, V., Kuschnig, N., Luckeneder, S., Tost, M., Sonter, L. J., & Bebbington, A. J. (2022). A pantropical assessment of deforestation caused by industrial mining. *Proc. Natl. Acad. Sci. U. S. A.*, *119*(38), e2118273119.

Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.*, *160*(3), 271–282.

Givnish, T. J. (1999). On the causes of gradients in tropical tree diversity. *J. Ecol.*, *87*(2), 193–210.

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks, In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings.

Goetz, S. J., Baccini, A., Laporte, N. T., Johns, T., Walker, W., Kellndorfer, J., Houghton, R. A., & Sun, M. (2009). Mapping and monitoring carbon stocks with satellite observations: A comparison of methods. *Carbon Balance Manag.*, *4*, 2.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Gora, E. M., & Esquivel-Muelbert, A. (2021). Implications of size-dependent tree mortality for tropical forest carbon dynamics. *Nature Plants*, *7*, 384–391. https://doi.org/10.1038/s41477-021-00879-0

Gourlet-Fleury, S., Guehl, J.-M., & Laroussinie, O. (2004). *Ecology and management of a neotropical rainforest : Lessons drawn from Paracou, a long-term experimental research site in French Guiana*. Elsevier.

Grabska, E., Frantz, D., & Ostapowicz, K. (2020). Evaluation of machine learning algorithms for forest stand species mapping using sentinel-2 imagery and environmental data in the polish carpathians. *Remote Sensing of Environment*, *251*, 112103.

Griscom, B. W., Adams, J., Ellis, P. W., Houghton, R. A., Lomax, G., Miteva, D. A., Schlesinger, W. H., Shoch, D., Siikamäki, J. V., Smith, P., Woodbury, P., Zganjar, C., Blackman, A., Campari, J., Conant, R. T., Delgado, C., Elias, P., Gopalakrishna, T., Hamsik, M. R., … Fargione, J. (2017). Natural climate solutions. *Proc. Natl. Acad. Sci. U. S. A.*, *114*(44), 11645–11650.

Grubb, P. J. (1977). The maintenance of species-richness in plant communities: The importance of the regeneration niche. *Biological Reviews*, *52*(1), 107–145.

Guan, K., Pan, M., Li, H., Wolf, A., Wu, J., Medvigy, D., Caylor, K. K., Sheffield, J., Wood, E. F., Malhi, Y., Liang, M., Kimball, J. S., Saleska, S. R., Berry, J., Joiner, J., & Lyapustin, A. I. (2015). Photosynthetic seasonality of global tropical forests constrained by hydroclimate. *Nat. Geosci.*, *8*(4), 284–289.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., … Sang, B. (2015). The EnMAP spaceborne imaging spectroscopy mission for Earth observation. *Remote Sensing*, *7*(7), 8830–8857.

Guisan, A., Weiss, S. B., & Weiss, A. D. (1999). GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol.*, *143*(1), 107–122.

Hall, D. O., & Rao, K. (1999). *Photosynthesis*. Cambridge University Press.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–853.

Hansen, M. C., Krylov, A., Tyukavina, A., Potapov, P. V., Turubanova, S., Zutta, B., Ifo, S., Margono, B., Stolle, F., & Moore, R. (2016). Humid tropical forest disturbance alerts using Landsat data. *Environ. Res. Lett.*, *11*(3), 034008.

Hao, Z., Lin, L., Post, C. J., Mikhailova, E. A., Li, M., Chen, Y., Yu, K., & Liu, J. (2021). Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (Mask R-CNN). *ISPRS Journal of Photogrammetry and Remote Sensing*, *178*, 112–123.

Harenčár, J. G., Ávila-Lovera, E., Goldsmith, G. R., Chen, G. F., & Kay, K. M. (2022). Flexible drought deciduousness in a neotropical understory herb. *Am. J. Bot.*, *109*(8), 1262–1272.

Harris, N. L., Brown, S., Hagen, S. C., Saatchi, S. S., Petrova, S., Salas, W., Hansen, M. C., Potapov, P. V., & Lotsch, A. (2012). Baseline map of carbon emissions from deforestation in tropical regions. *Science*, *336*(6088), 1573–1576.

Harris, N. L., Gibbs, D. A., Baccini, A., Birdsey, R. A., de Bruin, S., Farina, M., Fatoyinbo, L., Hansen, M. C., Herold, M., Houghton, R. A., Potapov, P. V., Suarez, D. R., Roman-Cuesta, R. M., Saatchi, S. S., Slay, C. M., Turubanova, S. A., & Tyukavina, A. (2021). Global maps of twenty-first century forest carbon fluxes. *Nat. Clim. Chang.*, *11*(3), 234–240.

Harris, N. L., Goldman, E., Gabris, C., Nordling, J., Minnemeyer, S., Ansari, S., Lippmann, M., Bennett, L., Raad, M., Hansen, M., & Potapov, P. (2017). Using spatial statistics to identify emerging hot spots of forest loss. *Environ. Res. Lett.*, *12*(2), 024012.

Harrison, D., Rivard, B., & Sánchez-Azofeifa, G. A. (2018). Classification of tree species based on longwave hyperspectral data from leaves, a case study for a tropical dry forest. *Int. J. Appl. Earth Obs. Geoinf.*, *66*, 93–105.

Hastings, J. H., Ollinger, S. V., Ouimette, A. P., Sanders-DeMott, R., Palace, M. W., Ducey, M. J., Sullivan, F. B., Basler, D., & Orwig, D. A. (2020). Tree species traits determine the success of LiDAR-Based crown mapping in a mixed temperate forest. *Remote Sensing*, *12*(2), 309.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN arXiv 1703.06870.

He, K., Zhang, X., Ren, S., & Sun, J. (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition, In *Computer vision – ECCV 2014*, Springer International Publishing.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition, In *Proceedings of the ieee conference on computer vision and pattern recognition*.

Hecht, S. B., & Cockburn, A. (2010). *The fate of the forest: Developers, destroyers, and defenders of the Amazon, updated edition*. University of Chicago Press.

Hennessy, A., Clarke, K., & Lewis, M. (2020). Hyperspectral classification of plants: A review of waveband selection generalisability. *Remote Sensing*, *12*(1), 113.

Hesketh, M., & Sánchez-Azofeifa, G. A. (2012). The effect of seasonal spectral variation on species classification in the Panamanian tropical forest. *Remote Sens. Environ.*, *118*, 73–82.

Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. *Am. Nat.*, *163*(2), 192–211.

Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines vinod nair.

Hoban, S., Bruford, M. W., Funk, W. C., Galbusera, P., Griffith, M. P., Grueber, C. E., Heuertz, M., Hunter, M. E., Hvilsom, C., Stroil, B. K., Kershaw, F., Khoury, C. K., Laikre, L., Lopes-Fernandes, M., MacDonald, A. J., Mergeay, J., Meek, M., Mittan, C., Mukassabi, T. A., … Vernesi, C. (2021). Global commitments to conserving and monitoring genetic diversity are now necessary and feasible. *Bioscience*, *71*(9), 964–976.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*, 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735

Hoeser, T., & Kuenzer, C. (2020). Object detection and image segmentation with deep learning on Earth observation data: A Review-Part i: Evolution and recent trends. *Remote Sensing*, *12*(10), 1667.

Hoffman, G. (2018). Introduction to LSTMs with TensorFlow [Accessed: 2023-10-25].

Hong, S., Kim, S., Joh, M., & Song, S.-k. (2017). Psique: Next sequence prediction of satellite images using a convolutional sequence-to-sequence network.

Horler, D. N. H., Dockray, M., & Barber, J. (1983). The red edge of plant leaf reflectance. *Int. J. Remote Sens.*, *4*(2), 273–288.

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset Large-Scale knowledge of biodiversity. *Annu. Rev. Ecol. Evol. Syst.*, *46*(Volume 46, 2015), 523–549.

Hosonuma, N., Herold, M., De Sy, V., De Fries, R. S., Brockhaus, M., Verchot, L., Angelsen, A., & Romijn, E. (2012). An assessment of deforestation and forest degradation drivers in developing countries. *Environ. Res. Lett.*, *7*(4), 044009.

Houghton, R. A., House, J. I., Pongratz, J., van der Werf, G. R., DeFries, R. S., Hansen, M. C., Le Quéré, C., & Ramankutty, N. (2012). Carbon emissions from land use and land-cover change. *Biogeosciences*, *9*(12), 5125–5142.

Hu, G., Wang, T., Wan, M., Bao, W., & Zeng, W. (2022). UAV remote sensing monitoring of pine forest diseases based on improved Mask R-CNN. *International Journal of Remote Sensing*, *43*(4), 1274–1305. https://doi.org/10.1080/01431161.2022.2032455

Hubau, W., Lewis, S. L., Phillips, O. L., Affum-Baffoe, K., Beeckman, H., Cuní-Sanchez, A., Daniels, A. K., Ewango, C. E. N., Fauset, S., Mukinzi, J. M., Sheil, D., Sonké, B., Sullivan, M. J. P., Sunderland, T. C. H., Taedoumg, H., Thomas, S. C., White, L. J. T., Abernethy, K. A., Adu-Bredu, S., … Zemagho, L. (2020). Asynchronous carbon sink saturation in African and Amazonian tropical forests. *Nature*, *579*, 80–87. https://doi.org/10.1038/s41586-020-2035-0

Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Princeton University Press.

Huertas, C., Sabatier, D., Derroire, G., Ferry, B., Jackson, T., Pélissier, R., & Vincent, G. (2022). Mapping tree mortality rate in a tropical moist forest using multi-temporal lidar. *International Journal of Applied Earth Observation and Geoinformation*, *109*, 102780. https://doi.org/10.1016/j.jag.2022.102780

Huete, A. R., Didan, K., Shimabukuro, Y. E., Ratana, P., Saleska, S. R., Hutyra, L. R., Yang, W., Nemani, R. R., & Myneni, R. (2006). Amazon rainforests green-up with sunlight in dry season. *Geophys. Res. Lett.*, *33*(6).

Huntingford, C., Fisher, R. A., Mercado, L., Booth, B. B. B., Sitch, S., Harris, P. P., Cox, P. M., Jones, C. D., Betts, R. A., Malhi, Y., Harris, G. R., Collins, M., & Moorcroft, P. (2008). Towards quantifying uncertainty in predictions of Amazon 'dieback'. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *363*(1498), 1857–1864.

Hurst, J. M., Allen, R. B., Coomes, D. A., & Duncan, R. P. (2011). Size-specific tree mortality varies with neighbourhood crowding and disturbance in a montane nothofagus forest. *PloS one*, *6*(10), e26670.

Ienco, D., Gaetano, R., Dupaquier, C., & Maurel, P. (2017). Land cover classification via multitemporal spatial data by deep recurrent neural networks. *IEEE Geoscience and*

Remote Sensing Letters, *14*(10), 1685–1689. https://doi.org/10.1109/LGRS.2017.2728698

Iglhaut, J., Cabo, C., Puliti, S., Piermattei, L., O'Connor, J., & Rosette, J. (2019). Structure from motion photogrammetry in forestry: A review. *Current Forestry Reports*, *5*, 155–168. https://doi.org/10.1007/s40725-019-00094-3

Iida, Y., Poorter, L., Sterck, F., Kassim, A. R., Potts, M. D., Kubo, T., & Kohyama, T. S. (2014). Linking size-dependent growth and mortality with architectural traits across 145 co-occurring tropical tree species. *Ecology*, *95*(2), 353–363.

Interdonato, R., Ienco, D., Gaetano, R., & Ose, K. (2019). DuPLO: A DUal view point deep learning architecture for time series classification. *ISPRS J. Photogramm. Remote Sens.*, *149*, 91–104.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

IPCC. (2021). *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press.

IPCC. (2022). Tropical forests. In *Climate change 2022 – impacts, adaptation and vulnerability: Working group II contribution to the sixth assessment report of the intergovernmental panel on climate change* (pp. 2369–2410). Cambridge University Press.

Isbell, F., Calcagno, V., Hector, A., Connolly, J., Harpole, W. S., Reich, P. B., Scherer-Lorenzen, M., Schmid, B., Tilman, D., van Ruijven, J., Weigelt, A., Wilsey, B. J., Zavaleta, E. S., & Loreau, M. (2011). High plant diversity is needed to maintain ecosystem services. *Nature*, *477*(7363), 199–202.

Itoh, A., Yamakura, T., Ohkubo, T., Kanzaki, M., Palmiotto, P., Tan, S., & Lee, H. S. (2003). Spatially aggregated fruiting in an emergent Bornean tree. *J. Trop. Ecol.*, *19*(5), 531–538.

Jablonski, D., Roy, K., & Valentine, J. W. (2006). Out of the tropics: Evolutionary dynamics of the latitudinal diversity gradient. *Science*, *314*(5796), 102–106.

Jacquemoud, S., & Baret, F. (1990). PROSPECT: A model of leaf optical properties spectra. *Remote Sens. Environ.*, *34*(2), 75–91.

Jactel, H., Bauhus, J., Boberg, J., Bonal, D., Castagneyrol, B., Gardiner, B., Gonzalez-Olabarria, J. R., Koricheva, J., Meurisse, N., & Brockerhoff, E. G. (2017). Tree diversity drives forest stand resistance to natural disturbances. *Current Forestry Reports*, *3*(3), 223–243.

Jalobeanu, A., & Gonçalves, G. R. (2014). Automated probabilistic LiDAR swath registration.

Jansson, G., & Angelstam, P. (1999). Threshold levels of habitat composition for the presence of the long-tailed tit (aegithalos caudatus) in a boreal landscape. *Landsc. Ecol.*, *14*(3), 283–290.

Jarecke, K. M., Bladon, K. D., & Wondzell, S. M. (2021). The influence of local and nonlocal factors on soil water content in a steep forested catchment. *Water Resour. Res.*, *57*(5).

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), https://science.sciencemag.org/content/353/6301/790.full.pdf, 790–794. https://doi.org/10.1126/science.aaf7894

Jenkins, C. N., Pimm, S. L., & Joppa, L. N. (2013). Global patterns of terrestrial vertebrate diversity and conservation. *Proc. Natl. Acad. Sci. U. S. A.*, *110*(28), E2602–10.

Ji, S., Zhang, C., Xu, A., Shi, Y., & Duan, Y. (2018). 3D convolutional neural networks for crop classification with Multi-Temporal remote sensing images. *Remote Sensing*, *10*(1), 75.

Jones, H. G., & Vaughan, R. A. (2010). *Remote sensing of vegetation: Principles, techniques, and applications*. OUP Oxford.

Jones, T. G., Coops, N. C., & Sharma, T. (2010). Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal pacific northwest, canada. *Remote Sens. Environ.*, *114*(12), 2841–2852.

Jucker, T., Bongalov, B., Burslem, D. F. R. P., Nilus, R., Dalponte, M., Lewis, S. L., Phillips, O. L., Qie, L., & Coomes, D. A. (2018). Topography shapes the structure, composition and function of tropical forest landscapes. *Ecol. Lett.*, *21*(7), 989–1000.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589.

Kalacska, M., & Sánchez-Azofeifa, G. A. (2008). *Hyperspectral remote sensing of tropical and Sub-Tropical forests*. CRC Press.

Kamoske, A. G., Dahlin, K. M., Read, Q. D., Record, S., Stark, S. C., Serbin, S. P., Zarnetske, P. L., & Dornelas, M. (2022). Towards mapping biodiversity from above: Can fusing lidar and hyperspectral remote sensing predict taxonomic, functional, and phylogenetic tree diversity in temperate forests? *Glob. Ecol. Biogeogr.*, *31*(7), 1440–1460.

Kanowski, P. J., McDermott, C. L., & Cashore, B. W. (2011). Implementing REDD+: Lessons from analysis of forest governance. *Environ. Sci. Policy*, *14*(2), 111–117.

Kaproth, M. A., Fredericksen, B. W., González-Rodriguez, A., Hipp, A. L., & Cavender-Bares, J. (2023). Drought response strategies are coupled with leaf habit in 35 evergreen and

deciduous oak (Quercus) species across a climatic gradient in the americas. *New Phytol.*, *239*(3), 888–904.

Karageorgou, P., & Manetas, Y. (2006). The importance of being red when young: Anthocyanins and the protection of young leaves of Quercus coccifera from insect herbivory and excess light. *Tree Physiol.*, *26*(5), 613–621.

Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.*, *173*, 24–49.

Kattenborn, T., Schiefer, F., Frey, J., Feilhauer, H., Mahecha, M. D., & Dormann, C. F. (2022). Spatially autocorrelated training and validation samples inflate performance assessment of convolutional neural networks. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, *5*, 100018.

Keenan, R. J. (2015). Climate change impacts and adaptation in forest management: A review. *Ann. For. Sci.*, *72*(2), 145–167.

Keenan, T. F., Hollinger, D. Y., Bohrer, G., Dragoni, D., Munger, J. W., Schmid, H. P., & Richardson, A. D. (2013). Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise. *Nature*, *499*(7458), 324–327.

Kellner, J. R., Albert, L. P., Burley, J. T., & Cushman, K. C. (2019). The case for remote sensing of individual plants. *American Journal of Botany*, *106*(9), 1139–1142. https://doi.org/10.1002/ajb2.1347

Ken, S., Sasaki, N., Entani, T., Ma, H. O., Thuch, P., & Tsusaka, T. W. (2020). Assessment of the local perceptions on the drivers of deforestation and forest degradation, agents of drivers, and appropriate activities in cambodia. *Sustain. Sci. Pract. Policy*, *12*(23), 9987.

Kennedy, C. E. J., & Southwood, T. R. E. (1984). The number of species of insects associated with british trees: A Re-Analysis. *J. Anim. Ecol.*, *53*(2), 455–478.

Kier, G., Kreft, H., Lee, T. M., Jetz, W., Ibisch, P. L., Nowicki, C., Mutke, J., & Barthlott, W. (2009). A global assessment of endemism and species richness across island and mainland regions. *Proc. Natl. Acad. Sci. U. S. A.*, *106*(23), 9322–9327.

Kikuzawa, K. (1991). A Cost-Benefit analysis of leaf habit and leaf longevity of trees and their geographical pattern. *Am. Nat.*, *138*(5), 1250–1263.

Kim, S., Hong, S., Joh, M., & Song, S.-K. (2017). DeepRain: ConvLSTM network for precipitation prediction using multichannel radar data arXiv 1711.02316.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization arXiv 1412.6980.

Kinnebrew, E., Ochoa-Brito, J. I., French, M., Mills-Novoa, M., Shoffner, E., & Siegel, K. (2022). Biases and limitations of global forest change and author-generated land cover maps in detecting deforestation in the Amazon. *PLoS One*, *17*(7), e0268970.

Kislov, D. E., Korznikov, K. A., Altman, J., Vozmishcheva, A. S., & Krestov, P. V. (2021). Extending deep learning approaches for forest disturbance segmentation on very high-resolution satellite images. *Remote Sens. Ecol. Conserv.*, *7*(3), 355–368.

Kissinger, G. (2020). Policy responses to direct and underlying drivers of deforestation: Examining rubber and coffee in the central highlands of Vietnam. *For. Trees Livelihoods*, *11*(7), 733.

Kitajima, K., Mulkey, S., & Wright, S. (1997). Decline of photosynthetic capacity with leaf age in relation to leaf longevities for five tropical canopy tree species. *Am. J. Bot.*, *84*(5), 702.

Kljun, N., Calanca, P., Rotach, M. W., & Schmid, H. P. (2004). A simple parameterisation for flux footprint predictions. *Bound.-Layer Meteorol.*, *112*(3), 503–523.

Kljun, N., Calanca, P., Rotach, M. W., & Schmid, H. P. (2015). A simple two-dimensional parameterisation for flux footprint prediction (FFP). *Geosci. Model Dev.*, *8*(11), 3695–3713.

Knipling, E. B. (1970). Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment*, *1*(3), 155–159. https://doi.org/10.1016/S0034-4257(70)80021-9

Knyazikhin, Y., Schull, M. A., Stenberg, P., Mõttus, M., Rautiainen, M., Yang, Y., Marshak, A., Latorre Carmona, P., Kaufmann, R. K., Lewis, P., Disney, M. I., Vanderbilt, V., Davis, A. B., Baret, F., Jacquemoud, S., Lyapustin, A., & Myneni, R. B. (2013). Hyperspectral remote sensing of foliar nitrogen content. *Proc. Natl. Acad. Sci. U. S. A.*, *110*(3), E185–92.

Kokaly, R. F., Asner, G. P., Ollinger, S. V., Martin, M. E., & Wessman, C. A. (2009). Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sens. Environ.*, *113*, S78–S91.

Köppen, V. (1884). The heat zones of the Earth, viewed according to the duration of the hot, temperate and cold periods and according to the effect of the heat on the organic world. *Meteorologische Zeitschrift*, *1*(21), 5–226.

Koven, C. D., Knox, R. G., Fisher, R. A., Chambers, J. Q., Christoffersen, B. O., Davies, S. J., Detto, M., Dietze, M. C., Faybishenko, B., Holm, J. Et al. (2020). Benchmarking and parameter sensitivity of physiological and vegetation dynamics using the functionally assembled terrestrial ecosystem simulator (fates) at barro colorado island, panama. *Biogeosciences*, *17*(11), 3017–3044.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Commun. ACM*, *60*(6), 84–90.

Krogh, A. (2020). *State of the tropical rainforest* (tech. rep.). Rainforest Foundation Norway.

Kumar, K. S., Tripathi, S. K., & Khanduri, V. P. (2023). Phenological patterns of tropical trees in relation to climatic factors in a mountain moist forest of Indo-Burma hotspot region. *Vegetos*, *36*(3), 1070–1079.

Kunert, N., Zailaa, J., Herrmann, V., Muller-Landau, H. C., Wright, S. J., Pérez, R., McMahon, S. M., Condit, R. C., Hubbell, S. P., Sack, L., Davies, S. J., & Anderson-Teixeira, K. J. (2021). Leaf turgor loss point shapes local and regional distributions of evergreen but not deciduous tropical trees. *New Phytol.*, *230*(2), 485–496.

Kursar, T. A., Dexter, K. G., Lokvam, J., Pennington, R. T., Richardson, J. E., Weber, M. G., Murakami, E. T., Drake, C., McGregor, R., & Coley, P. D. (2009). The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus Inga. *Proc. Natl. Acad. Sci. U. S. A.*, *106*(43), 18073–18078.

Kushwaha, C. P., Tripathi, S. K., Tripathi, B. D., & Singh, K. P. (2011). Patterns of tree phenological diversity in dry tropics. *Acta Ecol. Sin.*, *31*(4), 179–185.

Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, *14*(5), 778–782.

Laliberté, E., Schweiger, A. K., & Legendre, P. (2020). Partitioning plant spectral diversity into alpha and beta components. *Ecol. Lett.*, *23*(2), 370–380.

Laurance, W. F. (2007). Have we overstated the tropical biodiversity crisis? *Trends Ecol. Evol.*, *22*(2), 65–70.

Laurans, M., Martin, O., Nicolini, E., & Vincent, G. (2012). Functional traits and their plasticity predict tropical trees regeneration niche even among species with intermediate light requirements. *J. Ecol.*, *100*(6), 1440–1452.

Laybros, A., Aubry-Kientz, M., Féret, J.-B., Bedeau, C., Brunaux, O., Derroire, G., & Vincent, G. (2020). Quantitative airborne inventories in dense tropical forest using imaging spectroscopy. *Remote Sensing*, *12*(10).

Laybros, A., Schläpfer, D., Féret, J.-B., Descroix, L., Bedeau, C., Lefevre, M.-J., & Vincent, G. (2019). Across date species detection using airborne imaging spectroscopy. *Remote Sensing*, *11*(7).

Lee, H. S., Davies, S. J., LaFrankie, J. V., Tan, S., Yamakura, T., Itoh, A., Ohkubo, T., & Ashton, P. S. (2002). Floristic and structural diversity of mixed Dipterocarp forest in Lambir Hills National Park, Sarawak, Malaysia. *J. Trop. For. Sci.*, *14*(3), 379–400.

Lefcheck, J. S., Byrnes, J. E. K., Isbell, F., Gamfeldt, L., Griffin, J. N., Eisenhauer, N., Hensel, M. J. S., Hector, A., Cardinale, B. J., & Duffy, J. E. (2015). Biodiversity enhances ecosystem multifunctionality across trophic levels and habitats. *Nat. Commun.*, *6*, 6936.

Leigh, E. G., Jr, Davidar, P., Dick, C. W., Puyravaud, J.-P., Terborgh, J., ter Steege, H., & Wright, S. J. (2004). Why do some tropical forests have so many species of trees? *bitr*, *36*(4), 447–473.

Levin, N., Kyba, C. C. M., Zhang, Q., Sánchez de Miguel, A., Román, M. O., Li, X., Portnov, B. A., Molthan, A. L., Jechow, A., Miller, S. D., Wang, Z., Shrestha, R. M., & Elvidge, C. D. (2020). Remote sensing of night lights: A review and an outlook for the future. *Remote Sens. Environ.*, *237*, 111443.

Lewis, S. L., Edwards, D. P., & Galbraith, D. (2015). Increasing human dominance of tropical forests. *Science*, *349*(6250), 827–832.

Lewis, S. L., Lloyd, J., Sitch, S., Mitchard, E. T. A., & Laurance, W. F. (2009). Changing ecology of tropical forests: Evidence and drivers. *Annu. Rev. Ecol. Evol. Syst.*, *40*(1), 529–549.

Lewis, S. L., & Maslin, M. A. (2015). Defining the anthropocene. *Nature*, *519*(7542), 171–180.

Li, J., Chen, X., Niklas, K. J., Sun, J., Wang, Z., Zhong, Q., Hu, D., & Cheng, D. (2022). A whole-plant economics spectrum including bark functional traits for 59 subtropical woody plant species. *J. Ecol.*, *110*(1), 248–261.

Li, Q., Chen, X., Yuan, W., Lu, H., Shen, R., Wu, S., Gong, F., Dai, Y., Liu, L., Sun, Q., Zhang, C., & Su, Y. (2021a). Remote sensing of seasonal climatic constraints on leaf phenology across pantropical evergreen forest biome. *Earths Future*, *9*(9).

Li, X., Chen, S., Hu, X., & Yang, J. (2018). Understanding the disharmony between dropout and batch normalization by variance shift. *arXiv preprint arXiv:1801.05134*.

Li, X., Xiong, H., Li, X., Wu, X., Zhang, X., Liu, J., Bian, J., & Dou, D. (2021b). Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond arXiv 2103.10689.

Li, Y., Zhang, H., & Shen, Q. (2017). Spectral–Spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sensing*, *9*(1), 67.

Lieberman, D., & Lieberman, M. (1984). The causes and consequences of synchronous flushing in a dry tropical forest. *Biotropica*, *16*(3), 193–201.

Lieth, H. (1974). *Phenology and seasonality modeling* (Vol. 8). Springer Berlin Heidelberg.

Lim, C. L., Prescott, G. W., De Alban, J. D. T., Ziegler, A. D., & Webb, E. L. (2017). Untangling the proximate causes and underlying drivers of deforestation and forest degradation in Myanmar. *Conserv. Biol.*, *31*(6), 1362–1372.

Lima, A. L. A. d., Rodal, M. J. N., Castro, C. C., Antonino, A. C. D., Melo, A. L. d., Gonçalves-Souza, T., & Sampaio, E. V. d. S. B. (2021). Phenology of high- and low-density wood deciduous species responds differently to water supply in tropical semiarid regions. *J. Arid Environ.*, *193*, 104594.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection, In *Proceedings of the ieee conference on computer vision and pattern recognition*.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context, In *European conference on computer vision*. Springer.

Liu, F., & Kong, Y. (2015). Zoib: An R package for bayesian inference for beta regression and zero/one inflated beta regression. *R J.*, *7*(2), 34.

Liu, F., & Li, Q. (2016). A bayesian model for joint analysis of multivariate repeated measures and time to event data in crossover trials. *Stat. Methods Med. Res.*, *25*(5), 2180–2192.

Liu, Y., Zhan, Z., Ren, L., Ze, S., Yu, L., Jiang, Q., & Luo, Y. (2021). Hyperspectral evidence of early-stage pine shoot beetle attack in yunnan pine. *For. Ecol. Manage.*, *497*, 119505.

Longo, M., Knox, R. G., Medvigy, D. M., Levine, N. M., Dietze, M. C., Kim, Y., Swann, A. L. S., Zhang, K., Rollinson, C. R., Bras, R. L., Wofsy, S. C., & Moorcroft, P. R. (2019). The biophysics, ecology, and biogeochemistry of functionally diverse, vertically and horizontally heterogeneous ecosystems: The ecosystem demography model, version 2.2 – part 1: Model description. *Geoscientific Model Development*, *12*(10), 4309–4346.

Lopes, A. P., Nelson, B. W., Wu, J., Graça, P. M. L. d. A., Tavares, J. V., Prohaska, N., Martins, G. A., & Saleska, S. R. (2016). Leaf flush drives dry season green-up of the Central Amazon. *Remote Sens. Environ.*, *182*, 90–98.

Loreau, M., & Hector, A. (2001). Partitioning selection and complementarity in biodiversity experiments. *Nature*, *412*(6842), 72–76.

Loreau, M., & de Mazancourt, C. (2013). Biodiversity and ecosystem stability: A synthesis of underlying mechanisms. *Ecol. Lett.*, *16 Suppl 1*, 106–115.

Loubry, D. (1994). La phénologie des arbres caducifoliés en forêt guyanaise (5° de latitude nord): Illustration d'un déterminisme à composantes endogène et exogène. *Canadian journal of botany*, *72*(12), 1943–1957.

Louhaichi, M., Borman, M. M., & Johnson, D. E. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto Int.*, *16*(1), 65–70.

Lovejoy, T. E., & Nobre, C. (2019). Amazon tipping point: Last chance for action. *Sci Adv*, *5*(12), eaba2949.

Lovenduski, N. S., & Bonan, G. B. (2017). Reducing uncertainty in projections of terrestrial carbon uptake. *Environ. Res. Lett.*, *12*(4), 044020.

Lutz, J. A., Furniss, T. J., Johnson, D. J., Davies, S. J., Allen, D., Alonso, A., Anderson-Teixeira, K. J., Andrade, A., Baltzer, J., Becker, K. M. Et al. (2018). Global importance of large-diameter trees. *Global Ecology and Biogeography*, *27*(7), 849–864.

Lyapustin, A. I., Wang, Y., Laszlo, I., Hilker, T., Hall, F. G., Sellers, P. J., Tucker, C. J., & Korkin, S. V. (2012). Multi-angle implementation of atmospheric correction for MODIS (MAIAC): 3. atmospheric correction. *Remote Sens. Environ.*, *127*, 385–393.

Lyu, H., & Lu, H. (2016). Learning a transferable change detection method by recurrent neural network. https://doi.org/10.1109/IGARSS.2016.7730344

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.*, *152*, 166–177.

Ma, Y., Zhao, Y., Im, J., Zhao, Y., & Zhen, Z. (2024). A deep-learning-based tree species classification for natural secondary forests using unmanned aerial vehicle hyperspectral images and LiDAR. *Ecol. Indic.*, *159*, 111608.

Madritch, M. D., Kingdon, C. C., Singh, A., Mock, K. E., Lindroth, R. L., & Townsend, P. A. (2014). Imaging spectroscopy links aspen genotype with below-ground processes at landscape scales. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *369*(1643), 20130194.

Maire, V., Gross, N., Börger, L., Proulx, R., Wirth, C., Pontes, L. d. S., Soussana, J.-F., & Louault, F. (2012). Habitat filtering and niche differentiation jointly explain species relative abundance within grassland communities along fertility and disturbance gradients. *New Phytol.*, *196*(2), 497–509.

Malhi, Y. (2012). The productivity, metabolism and carbon cycle of tropical forest vegetation. *J. Ecol.*, *100*(1), 65–75.

Malhi, Y., Aragão, L. E. O. C., Galbraith, D., Huntingford, C., Fisher, R., Zelazowski, P., Sitch, S., McSweeney, C., & Meir, P. (2009). Exploring the likelihood and mechanism of a climate-change-induced dieback of the Amazon rainforest. *Proc. Natl. Acad. Sci. U. S. A.*, *106*(49), 20610–20615.

Malhi, Y., Roberts, J. T., Betts, R. A., Killeen, T. J., Li, W., & Nobre, C. A. (2008). Climate change, deforestation, and the fate of the Amazon. *Science*, *319*(5860), 169–172.

Manoli, G., Ivanov, V. Y., & Fatichi, S. (2018). Dry-season greening and water stress in Amazonia: The role of modeling leaf phenology. *J. Geophys. Res. Biogeosci.*, *123*(6), 1909–1926.

Marconi, S., Weinstein, B. G., Zou, S., Bohlman, S. A., Zare, A., Singh, A., Stewart, D., Harmon, I., Steinkraus, A., & White, E. P. (2022). Continental-scale hyperspectral

tree species classification in the united states national ecological observatory network. *Remote Sens. Environ.*, *282*, 113264.

Maréchaux, I., & Chave, J. (2017). An individual-based forest model to jointly simulate carbon and tree diversity in Amazonia: Description and applications. *Ecol. Monogr.*, *87*(4), 632–664.

Markesteijn, L., Poorter, L., Bongers, F., Paz, H., & Sack, L. (2011). Hydraulics and life history of tropical dry forest tree species: Coordination of species' drought and shade tolerance. *New Phytol.*, *191*(2), 480–495.

Martin, F.-M., Müllerová, J., Borgniet, L., Dommanget, F., Breton, V., & Evette, A. (2018). Using single- and Multi-Date UAV and satellite imagery to accurately monitor invasive knotweed species. *Remote Sensing*, *10*(10), 1662.

Marvin, D. C., Asner, G. P., Knapp, D. E., Anderson, C. B., Martin, R. E., Sinca, F., & Tupayachi, R. (2014). Amazonian landscapes and the bias in field studies of forest structure and biomass. *Proceedings of the National Academy of Sciences*, *111*(48), https://www.pnas.org/doi/pdf/10.1073/pnas.1412999111, E5224–E5232. https://doi.org/10.1073/pnas.1412999111

Mauricio, J., Domingues, I., & Bernardino, J. (2023). Comparing vision transformers and convolutional neural networks for image classification: A literature review. *NATO Adv. Sci. Inst. Ser. E Appl. Sci.*, *13*(9), 5521.

Mayfield, H., Smith, C., Gallagher, M., & Hockings, M. (2017). Use of freely available datasets and machine learning methods in predicting deforestation. *Environmental Modelling & Software*, *87*, 17–28.

Mäyrä, J., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Hurskainen, P., Kullberg, P., Poikolainen, L., Viinikka, A., Tuominen, S., Kumpula, T., & Vihervaara, P. (2021). Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sens. Environ.*, *256*, 112322.

Mazlan, S. M., Wan Mohd Jaafar, W. S., Muhmad Kamarulzaman, A. M., Saad, S. N. M., Mohd Ghazali, N., Adrah, E., Abdul Maulud, K. N., Omar, H., Teh, Y. A., Dzulkifli, D., & Mahmud, M. R. (2022). A review on the use of LiDAR remote sensing for forest landscape restoration. In M. N. Suratman (Ed.), *Concepts and applications of remote sensing in forestry* (pp. 49–74). Singapore, Springer Nature Singapore.

McCann, K. S. (2000). The diversity-stability debate. *Nature*, *405*(6783), 228–233.

McGrath, L. J., van Riper, C., 3rd, & Fontaine, J. J. (2009). Flower power: Tree flowering phenology as a settlement cue for migrating birds. *J. Anim. Ecol.*, *78*(1), 22–30.

McManus, K. M., Asner, G. P., Martin, R. E., Dexter, K. G., Kress, W. J., & Field, C. B. (2016). Phylogenetic structure of foliar spectral traits in tropical forest canopies. *Remote Sensing*, *8*(3), 196.

Meakem, V., Tepley, A. J., Gonzalez-Akre, E. B., Herrmann, V., Muller-Landau, H. C., Wright, S. J., Hubbell, S. P., Condit, R., & Anderson-Teixeira, K. J. (2018). Role of tree size in moist tropical forest carbon cycling and water deficit responses. *New Phytologist*, *219*(3), 947–958. https://doi.org/10.1111/nph.14633

Meireles, J. E., Cavender-Bares, J., Townsend, P. A., Ustin, S., Gamon, J. A., Schweiger, A. K., Schaepman, M. E., Asner, G. P., Martin, R. E., Singh, A., Schrodt, F., Chlus, A., & O'Meara, B. C. (2020). Leaf reflectance spectra capture the evolutionary history of seed plants. *New Phytol.*, *228*(2), 485–493.

Mena, C. F., Laso, F., Martinez, P., & Sampedro, C. (2017). Modeling road building, deforestation and carbon emissions due deforestation in the Ecuadorian Amazon: The potential impact of oil frontier growth. *J. Land Use Sci.*, *12*(6), 477–492.

Méndez-Alonzo, R., Pineda-Garcia, F., Paz, H., Rosell, J. A., & Olson, M. E. (2013). Leaf phenology is associated with soil water availability and xylem traits in a tropical dry forest. *Trees*, *27*(3), 745–754.

Midgley, J. J. (2003). Is bigger better in plants? The hydraulic costs of increasing size in trees. *Trends Ecol. Evol.*, *18*(1), 5–6.

Mitchard, E. T. A. (2018). The tropical forest carbon cycle and climate change. *Nature*, *559*(7715), 527–534.

Mitchard, E. T. A., Feldpausch, T. R., Brienen, R. J. W., Lopez-Gonzalez, G., Monteagudo, A., Baker, T. R., Lewis, S. L., Lloyd, J., Quesada, C. A., Gloor, M., Ter Steege, H., Meir, P., Alvarez, E., Araujo-Murakami, A., Aragão, L. E. O. C., Arroyo, L., Aymard, G., Banki, O., Bonal, D., … Phillips, O. L. (2014). Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. *Glob. Ecol. Biogeogr.*, *23*(8), 935–946.

Miyamoto, M., Mohd Parid, M., Noor Aini, Z., & Michinaka, T. (2014). Proximate and underlying causes of forest cover change in Peninsular Malaysia. *For. Policy Econ.*, *44*, 18–25.

Moffette, F., Alix-Garcia, J., Shea, K., & Pickens, A. H. (2021). The impact of near-real-time deforestation alerts across the tropics. *Nat. Clim. Chang.*, *11*(2), 172–178.

Molina-Venegas, R., & Rodriguez, M. A. (2017). Revisiting phylogenetic signal; strong or negligible impacts of polytomies and branch length information? *BMC Evol. Biol.*, *17*(1), 53.

Montes, R., & Ureña, C. (2012). An overview of BRDF models. *University of Grenada, Technical Report LSI-2012*, *1*, 19.

Morel, H., Mangenet, T., Beauchêne, J., Ruelle, J., Nicolini, E., Heuret, P., & Thibaut, B. (2015). Seasonal variations in phenological traits: Leaf shedding and cambial activity in parkia nitida miq. and parkia velutina benoist (fabaceae) in tropical rainforest. *Trees*, *29*(4), 973–984.

Morellato, L. P. C., Alberton, B., Alvarado, S. T., Borges, B., Buisson, E., Camargo, M. G. G., Cancian, L. F., Carstensen, D. W., Escobar, D. F. E., Leite, P. T. P., Mendoza, I., Rocha, N. M. W. B., Soares, N. C., Silva, T. S. F., Staggemeier, V. G., Streher, A. S., Vargas, B. C., & Peres, C. A. (2016). Linking plant phenology to conservation biology. *Biol. Conserv.*, *195*, 60–72.

Morley, R. J. (2011). Cretaceous and Tertiary climate change and the past distribution of megathermal rainforests. In M. Bush, J. Flenley, & W. Gosling (Eds.), *Tropical rainforest responses to climatic change* (pp. 1–34). Berlin, Heidelberg, Springer Berlin Heidelberg.

Morley, R. J. Et al. (2000). *Origin and evolution of tropical rain forests*. John Wiley & Sons.

Morton, D. C., Nagol, J., Carabajal, C. C., Rosette, J., Palace, M., Cook, B. D., Vermote, E. F., Harding, D. J., & North, P. R. J. (2014). Amazon forests maintain consistent canopy structure and greenness during the dry season. *Nature*, *506*(7487), 221–224.

Mou, L., Bruzzone, L., & Zhu, X. X. (2019). Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(2), 924–935. https://doi.org/10.1109/tgrs.2018.2863224

Mugabowindekwe, M., Brandt, M., Chave, J., Reiner, F., Skole, D. L., Kariryaa, A., Igel, C., Hiernaux, P., Ciais, P., Mertz, O., Tong, X., Li, S., Rwanyiziri, G., Dushimiyimana, T., Ndoli, A., Uwizeyimana, V., Lillesø, J.-P. B., Gieseke, F., Tucker, C. J., … Fensholt, R. (2022). Nation-wide mapping of tree-level aboveground carbon stocks in rwanda. *Nature Climate Change*. https://doi.org/10.1038/s41558-022-01544-w

Muller-Landau, H. C., Condit, R. S., Chave, J., Thomas, S. C., Bohlman, S. A., Bunyavejchewin, S., Davies, S., Foster, R., Gunatilleke, S., Gunatilleke, N. Et al. (2006). Testing metabolic ecology theory for allometric scaling of tree size, growth and mortality in tropical forests. *Ecology letters*, *9*(5), 575–588.

Murali, K. S., & Sukumar, R. (1993). Leaf flushing phenology and herbivory in a tropical dry deciduous forest, southern india. *Oecologia*, *94*(1), 114–119.

Myneni, R. B., Yang, W., Nemani, R. R., Huete, A. R., Dickinson, R. E., Knyazikhin, Y., Didan, K., Fu, R., Negrón Juárez, R. I., Saatchi, S. S., Hashimoto, H., Ichii, K., Shabanov,

N. V., Tan, B., Ratana, P., Privette, J. L., Morisette, J. T., Vermote, E. F., Roy, D. P., ... Salomonson, V. V. (2007). Large seasonal swings in leaf area of Amazon rainforests. *Proc. Natl. Acad. Sci. U. S. A.*, *104*(12), 4820–4823.

Naeem, S., Bunker, D. E., Hector, A., Loreau, M., & Perrings, C. (2009). *Biodiversity, ecosystem functioning, and human wellbeing: An ecological and economic perspective*. Oxford University Press.

Newstrom, L. E., Frankie, G. W., & Baker, H. G. (1994). A new classification for plant phenology based on flowering patterns in lowland tropical rain forest trees at La Selva, Costa Rica. *Biotropica*, *26*(2), 141–159.

Newton, P., Miller, D. C., Byenkya, M. A. A., & Agrawal, A. (2016). Who are forest-dependent people? a taxo nomy to aid livelihood and land use decision-making in forested regions. *Land use policy*, *57*, 388–395.

Nicolau, A. P., Herndon, K., Flores-Anderson, A., & Griffin, R. (2019). A spatial pattern analysis of forest loss in the Madre de Dios region, Peru. *Environ. Res. Lett.*, *14*(12), 124045.

Nieding, R. (2023). *Unsupervised machine learning via feature extraction and clustering to classify tree species from High-Resolution UAV-based RGB image data* (Doctoral dissertation). Otto-Friedrich-Universität Bamberg, Germany.

Nilus, R., Maycock, C. R., Majalap-Lee, N., & Burslem, D. (2011). Nutrient limitation of tree seedling growth in three soil types found in sabah. *Journal of Tropical Forest Science*, 133–142.

Nkem, J. N., Somorin, O. A., Jum, C., Idinoba, M. E., Bele, Y. M., & Sonwa, D. J. (2013). Profiling climate change vulnerability of forest indigenous communities in the congo basin. *Mitigation and Adaptation Strategies for Global Change*, *18*(5), 513–533.

Noh, S.-H. (2021). Analysis of gradient vanishing of RNNs and performance comparison. *Information*, *12*(11), 442.

Nunes, M. H., Davey, M. P., & Coomes, D. A. (2017). On the challenges of using field spectroscopy to measure the impact of soil type on leaf traits. *Biogeosciences*, *14*(13), 3371–3385.

Ocer, N. E., Kaplan, G., Erdem, F., Matci, D. K., & Avdan, U. (2020). Tree extraction from multi-scale uav images using Mask R-CNN with FPN. *Remote Sensing Letters*, *11*(9), 847–856. https://doi.org/10.1080/2150704X.2020.1784491

Oliveira, R. S., Eller, C. B., Barros, F. d. V., Hirota, M., Brum, M., & Bittencourt, P. (2021). Linking plant hydraulics and the fast-slow continuum to understand resilience to drought in tropical ecosystems. *New Phytol.*, *230*(3), 904–923.

Ollinger, S. V. (2011). Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytol.*, *189*(2), 375–394.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., & Kassem, K. R. (2001). Terrestrial ecoregions of the world: A new map of life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *Bioscience*, *51*(11), 933–938.

Ordway, E. M., Asner, G. P., Burslem, D. F., Lewis, S. L., Nilus, R., Martin, R. E., O'Brien, M. J., Phillips, O. L., Qie, L., Vaughn, N. R. Et al. (2022). Mapping tropical forest functional variation at satellite remote sensing resolutions depends on key traits. *Communications Earth & Environment*, *3*(1), 247.

Ouédraogo, D.-Y., Fayolle, A., Gourlet-Fleury, S., Mortier, F., Freycon, V., Fauvet, N., Rabaud, S., Cornu, G., Bénédet, F., Gillet, J.-F., Oslisly, R., Doucet, J.-L., Lejeune, P., & Favier, C. (2016). The determinants of tropical forest deciduousness: Disentangling the effects of rainfall and geology in central Africa. *J. Ecol.*, *104*(4), 924–935.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, *401*(6756), 877–884.

Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., Phillips, O. L., Shvidenko, A., Lewis, S. L., Canadell, J. G., Ciais, P., Jackson, R. B., Pacala, S. W., McGuire, A. D., Piao, S., Rautiainen, A., Sitch, S., & Hayes, D. (2011). A large and persistent carbon sink in the world's forests. *Science*, *333*(6045), 988–993.

Park, J. Y., Muller-Landau, H. C., Lichstein, J. W., Rifai, S. W., Dandois, J. P., & Bohlman, S. A. (2019). Quantifying leaf phenology of individual trees and species in a tropical forest using unmanned aerial vehicle (UAV) images. *Remote Sensing*, *11*(13), 1534.

Parsons, S. A., Valdez-Ramirez, V., Congdon, R. A., & Williams, S. E. (2014). Contrasting patterns of litterfall seasonality and seasonal changes in litter decomposability in a tropical rainforest region. *Biogeosciences*, *11*(18), 5047–5056.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., van Ingen, C., Vuichard, N., Zhang, L., Amiro, B., Ammann, C., … Papale, D. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Sci Data*, *7*(1), 225.

Pau, S., Okin, G. S., & Gillespie, T. W. (2010). Asynchronous response of tropical forest leaf phenology to seasonal and El Niño-driven drought. *PLoS One*, *5*(6), e11325.

Pau, S., Wolkovich, E. M., Cook, B. I., Davies, T. J., Kraft, N. J. B., Bolmgren, K., Betancourt, J. L., & Cleland, E. E. (2011). Predicting phenology by integrating ecology, evolution and climate science. *Glob. Chang. Biol.*, *17*(12), 3633–3643.

Pausas, J. G., Austin, M. P., & Noble, I. R. (1997). A forest simulation model for predicting eucalypt dynamics and habitat quality for arboreal marsupials. *Ecol. Appl.*, *7*(3), 921–933.

Pereira, E. J. d. A. L., de Santana Ribeiro, L. C., da Silva Freitas, L. F., & de Barros Pereira, H. B. (2020). Brazilian policy and agribusiness damage the Amazon rainforest. *Land use policy*, *92*, 104491.

Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., Bruford, M. W., Brummitt, N., Butchart, S. H. M., Cardoso, A. C., Coops, N. C., Dulloo, E., Faith, D. P., Freyhof, J., Gregory, R. D., Heip, C., Höft, R., Hurtt, G., Jetz, W., … Wegmann, M. (2013). Ecology. essential biodiversity variables. *Science*, *339*(6117), 277–278.

Perz, S. G., Caldas, M. M., Arima, E., & Walker, R. J. (2007). Unofficial road building in the Amazon: Socioeconomic and biophysical explanations. *Dev. Change*, *38*(3), 529–551.

Phillips, O. L., Malhi, Y., Higuchi, N., Laurance, W. F., Nunez, P. V., Vasquez, R. M., Laurance, S. G., Ferreira, L. V., Stern, M., Brown, S., & Grace, J. (1998). Changes in the carbon balance of tropical forests: Evidence from long-term plots. *Science*, *282*(5388), 439–442.

Piao, S., Liu, Q., Chen, A., Janssens, I. A., Fu, Y., Dai, J., Liu, L., Lian, X., Shen, M., & Zhu, X. (2019). Plant phenology and global climate change: Current progresses and challenges. *Glob. Chang. Biol.*, *25*(6), 1922–1940.

Piao, S., Sitch, S., Ciais, P., Friedlingstein, P., Peylin, P., Wang, X., Ahlström, A., Anav, A., Canadell, J. G., Cong, N., Huntingford, C., Jung, M., Levis, S., Levy, P. E., Li, J., Lin, X., Lomas, M. R., Lu, M., Luo, Y., … Zeng, N. (2013). Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO2 trends. *Glob. Chang. Biol.*, *19*(7), 2117–2132.

Piotrowski, M. (2019). Nearing the tipping point: Drivers of deforestation in the Amazon Region. *Inter-American Dialogue: Washington, WA, USA*.

Piponiot, C., Anderson-Teixeira, K. J., Davies, S. J., Allen, D., Bourg, N. A., Burslem, D. F. R. P., Cárdenas, D., Chang-Yang, C.-H., Chuyong, G., Cordell, S., Dattaraja, H. S., Duque, Á., Ediriweera, S., Ewango, C., Ezedin, Z., Filip, J., Giardina, C. P., Howe, R., Hsieh, C.-F., … Muller-Landau, H. C. (2022). Distribution of biomass dynamics in relation to tree size in forests across the world. *New Phytologist*, *234*(5),

https://nph.onlinelibrary.wiley.com/doi/pdf/10.1111/nph.17995, 1664–1677. https://doi.org/10.1111/nph.17995

Plotkin, J. B., Potts, M. D., Yu, D. W., Bunyavejchewin, S., Condit, R., Foster, R., Hubbell, S., LaFrankie, J., Manokaran, N., Seng, L. H., Sukumar, R., Nowak, M. A., & Ashton, P. S. (2000). Predicting species diversity in tropical forests. *Proc. Natl. Acad. Sci. U. S. A.*, *97*(20), 10850–10854.

Poorter, L., Bongers, F., Aide, T. M., Almeyda Zambrano, A. M., Balvanera, P., Becknell, J. M., Boukili, V., Brancalion, P. H. S., Broadbent, E. N., Chazdon, R. L., Craven, D., de Almeida-Cortez, J. S., Cabral, G. A. L., de Jong, B. H. J., Denslow, J. S., Dent, D. H., DeWalt, S. J., Dupuy, J. M., Durán, S. M., … Rozendaal, D. M. A. (2016). Biomass resilience of neotropical secondary forests. *Nature*, *530*(7589), 211–214.

Potapov, P., Hansen, M. C., Laestadius, L., Turubanova, S., Yaroshenko, A., Thies, C., Smith, W., Zhuravleva, I., Komarova, A., Minnemeyer, S., & Esipova, E. (2017). The last frontiers of wilderness: Tracking loss of intact forest landscapes from 2000 to 2013. *Sci Adv*, *3*(1), e1600821.

Prospere, K., McLaren, K., & Wilson, B. (2014). Plant species discrimination in a tropical wetland using in situ hyperspectral data. *Remote Sensing*, *6*(9), 8494–8523.

Purves, D. W., & Turnbull, L. A. (2010). Different but equal: The implausible assumption at the heart of neutral theory. *J. Anim. Ecol.*, *79*(6), 1215–1225.

Qie, L., Lewis, S. L., Sullivan, M. J. P., Lopez-Gonzalez, G., Pickavance, G. C., Sunderland, T., Ashton, P., Hubau, W., Abu Salim, K., Aiba, S.-I., Banin, L. F., Berry, N., Brearley, F. Q., Burslem, D. F. R. P., Dančák, M., Davies, S. J., Fredriksson, G., Hamer, K. C., Hédl, R., … Phillips, O. L. (2017). Long-term carbon sink in Borneo's forests halted by drought and vulnerable to edge effects. *Nat. Commun.*, *8*(1), 1966.

Rand, T. A., Tylianakis, J. M., & Tscharntke, T. (2006). Spillover edge effects: The dispersal of agriculturally subsidized insect natural enemies into adjacent natural habitats. *Ecol. Lett.*, *9*(5), 603–614.

Reich, P. B. (1995). Phenology of tropical forests: Patterns, causes, and consequences. *Can. J. Bot.*, *73*(2), 164–174.

Reich, P. B. (2014). The world-wide 'fast–slow' plant economics spectrum: A traits manifesto. *J. Ecol.*, *102*(2), 275–301.

Reich, P. B., & Borchert, R. (1982). Phenology and ecophysiology of the tropical tree, tabebuia neochrysantha (bignoniaceae). *Ecology*, *63*(2), 294–299.

Reich, P. B., & Borchert, R. (1984). Water stress and tree phenology in a tropical dry forest in the lowlands of Costa Rica. *J. Ecol.*, *72*(1), 61–74.

Reich, P. B., & Borchert, R. (1988). Changes with leaf age in stomatal function and water status of several tropical tree species. *Biotropica*, *20*(1), 60–69.

Reich, P. B., Uhl, C., Walters, M. B., & Ellsworth, D. S. (1991). Leaf lifespan as a determinant of leaf structure and function among 23 Amazonian tree species. *Oecologia*, *86*(1), 16–24.

Reich, P. B., Walters, M. B., & Ellsworth, D. S. (1997). From tropics to tundra: Global convergence in plant functioning. *Proc. Natl. Acad. Sci. U. S. A.*, *94*(25), 13730–13734.

Reiche, J., Mullissa, A., Slagter, B., Gou, Y., Tsendbazar, N.-E., Odongo-Braun, C., Vollrath, A., Weisse, M. J., Stolle, F., Pickens, A., Donchyts, G., Clinton, N., Gorelick, N., & Herold, M. (2021). Forest disturbance alerts for the congo basin using sentinel-1. *Environ. Res. Lett.*, *16*(2), 024005.

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M. D., Seneviratne, S. I., Zscheischler, J., Beer, C., Buchmann, N., Frank, D. C., Papale, D., Rammig, A., Smith, P., Thonicke, K., van der Velde, M., Vicca, S., Walz, A., & Wattenbach, M. (2013). Climate extremes and the carbon cycle. *Nature*, *500*(7462), 287–295.

Ren, S., He, K., Girshick, R. B., & Sun, J. (2015). Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, *abs/1506.01497*arXiv 1506.01497. http://arxiv.org/abs/1506.01497

Restrepo-Coupe, N., Albert, L. P., Longo, M., Baker, I., Levine, N. M., Mercado, L. M., da Araujo, A. C., Christoffersen, B. O., Costa, M. H., Fitzjarrald, D. R., Galbraith, D., Imbuzeiro, H., Malhi, Y., von Randow, C., Zeng, X., Moorcroft, P., & Saleska, S. R. (2021). Understanding water and energy fluxes in the Amazonia: Lessons from an observation-model intercomparison. *Glob. Chang. Biol.*, *27*(9), 1802–1819.

Restrepo-Coupe, N., Levine, N. M., Christoffersen, B. O., Albert, L. P., Wu, J., Costa, M. H., Galbraith, D., Imbuzeiro, H., Martins, G., da Araujo, A. C., Malhi, Y. S., Zeng, X., Moorcroft, P., & Saleska, S. R. (2017). Do dynamic global vegetation models capture the seasonality of carbon fluxes in the Amazon basin? A data-model intercomparison. *Glob. Chang. Biol.*, *23*(1), 191–208.

Revell, L. J. (2009). Size-correction and principal components for interspecific comparative studies. *Evolution*, *63*(12), 3258–3268.

Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223.

Ribeiro, D. R., Silva, J. L. A., do Nascimento, M. T., & Vitória, A. P. (2022). Leaf habits and their relationship with leaf and wood traits in tropical dry forests. *Trees*, *36*(1), 7–24.

Richardson, S. J., Smale, M. C., Hurst, J. M., Fitzgerald, N. B., Peltzer, D. A., Allen, R. B., Bellingham, P. J., & McKelvey, P. J. (2009). Large-tree growth and mortality rates

in forests of the central north island, new zealand. *New Zealand Journal of Ecology*, 208–215.

Ritchie, H., & Roser, M. (2024). Drivers of deforestation. *Our World in Data*.

Rivera, G., Elliott, S., Caldas, L. S., Nicolossi, G., Coradin, V. T., & Borchert, R. (2002). Increasing day-length induces spring flushing of tropical dry forest trees in the absence of rain. *Trees*, *16*(7), 445–456.

Rohde, K. (1992). Latitudinal gradients in species diversity: The search for the primary cause. *Oikos*, *65*(3), 514–527.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation arXiv 1505.04597.

Rosa, I. M. D., Ahmed, S. E., & Ewers, R. M. (2014). The transparency, reliability and utility of tropical rainforest land-use and land-cover change models. *Glob. Chang. Biol.*, *20*(6), 1707–1722.

Roussel, J.-R., Caspersen, J., Béland, M., Thomas, S., & Achim, A. (2017). Removing bias from lidar-based estimates of canopy height: Accounting for the effects of pulse density and footprint size. *Remote Sensing of Environment*, *198*, 1–16.

Rusinkiewicz, S., & Levoy, M. (2001). Efficient variants of the ICP algorithm, In *Proceedings third international conference on 3-D digital imaging and modeling*.

Rußwurm, M., & Körner, M. (2017). Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images, In *2017 ieee conference on computer vision and pattern recognition workshops (cvprw)*. https://doi.org/10.1109/CVPRW.2017.193

Rußwurm, M., & Körner, M. (2018a). Multi-Temporal land cover classification with sequential recurrent encoders. *ISPRS International Journal of Geo-Information*, *7*(4), 129.

Rußwurm, M., & Körner, M. (2018b). Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS International Journal of Geo-Information*, *7*(4). https://doi.org/10.3390/ijgi7040129

Ryan, C. M., Williams, M., Grace, J., Woollen, E., & Lehmann, C. E. R. (2017). Pre-rain green-up is ubiquitous across southern tropical Africa: Implications for temporal niche separation and model representation. *New Phytol.*, *213*(2), 625–633.

Ryan, M. G., & Yoder, B. J. (1997). Hydraulic limits to tree height and tree growth. *Bioscience*, *47*(4), 235–242.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M., & Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc. Natl. Acad. Sci. U. S. A.*, *108*(24), 9899–9904.

Sabatier, D. (1983). *Fructification et dissémination en forêt guyanaise : L'exemple de quelques espèces ligneuses* (Doctoral dissertation). Université des Sciences et Techniques du Languedoc. Montpellier.

Sabat-Tomala, A., Raczko, E., & Zagajewski, B. (2020). Comparison of support vector machine and random forest algorithms for invasive and expansive species classification using airborne hyperspectral data. *Remote Sensing*, *12*(3), 516.

Saha, S., Saha, M., Mukherjee, K., Arabameri, A., Ngo, P. T. T., & Paul, G. C. (2020). Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, REPTree: A case study at the gumani river basin, india. *Sci. Total Environ.*, *730*, 139197.

Sakschewski, B., von Bloh, W., Drüke, M., Sörensson, A. A., Ruscica, R., Langerwisch, F., Billing, M., Bereswill, S., Hirota, M., Oliveira, R. S., Heinke, J., & Thonicke, K. (2021). Variable tree rooting strategies are key for modelling the distribution, productivity and evapotranspiration of tropical evergreen forests. *Biogeosciences*, *18*(13), 4091–4116.

Saleska, S. R., Miller, S. D., Matross, D. M., Goulden, M. L., Wofsy, S. C., da Rocha, H. R., de Camargo, P. B., Crill, P., Daube, B. C., de Freitas, H. C., Hutyra, L., Keller, M., Kirchhoff, V., Menton, M., Munger, J. W., Pyle, E. H., Rice, A. H., & Silva, H. (2003). Carbon in Amazon forests: Unexpected seasonal fluxes and disturbance-induced losses. *Science*, *302*(5650), 1554–1557.

Saleska, S. R., Wu, J., Guan, K., Araujo, A. C., Huete, A., Nobre, A. D., & Restrepo-Coupe, N. (2016). Dry-season greening of Amazon forests. *Nature*, *531*(7594), E4–5.

Samanta, A., Ganguly, S., Hashimoto, H., Devadiga, S., Vermote, E., Knyazikhin, Y., Nemani, R. R., & Myneni, R. B. (2010). Amazon forests did not green-up during the 2005 drought. *Geophys. Res. Lett.*, *37*(5).

Santopaolo, A., Saif, S. S., Pietrabissa, A., & Giuseppi, A. (2021). Forest fire risk prediction from satellite data with convolutional neural networks, In *2021 29th mediterranean conference on control and automation (MED)*, IEEE.

Santos, A. A. d., Marcato Junior, J., Araújo, M. S., Di Martini, D. R., Tetila, E. C., Siqueira, H. L., Aoki, C., Eltner, A., Matsubara, E. T., Pistori, H. Et al. (2019). Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to uavs. *Sensors*, *19*(16), 3595.

Sastry, A., & Barua, D. (2017). Leaf thermotolerance in tropical trees from a seasonally dry climate varies along the slow-fast resource acquisition spectrum. *Sci. Rep.*, *7*(1), 11246.

Scheiter, S., Langan, L., & Higgins, S. I. (2013). Next-generation dynamic global vegetation models: Learning from community ecology. *New Phytol.*, *198*(3), 957–969.

Schläpfer, D., Hueni, A., & Richter, R. (2018). Cast shadow detection to quantify the aerosol optical thickness for atmospheric correction of high spatial resolution optical imagery. *Remote Sensing*, *10*(2).

Schläpfer, D., & Richter, R. (2014). Evaluation of brefcor BRDF effects correction for HYSPEX, CASI, and APEX imaging spectroscopy data, In *2014 6th workshop on hyperspectral image and signal processing: Evolution in remote sensing (WHISPERS)*.

Schläpfer, D., Richter, R., & Feingersh, T. (2015). Operational BRDF effects correction for Wide-Field-of-View optical scanners (BREFCOR). *IEEE Trans. Geosci. Remote Sens.*, *53*(4), 1855–1864.

Schmitt, S., Trueba, S., Coste, S., Ducouret, É., Tysklind, N., Heuertz, M., Bonal, D., Burban, B., Hérault, B., & Derroire, G. (2022). Seasonal variation of leaf thickness: An overlooked component of functional trait variability. *Plant Biol.*, *24*(3), 458–463.

Schweiger, A. K., Cavender-Bares, J., Kothari, S., Townsend, P. A., Madritch, M. D., Grossman, J. J., Gholizadeh, H., Wang, R., & Gamon, J. A. (2021). Coupling spectral and resource-use complementarity in experimental grassland and forest communities. *Proc. Biol. Sci.*, *288*(1958), 20211290.

Serbin, S. P., Singh, A., McNeil, B. E., Kingdon, C. C., & Townsend, P. A. (2016). Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species. *Ecol. Appl.*, *24*(7), 1651–1669.

Serrano, L., Peñuelas, J., & Ustin, S. L. (2002). Remote sensing of nitrogen and lignin in mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. *Remote Sens. Environ.*, *81*(2), 355–364.

Settles, B. (2009). Active learning literature survey (computer sciences technical report 1648) university of Wisconsin-Madison. *Madison, WI, USA: Jan*.

Seymour, F., & Harris, N. L. (2019). Reducing tropical deforestation. *Science*, *365*(6455), 756–757.

Seyoum, Y., Fetene, M., Strobl, S., & Beck, E. (2012). Foliage dynamics, leaf traits, and growth of coexisting evergreen and deciduous trees in a tropical montane forest in ethiopia. *Trees*, *26*(5), 1495–1512.

Shamaoma, H., Chirwa, P. W., Zekeng, J. C., Ramoelo, A., Hudak, A. T., Handavu, F., & Syampungani, S. (2023). Use of Multi-Date and Multi-Spectral UAS imagery to classify dominant tree species in the wet miombo woodlands of zambia. *Sensors*, *23*(4).

Shang, X., & Chisholm, L. A. (2014). Classification of Australian native forest species using hyperspectral remote sensing and Machine-Learning classification algorithms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(6), 2481–2489.

Shenkin, A., Chandler, C. J., Boyd, D. S., Jackson, T., Disney, M., Majalap, N., Nilus, R., Foody, G., bin Jami, J., Reynolds, G. Et al. (2019). The world's tallest tropical tree in three dimensions. *Frontiers in Forests and Global Change*, *2*, 32.

SHI, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., & WOO, W.-c. (2015). Convolutional lstm network: A machine learning approach for precipitation nowcasting. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems 28* (pp. 802–810). Curran Associates, Inc.

Silva, F. B., Shimabukuro, Y. E., Aragão, L. E. O., Anderson, L. O., Pereira, G., Cardozo, F., & Arai, E. (2013). Large-scale heterogeneity of Amazonian phenology revealed from 26-year long AVHRR/NDVI time-series. *Environ. Res. Lett.*, *8*(2), 024011.

Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.*, *81*(2), 337–354.

Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., Doney, S. C., Graven, H., Heinze, C., Huntingford, C., Levis, S., Levy, P. E., Lomas, M., Poulter, B., Viovy, N., Zaehle, S., Zeng, N., Arneth, A., Bonan, G., … Myneni, R. (2015). Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences*, *12*(3), 653–679.

Slik, J. W. F., Arroyo-Rodriguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L. F., Ashton, P., Balvanera, P., Bastian, M. L., Bellingham, P. J., van den Berg, E., Bernacci, L., da Conceição Bispo, P., Blanc, L., Böhning-Gaese, K., Boeckx, P., Bongers, F., Boyle, B., Bradford, M., Brearley, F. Q., … Venticinque, E. M. (2015). An estimate of the number of tropical tree species. *Proc. Natl. Acad. Sci. U. S. A.*, *112*(24), 7472–7477.

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms (F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger, Eds.). In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, Curran Associates, Inc.

Sobrado, M. A. (1991). Cost-Benefit relationships in deciduous and evergreen leaves of tropical dry forest species. *Funct. Ecol.*, *5*(5), 608–616.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, *15*(56), 1929–1958.

Ssali, F., & Sheil, D. (2023). Seasonality in the equatorial tropics: Flower, fruit, and leaf phenology of montane trees in the highlands of southwest uganda. *Biotropica*, *55*(3), 680–698.

Staggemeier, V. G., Camargo, M. G. G., Diniz-Filho, J. A. F., Freckleton, R., Jardim, L., & Morellato, L. P. C. (2020). The circular nature of recurrent life cycle events: A test comparing tropical and temperate phenology. *J. Ecol.*, *108*(2), 393–404.

Stephenson, P. J., & Stengel, C. (2020). An inventory of biodiversity data sources for conservation monitoring. *PLoS One*, *15*(12), e0242923.

Stevens, N., Archibald, S. A., Nickless, A., Swemmer, A., & Scholes, R. J. (2016). Evidence for facultative deciduousness in *Colophospermum mopane* in semi-arid African savannas. *Austral Ecol.*, *41*(1), 87–96.

Stovall, A. E. L., Shugart, H., & Yang, X. (2019). Tree height explains mortality risk during an intense drought. *Nature Communications*, *10*, 4385. https://doi.org/10.1038/s41467-019-12380-6

Sun, Z., Wang, X., Wang, Z., Yang, L., Xie, Y., & Huang, Y. (2021). UAVs as remote sensing platforms in plant ecology: Review of applications and challenges. *J Plant Ecol*, *14*(6), 1003–1023.

Taffo, J. B. W., Nguetsop, V. F., Anja, G. M., Solefack, M. C. M., Tacha, W. N., & Feukeng, S. S. K. (2019). Phenological behaviour of tropical tree species in three altitudinal zones of Bambouto mountains, west Cameroon. *J. Appl. Sci.*, *19*(2), 68–76.

Tengö, M., Hill, R., Malmer, P., Raymond, C. M., Spierenburg, M., Danielsen, F., Elmqvist, T., & Folke, C. (2017). Weaving knowledge systems in IPBES, CBD and beyond—lessons learned for sustainability. *Current Opinion in Environmental Sustainability*, *26-27*, 17–25.

Ter Steege, H., Pitman, N., Sabatier, D., Castellanos, H., Van Der Hout, P., Daly, D. C., Silveira, M., Phillips, O., Vasquez, R., Van Andel, T., Duivenvoorden, J., De Oliveira, A. A., Ek, R., Lilwah, R., Thomas, R., Van Essen, J., Baider, C., Maas, P., Mori, S., … Morawetz, W. (2003). A spatial model of tree α-diversity and tree density for the Amazon. *Biodiversity & Conservation*, *12*(11), 2255–2277.

Ter Steege, H., Prado, P. I., Lima, R. A. F. d., Pos, E., de Souza Coelho, L., de Andrade Lima Filho, D., Salomão, R. P., Amaral, I. L., de Almeida Matos, F. D., Castilho, C. V., Phillips, O. L., Guevara, J. E., de Jesus Veiga Carim, M., Cárdenas López, D., Magnusson, W. E., Wittmann, F., Martins, M. P., Sabatier, D., Irume, M. V., … Pickavance, G. (2020). Biased-corrected richness estimates for the Amazonian tree flora. *Sci. Rep.*, *10*(1), 10130.

Ter Steege, H., Vaessen, R. W., Cárdenas-López, D., Sabatier, D., Antonelli, A., de Oliveira, S. M., Pitman, N. C. A., Jørgensen, P. M., & Salomão, R. P. (2016). The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. *Sci. Rep.*, *6*, 29549.

ter Steege, H., Pitman, N. C. A., Sabatier, D., Baraloto, C., Salomão, R. P., Guevara, J. E., Phillips, O. L., Castilho, C. V., Magnusson, W. E., Molino, J.-F., Monteagudo, A., Núñez Vargas, P., Montero, J. C., Feldpausch, T. R., Coronado, E. N. H., Killeen, T. J., Mostacedo, B., Vasquez, R., Assis, R. L., … Silman, M. R. (2013). Hyperdominance in the Amazonian tree flora. *Science*, *342*(6156), 1243092.

Teuling, A. J., Seneviratne, S. I., Stöckli, R., Reichstein, M., Moors, E., Ciais, P., Luyssaert, S., van den Hurk, B., Ammann, C., Bernhofer, C., Dellwik, E., Gianelle, D., Gielen, B., Grünwald, T., Klumpp, K., Montagnani, L., Moureaux, C., Sottocornola, M., & Wohlfahrt, G. (2010). Contrasting response of European forest and grassland energy exchange to heatwaves. *Nat. Geosci.*, *3*(10), 722–727.

Theiler, J., Ziemann, A., Matteoli, S., & Diani, M. (2019). Spectral variability of remotely sensed target materials: Causes, models, and strategies for mitigation and robust exploitation. *IEEE Geoscience and Remote Sensing Magazine*, *7*(2), 8–30.

Thomas, J. R., & Gausman, H. W. (1977). Leaf reflectance vs. leaf chlorophyll and carotenoid concentrations for eight crops [1]. *Agron. J.*, *69*(5), 799–802.

Thompson, I. D., Okabe, K., Parrotta, J. A., Brockerhoff, E., Jactel, H., Forrester, D. I., & Taki, H. (2014). Biodiversity and ecosystem services: Lessons from nature to improve management of planted forests for REDD-plus. *Biodivers. Conserv.*, *23*(10), 2613–2635.

Tong, K., & Wu, Y. (2022). Deep learning-based detection from the perspective of small or tiny objects: A survey. *Image and Vision Computing*, *123*, 104471. https://doi.org/10.1016/j.imavis.2022.104471

Torres, D. L., Turnes, J. N., Soto Vega, P. J., Feitosa, R. Q., Silva, D. E., Marcato Junior, J., & Almeida, C. (2021). Deforestation detection with fully convolutional networks in the Amazon Forest from Landsat-8 and Sentinel-2 images. *Remote Sensing*, *13*(24), 5084.

Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). Learning spatiotemporal features with 3d convolutional networks, In *In iccv*.

Troles, J.-D., Nieding, R., Simons, S., & Schmid, U. (2023). Task planning support for arborists and foresters: Comparing deep learning approaches for tree inventory and tree vitality assessment based on UAV-Data arXiv 2307.01651.

Turubanova, S., Potapov, P. V., Tyukavina, A., & Hansen, M. C. (2018). Ongoing primary forest loss in brazil, democratic republic of the congo, and indonesia. *Environ. Res. Lett.*, *13*(7), 074028.

Tydecks, L., Jeschke, J. M., Wolf, M., Singer, G., & Tockner, K. (2018). Spatial and topical imbalances in biodiversity research. *PLoS One*, *13*(7), e0199327.

Tyukavina, A., Baccini, A., Hansen, M. C., Potapov, P. V., Stehman, S. V., Houghton, R. A., Krylov, A. M., Turubanova, S., & Goetz, S. J. (2015). Aboveground carbon loss in natural and managed tropical forests from 2000 to 2012. *Environ. Res. Lett.*, *10*(7), 074002.

Uckele, K. A., Jahner, J. P., Tepe, E. J., Richards, L. A., Dyer, L. A., Ochsenrider, K. M., Philbin, C. S., Kato, M. J., Yamaguchi, L. F., Forister, M. L., Smilanich, A. M., Dodson, C. D., Jeffrey, C. S., & Parchman, T. L. (2021). Phytochemistry reflects different evolutionary history in traditional classes versus specialized structural motifs. *Sci. Rep.*, *11*(1), 17247.

Ustin, S. L. (2013). Remote sensing of canopy chemistry. *Proc. Natl. Acad. Sci. U. S. A.*, *110*(3), 804–805.

Ustin, S. L., Gitelson, A. A., Jacquemoud, S., Schaepman, M., Asner, G. P., Gamon, J. A., & Zarco-Tejada, P. (2009). Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sens. Environ.*, *113*, S67–S77.

Ustin, S. L., Roberts, D. A., Gamon, J. A., Asner, G. P., & Green, R. O. (2004). Using imaging spectroscopy to study ecosystem processes and properties. *Bioscience*, *54*(6), 523–534.

Vaglio Laurin, G., Cheung-Wai Chan, J., Chen, Q., Lindsell, J. A., Coomes, D. A., Guerriero, L., Del Frate, F., Miglietta, F., & Valentini, R. (2014). Biodiversity mapping in a tropical West African forest with airborne hyperspectral data. *PLoS One*, *9*(6), e97910.

Valencia, R., Balslev, H., & Paz Y Miño C, G. (1994). High tree alpha-diversity in Amazonian Ecuador. *Biodiversity & Conservation*, *3*(1), 21–28.

van der Plas, F. (2019). Biodiversity and ecosystem functioning in naturally assembled communities. *Biol. Rev. Camb. Philos. Soc.*, *94*(4), 1220–1245.

van Aardt, J. A. N., & Wynne, R. H. (2007). Examining pine spectral separability using hyperspectral data from an airborne sensor: An extension of field-based results. *Int. J. Remote Sens.*, *28*(2), 431–436.

van Ewijk, K. Y., Randin, C. F., Treitz, P. M., & Scott, N. A. (2014). Predicting fine-scale tree species abundance patterns using biotic variables derived from LiDAR and high spatial resolution imagery. *Remote Sens. Environ.*, *150*, 120–131.

van Schaik, C. P., Terborgh, J. W., & Wright, S. J. (1993). The phenology of tropical forests: Adaptive significance and consequences for primary consumers. *Annu. Rev. Ecol. Syst.*, *24*(1), 353–377.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. U., & Polosukhin, I. (2017). Attention is all you need (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett, Eds.). In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems*, Curran Associates, Inc.

Vauhkonen, J., Ørka, H. O., Holmgren, J., Dalponte, M., Heinzel, J., & Koch, B. (2014). Tree species recognition based on airborne laser scanning and complementary data sources. In M. Maltamo, E. Næsset, & J. Vauhkonen (Eds.), *Forestry applications of airborne laser scanning: Concepts and case studies* (pp. 135–156). Dordrecht, Springer Netherlands.

Vicari, M. B., Disney, M., Wilkes, P., Burt, A., Calders, K., & Woodgate, W. (2019). Leaf and wood classification framework for terrestrial LiDAR point clouds. *Methods Ecol. Evol.*, *10*(5), 680–694.

Vincent, G., Antin, C., Laurans, M., Heurtebize, J., Durrieu, S., Lavalley, C., & Dauzat, J. (2017). Mapping plant area index of tropical evergreen forest by airborne laser scanning. a cross-validation study using LAI2200 optical sensor. *Remote Sens. Environ.*, *198*, 254–266.

Vincent, G., Pimont, F., & Verley, P. (2021). Various PAD/LAD estimators implemented in AMAPVox 1.8. DataSuds.

Vincent, G., Verley, P., Brede, B., Delaitre, G., Maurent, E., Ball, J. G. C., Clocher, I., & Barbier, N. (2023). Multi-sensor airborne lidar requires intercalibration for consistent estimation of light attenuation and plant area density. *Remote Sensing of Environment*, *286*, 113442.

Visser, M. E., & Both, C. (2005). Shifts in phenology due to global climate change: The need for a yardstick. *Proc. Biol. Sci.*, *272*(1581), 2561–2569.

Vorovencii, I., Dinca, L., Crisan, V., Postolache, R.-G., Codrean, C.-L., Catalin, C., Gresița, C. I., Chima, S., & Gavrilescu, I. (2023). Local-scale mapping of tree species in a lower mountain area using sentinel-1 and -2 multitemporal images, vegetation indices, and topographic information. *Frontiers in Forests and Global Change*, *6*.

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.*, *2018*, 7068349.

Wagner, F., Hérault, B., Stahl, C., Bonal, D., & Rossi, V. (2011). Modeling water availability for trees in tropical forests. *Agric. For. Meteorol.*, *151*(9), 1202–1213.

Wagner, F., Rutishauser, E., Blanc, L., & Herault, B. (2010). Effects of plot size and census interval on descriptors of forest structure and dynamics. *Biotropica*, *42*(6), 664–671. https://doi.org/10.1111/j.1744-7429.2010.00644.x

Wagner, F. H., Hérault, B., Rossi, V., Hilker, T., Maeda, E. E., Sanchez, A., Lyapustin, A. I., Galvão, L. S., Wang, Y., & Aragão, L. E. O. C. (2017). Climate drivers of the Amazon forest greening. *PLoS One*, *12*(7), e0180932.

Wang, H., Prentice, I. C., Wright, I. J., Warton, D. I., Qiao, S., Xu, X., Zhou, J., Kikuzawa, K., & Stenseth, N. C. (2023). Leaf economics fundamentals explained by optimality principles. *Sci Adv*, *9*(3), eadd5667.

Wang, J., Yang, D., Detto, M., Nelson, B. W., Chen, M., Guan, K., Wu, S., Yan, Z., & Wu, J. (2020). Multi-scale integration of satellite remote sensing improves characterization of dry-season green-up in an Amazon tropical evergreen forest. *Remote Sens. Environ.*, *246*, 111865.

Wardle, D. A. (2001). Experimental demonstration that plant diversity reduces invasibility: Evidence of a biological mechanism or a consequence of sampling effect? *Oikos*, *95*(1), 161–170.

Watt, M. S., Leonardo, E. M. C., Estarija, H. J. C., Massam, P., de Silva, D., O'Neill, R., Lane, D., McDougal, R., Buddenbaum, H., & Zarco-Tejada, P. J. (2021). Long-term effects of water stress on hyperspectral remote sensing indicators in young radiata pine. *For. Ecol. Manage.*, *502*, 119707.

Weinstein, B., Marconi, S., Zare, A., Bohlman, S., Graves, S., Singh, A., & White, E. (2020). NEON tree crowns dataset. Zenodo.

Weinstein, B. G., Marconi, S., Bohlman, S., Zare, A., & White, E. (2019). Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sensing*, *11*(11), 1309.

Weinstein, B. G., Marconi, S., Bohlman, S. A., Zare, A., Singh, A., Graves, S. J., & White, E. P. (2021). A remote sensing derived data set of 100 million individual tree crowns for the national ecological observatory network. *Elife*, *10*, e62922.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, *3*(1), 1–40.

Weisse, M., Goldman, L., & Carter, S. (2023). Tropical primary forest loss worsened in 2022, despite international commitments to end deforestation [Accessed: 2023-11-12].

Weng, J.-K., & Chapple, C. (2010). The origin and evolution of lignin biosynthesis. *New Phytol.*, *187*(2), 273–285.

Wessel, M., Brandmeier, M., & Tiede, D. (2018). Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on sentinel-2 data. *Remote Sensing*, *10*(9), 1419.

Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., & Reynolds, J. M. (2012). structure-from-motion'photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology*, *179*, 300–314.

Wikipedia contributors. (2024). Precision and recall [Accessed: NA-NA-NA].

Williams, L. J., Bunyavejchewin, S., & Baker, P. J. (2008). Deciduousness in a seasonal tropical forest in Western Thailand: Interannual and intraspecific variation in timing, duration and environmental cues. *Oecologia*, *155*(3), 571–582.

Williams, M. (2003). *Deforesting the Earth: From prehistory to global crisis*. University of Chicago Press.

Wilson, K. A., Underwood, E. C., Morrison, S. A., Klausmeyer, K. R., Murdoch, W. W., Reyers, B., Wardell-Johnson, G., Marquet, P. A., Rundel, P. W., McBride, M. F., Pressey, R. L., Bode, M., Hoekstra, J. M., Andelman, S., Looker, M., Rondinini, C., Kareiva, P., Shaw, M. R., & Possingham, H. P. (2007). Conserving biodiversity efficiently: What to do, where, and when. *PLoS Biol.*, *5*(9), e223.

Woolston, C., & Ball, J. G. C. (2021). Unpicking the rhythms of the Amazon rainforest: (Where I work). *Nature*, *591*, 494–494.

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H. C., Diemer, M., Flexas, J., Garnier, E., Groom, P. K., Gulias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., … Villar, R. (2004). The worldwide leaf economics spectrum. *Nature*, *428*(6985), 821–827.

Wright, J. S. (2002). Plant diversity in tropical forests: A review of mechanisms of species coexistence. *Oecologia*, *130*(1), 1–14.

Wright, S. J., & Cornejo, F. H. (1990). Seasonal drought and leaf fall in a tropical forest. *Ecology*, *71*(3), 1165–1175.

Wright, S. J., & Muller-Landau, H. C. (2006). The future of tropical forest species. *Biotropica*, *38*(3), 287–301.

Wright, S. J., & van Schaik, C. P. (1994). Light and the phenology of tropical trees. *Am. Nat.*, *143*(1), 192–199.

Wu, J., Su, Y., Chen, X., Liu, L., Yang, X., Gong, F., Zhang, H., Xiong, X., & Zhang, D. (2021). Leaf shedding of Pan-Asian tropical evergreen forests depends on the synchrony of seasonal variations of rainfall and incoming solar radiation. *Agric. For. Meteorol.*, *311*, 108691.

Wu, J., Albert, L. P., Lopes, A. P., Restrepo-Coupe, N., Hayek, M., Wiedemann, K. T., Guan, K., Stark, S. C., Christoffersen, B., Prohaska, N., Tavares, J. V., Marostica, S., Kobayashi, H., Ferreira, M. L., Campos, K. S., da Silva, R., Brando, P. M., Dye, D. G., Huxman, T. E., … Saleska, S. R. (2016). Leaf development and demography explain photosynthetic seasonality in Amazon evergreen forests. *Science*, *351*(6276), 972–976.

Wu, J., Guan, K., Hayek, M., Restrepo-Coupe, N., Wiedemann, K. T., Xu, X., Wehr, R., Christoffersen, B. O., Miao, G., da Silva, R., de Araujo, A. C., Oliviera, R. C., Camargo, P. B., Monson, R. K., Huete, A. R., & Saleska, S. R. (2017a). Partitioning controls on

Amazon forest photosynthesis between environmental and biotic factors at hourly to interannual timescales. *Glob. Chang. Biol.*, *23*(3), 1240–1257.

Wu, J., Rogers, A., Albert, L. P., Ely, K., Prohaska, N., Wolfe, B. T., Oliveira, R. C., Jr, Saleska, S. R., & Serbin, S. P. (2019a). Leaf reflectance spectroscopy captures variation in carboxylation capacity across species, canopy environment and leaf age in lowland moist tropical forests. *New Phytol.*, *224*(2), 663–674.

Wu, J., Serbin, S. P., Xu, X., Albert, L. P., Chen, M., Meng, R., Saleska, S. R., & Rogers, A. (2017b). The phenology of leaf quality and its within-canopy variation is essential for accurate modelling of photosynthesis in tropical evergreen forests. *Glob. Chang. Biol.*, *23*(11), 4814–4827.

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., & Girshick, R. (2019b). Detectron2.

WWF. (2020). Predicting deforestation with an early warning system [Accessed: 2021-12-9].

Xiao, X., Hagen, S., Zhang, Q., Keller, M., & Moore, B. (2006). Detecting leaf phenology of seasonally moist tropical forests in South America with multi-temporal MODIS images. *Remote Sens. Environ.*, *103*(4), 465–473.

Xu, Z., Guan, K., Casler, N., Peng, B., & Wang, S. (2018). A 3D convolutional neural network method for land cover classification using LiDAR and multi-temporal landsat imagery. *ISPRS J. Photogramm. Remote Sens.*, *144*, 423–434.

Yang, X., Wu, J., Chen, X., Ciais, P., Maignan, F., Yuan, W., Piao, S., Yang, S., Gong, F., Su, Y., Dai, Y., Liu, L., Zhang, H., Bonal, D., Liu, H., Chen, G., Lu, H., Wu, S., Fan, L., … Wright, S. J. (2021). A comprehensive framework for seasonal controls of leaf abscission and productivity in evergreen broadleaved tropical and subtropical forests. *Innovation (Camb)*, *2*(4), 100154.

Yang, Y., Zhu, Q., Peng, C., Wang, H., Xue, W., Lin, G., Wen, Z., Chang, J., Wang, M., Liu, G., & Li, S. (2016). A novel approach for modelling vegetation distributions and analysing vegetation sensitivity through trait-climate relationships in china. *Sci. Rep.*, *6*, 24110.

Yokoyama, S., Xing, J., Liu, Y., Faggionato, D., Altun, A., & Starmer, W. T. (2014). Epistatic adaptive evolution of human color vision. *PLoS Genet.*, *10*(12), e1004884.

Yu, K., Ciais, P., Seneviratne, S. I., Liu, Z., Chen, H. Y. H., Barichivich, J., Allen, C. D., Yang, H., Huang, Y., & Ballantyne, A. P. (2022a). Field-based tree mortality constraint reduces estimates of model-projected forest carbon sinks. *Nat. Commun.*, *13*(1), 2094.

Yu, K., Hao, Z., Post, C. J., Mikhailova, E. A., Lin, L., Zhao, G., Tian, S., & Liu, J. (2022b). Comparison of classical methods and Mask R-CNN for automatic tree detection and mapping using uav imagery. *Remote Sensing*, *14*(2). https://doi.org/10.3390/rs14020295

Zarin, D. J., Harris, N. L., Baccini, A., Aksenov, D., Hansen, M. C., Azevedo-Ramos, C., Azevedo, T., Margono, B. A., Alencar, A. C., Gabris, C., Allegretti, A., Potapov, P.,

Farina, M., Walker, W. S., Shevade, V. S., Loboda, T. V., Turubanova, S., & Tyukavina, A. (2016). Can carbon emissions from tropical deforestation drop by 50% in 5 years? *Glob. Chang. Biol.*, *22*(4), 1336–1347.

Zeileis, A., & Grothendieck, G. (2005). Zoo: S3 infrastructure for regular and irregular time series arXiv math/0505527.

Zhang, H., Yuan, W., Dong, W., & Liu, S. (2014). Seasonal patterns of litterfall in forest ecosystem worldwide. *Ecol. Complex.*, *20*, 240–247.

Zhang, J., Rivard, B., Sánchez-Azofeifa, G. A., & Castro-Esau, K. (2006). Intra- and inter-class spectral variability of tropical tree species at La Selva, Costa Rica: Implications for species identification using HYDICE imagery. *Remote Sens. Environ.*, *105*(2), 129–141.

Zhang, Y., Liu, H., & Hu, Q. (2021). TransFuse: Fusing transformers and CNNs for medical image segmentation, In *Medical image computing and computer assisted intervention – MICCAI 2021*, Springer International Publishing.

Zhao, Y.-T., Ali, A., & Yan, E.-R. (2017). The plant economics spectrum is structured by leaf habits and growth forms across subtropical species. *Tree Physiol.*, *37*(2), 173–185.

Zhen, Z., Quackenbush, L. J., & Zhang, L. (2016). Trends in automatic individual tree crown detection and delineation—evolution of lidar data. *Remote Sensing*, *8*(4). https://doi.org/10.3390/rs8040333

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 8–36.

Zou, L., Stan, K., Cao, S., & Zhu, Z. (2023). Dynamic global vegetation models may not capture the dynamics of the leaf area index in the tropical rainforests: A data-model intercomparison. *Agric. For. Meteorol.*, *339*, 109562.

Zuidema, P. A., & van der Sleen, P. (2022). Seeing the forest through the trees: How tree-level measurements can help understand forest dynamics. *New Phytologist*, *234*(5), https://nph.onlinelibrary.wiley.com/doi/pdf/10.1111/nph.18144, 1544–1546. https://doi.org/10.1111/nph.18144

Zuppinger-Dingley, D., Schmid, B., Petermann, J. S., Yadav, V., De Deyn, G. B., & Flynn, D. F. B. (2014). Selection for niche differentiation in plant communities increases biodiversity effects. *Nature*, *515*(7525), 108–111.

Appendix A

Chapter 2 Supplementary Materials

A.1 Delimitation of study areas

A.1.1 Madre de Dios

The boundary of Madre de Dios was taken from a shapefile of Peru's administrative Departments[1].

A.1.2 Junin region

The study region was defined by the following steps: Starting layers were:
1) Shapefile of Peru's administrative Departments
2) WWF map of the entire Amazon Ecoregion

Junin, and the five surrounding departments, Pasco, Apurimac, Ayachcho, Cusco and Huancavelica, were selected and merged (dissolved) into one area. Then, the intersection of the dissolved departments and the Amazon Ecoregion was calculated. Some small islands detached from the main area were removed. To allow for predictions close to the edge of the region, a second region was also defined by adding an additional buffer of 0.09 degrees and the removing any internal islands that had not been included.

[1] https://data.humdata.org/dataset/limites-de-peru

A.2 Datasets and feature extraction

A.2.1 Global Forest Change dataset and Landsat imagery

The Global Forest Change dataset (Hansen et al., 2013) gives areas (at approximately 30×30 meter resolution) that have been detected from Landsat imagery as undergoing forest loss. The first release of the global map of forest cover loss, was made in 2013 and contains a map of forest extent (in 2000), annual loss (from 2000 to 2012), and overall gain (from 2000 to 2012). Since then, the data set has been updated annually with the same methodologies being used as in 2013. The current Version 1.8, includes forest loss from 2000 through to 2020. The features of the data is given in Table A.1.

"Tree cover" is defined as "all vegetation greater than 5 meters in height, and may take the form of natural forests or plantations across a range of canopy densities". Forest loss is defined as "the disturbance or complete removal of tree cover canopy (below 30% tree canopy cover)". The forest loss detection does not differentiate between permanent tree cover loss or temporary loss from which the forest will recover. It also does not determine whether the cause of the loss is natural or human induced. The dataset comes with a layer of percentage tree canopy cover observed in the year 2000, a data mask map (describing areas of land or permanent water) and a cloud-free multi-spectral (Landsat) satellite image from the year 2000. Additionally, each annual release comes with the most recent available cloud-free Landsat image composite from that year (the "last" layer) and updated layer an annual forest loss ("lossyear").

The latest data was extracted for the Madre de Dios area and the Junin area described above. Consistent cloud-free Landsat imagery was available from 2014. Annual forest loss events were available from 2001 to 2020.

Table A.1 Layers of the Global Forest Change dataset.

Variable	Description	Value
treecover2000	Percentage of tree cover in the pixel observed in 2000.	0 - 100
gain	One if gain happens during the period: 2000 - 2012, zero otherwise.	0 or 1
lossyear	The year when loss was detected, one-indexed from year 2001, or zero if no loss occurred.	0 - 18
datamask	No data (0), mapped land surface (1), and permanent water bodies (2).	0,1 or 2
first_b30	The Landsat 7 red band built from the first cloud free pixels in 2000.	0 - 255
first_b40	The Landsat 7 near infrared band built from the first valid pixels in 2000.	0 - 255
first_b50	The first Landsat 7 short wave infrared band built from the first valid pixels in 2000.	0 - 255
first_b70	The second Landsat 7 short wave infrared band built from the first valid pixels in 2000.	0 - 255
last_b30_2013 (/2014/..)	The Landsat 7 red band built from the latest valid pixels in 2013 (/2014...).	0 - 255
last_b40_2013 (/2014...)	The Landsat 7 near infrared band built from the latest valid pixels in 2013(/2014...).	0 - 255
last_b50_2013 (/2014/..)	The first Landsat 7 short wave infrared band built from the latest valid pixels 2013 (/2014...).	0 - 255
last_b70_2013 (/2014/..)	The second Landsat 7 short wave infrared band built from the latest valid pixels 2013 (/2014...).	0 - 255

Since the utility of optical satellite imagery is highly reliant on cloud cover (which can be persistent across the Andean Amazon), some late-in the year losses may only detected in the following year. The assignment of deforestation events to a particular year should therefore be treated with some caution.

Additionally, two different algorithms were used to generate the the measurements of tree cover loss - one for 2001-2010 and another for 2011-2018. The new algorithm is more sensitive to small-scale agricultural, fire-caused, or other forest losses.

Global Forest Watch noted that they observe large spike in the tree cover loss in the years 2016 and 2017 globally, and that the causes for that are determined to be fires. While this can be considered as an anomaly in their dataset, they also suggest that other pre-2011 fire-related losses may not detected by their initial algorithm. The scientists involved in this project are working to back-cast the new algorithm to generate one consistent time series of forest loss events.

A.2.2 Digital surface model

The second set of models for the Junin/Ashaninka area were further developed by the inclusion of an elevation predictor layer to allow the models to learn from features of the topology.

A Digital Elevation Model (DEM) or Digital Surface Model provides a digital representation of the Earth's surface. In the We included the Japan Aerospace Exploration Agency's (JAXA) 30-m resolution ALOS Global DSM as a layer for the models. Including a DSM as a layer gives another dimension to the data and allows the deep networks to learn and predict from features of the topology of the areas of interest.

We compared the available DEM/DSM sources. Alternatives included NASA's SRTM DSM, the MERIT DEM and the TanDEM-X DEM. The later two were excluded as their resolution was too low. SRTM is said to struggle in sloping regions with foreshortening, layover and shadow.

ALOS PALSAR is an L-band product and so has better canopy penetration when compared to C-band products. This means it is more reflective of topology rather than canopy surface.

JAXA's ALOS Global DSM was judged to be the most precise and suitable to complement the Global Forest Change and satellite data described above.

A.2.3 Feature extraction

Global Forest Change dataset (Hansen et al., 2013) is divided into 10x10 degree tiles, each of which comes with six raster files per tile: treecover, gain, data mask, loss year, first and last (see Table A.1). All files contain unsigned 8-bit values and have a spatial resolution of

1 arc-second per pixel, which correspond to approximately 30 meters per pixel around the equator. After 2013 loss year and last files were updated annually. The last 2020 loss year file assign an integer value 0-20 to each pixel. 1-20 corresponds to the year (2001-2020) at which a forest to non-forest event was observed at this location. 0 is assigned if no change is detected in the period 2001-2020. The dataset is encoded such that once a pixel is assigned as deforested, it does not go back to forested at any time in the future. We collected the following ten tif files: treecover, gain, datamask, all "last" files from 2014 to 2020 and the most recent, 2020, loss year file. Since we wish our models to be able to predict the label of each pixel of the regions by analyzing an image, or time series of images, that captures its local region, we also included pixels lying in a buffer area of 0.09 degree (or approximately 10km) in our dataset. This allowed us to extract features from images that cover area up to 10km away from within region pixel.

From the processed dataset we then assigned 7-8 predictor values to each pixel, 2 stationary and 5 that vary each of the years 2014-2020. Table A.2 provides explanation for each of them. Each pixel also has a corresponding lossyear value $\in 0, 1, 2, 3, 4, ...20$, where 1-20 indicate the year at which it was marked as deforested or 0 if it did not experience deforestation up to year 2020.

Table A.2 Predictor layers for the forecasts.

Variable	Description	Value
datamask	Mapped land surface (0), and permanent water bodies (1).	0, 1
treecover2000	Percentage of tree cover in the pixel observed in 2000.	0 - 100
elevation*	Height above sea level in meters	0 - max(h)
recent loss1(t)	If pixel transitioned from forest to non-forest in years [t, t - 2)	0, 1
recent loss2(t)	If pixel transitioned from forest to non-forest in years [t-2, t - 5)	0, 1
recent loss3(t)	If pixel transitioned from forest to non-forest in years [t-2, t - 5)	0, 1
recent loss4(t)	If pixel transitioned from forest to non-forest in years [t-2, t - 5)	0, 1
last_b30(t)†	Normalised TAO reflectance Landsat 7 band 3 (red) from the latest valid pixels in year t	0 - 255
last_b40(t)†	Normalised TAO reflectance Landsat 7 band 4 (NIR) from the latest valid pixels in year t	0 - 255
last_b50(t)†	Normalised TAO reflectance Landsat 7 band 5 (SWIR) from the latest valid pixels in year t	0 - 255
last_b70(t)†	Normalised TAO reflectance Landsat 7 band 7 (SWIR) from the latest valid pixels in year t	0 - 255

* optional layer

†latest available cloud-free observation at pixel

We constructed an additional feature from Hansen et al. (2013) data called $recent\_loss(t)$. Deforestation tends to cluster around an emergent point (contagion) so we wanted to encode the proximity, in time and space, of recent loss. Therefore, we wanted to have feature that summarised the information of neighbouring pixels' deforestation state. We chose to represent recent loss as four, one-hot encoded layers, each layer representing loss within a specified period in the recent past.

All our models were build so that they can take two or more tensors with the same spatial dimensions, which we define below, and forecast if deforestation is observed in the following year at the locating corresponding to the spatially-central pixel of these tensors.

The first 3D tensor that any of our models receives, which we named **"Static"**, is tensor of shape $\mathbf{S} \in \mathbb{R}^{2\times(2r+1)\times(2r+1)}$ where $(2r+1)\times(2r+1)$ is its spatial dimension and r is a predefined hyperparameter indicating the number of pixels the input tensor have in each spatial direction from the target central pixel (which has spatial coordinates $(r+1\times r+1)$ for an image of spatial size $(2r+1)\times(2r+1)$). The two channels of this tensor are $treecover2000$ and $datamask$. The third optional DSM layer may also be included and in this case there are three channels.

Our second set of tensors is a time series of 3D tensors $\mathbf{X}_{t-3}, \mathbf{X}_{t-2}, \mathbf{X}_t \in \mathbb{R}^{5\times(2r+1)\times(2r+1)}$, where again each tensor has spatial dimensions $(2r+1)\times(2r+1)$ but depth 5. The five channels of a tensor with time index t are $recentloss(t)$, $last\_b30(t)$, $last\_b40(t)$, $last\_b50(t)$ and $last\_b70(t)$ as defined in Table A.2. The depth of the time series stack can be varied but for the sake of illustration we will use a time series of 3 years. Finally, each tensor with time index t comes with a label $Y_{t+1} \in \{0,1\}$ which takes value 1 only if the target central pixel (at spatial location $r+1\times r+1$) is marked as deforested exactly in year $t+1$. To clarify this, here we note that if this pixel was labeled as deforested in any other year $t_j \neq t+1$,$lossyear_{t_j} = 1$, or was never labeled as deforested in the study period 2001-2018, $lossyear_t = 0 \forall t_j \in 1,2,..20$, then $Y_{t+1} = 0$.

Due to the characteristics of Hansen et al. (2013) dataset, we know that if a pixel is labelled as deforested in year t_j then the pixel never returns to the state of being forested. Additionally, if its the percentage of tree cover observed in 2000 was below 30%, than this location is not considered as forest. Only if a pixel with $treecover2000 < 30\%$ experience "gain" in the study period 2001-2012 we may assume it corresponds to a forested area from 2013 onward. Finally, if it has $datamask = 1$ then we know it is a permanent water body. Having stated this facts, we note that if our models aim to forecast the label of a pixel with index j , $Y^j_{t+1} \equiv \mathbb{I}\{lossyear_j = t+1\}$, they would not be of any use if we know that this pixel j is not a forested area in year t. It will never be reverted to forest and therefore detecting deforestation at this location in year $t+1$ doesn't make sense. Therefore, when predicting the labels of pixels ${Y^j}_{t+1}$ in year $t+1$, we restricted these set of pixels to be:

$$\mathbf{J}_t =: \{j \in \mathbb{M} : (lossyear_j > t \ \cup \ lossyear_j = 0)$$
$$\cap (datamask_j = 0) \cap (treecover_j > 30\% \ \cup \ gain_j = 1)\}$$

where $\mathbb{M}$ is the index set of pixels lying within Madre de Dios boundaries.

Since channel *treecover*2000 has range 0:100, and the Landsat bands 0:255, we rescaled each of them to be in the range 0:1.

For our last 3 models that utilize a time series of tensors we worked with the following dataset: $[\mathbf{S^j}, \mathbf{X}^j_{2014}, \mathbf{X}^j_{2015}, \mathbf{X}^j_{2016}]$ as set of input tensors and Y^j_{2017} as the set of labels to be predicted where $j \in \mathbf{J}_{2016}$. We split the data into train,validation and test with ratio 6:2:2 to select the best model of each class. We evaluated their performance on $[\mathbf{S^j}, \mathbf{X}^j_{2015}, \mathbf{X}^j_{2016}, \mathbf{X}^j_{2017}]$ as the set of input tensors and Y^j_{2018} as the set of labels to be predicted where $j \in \mathbf{J}_{2017}$.

Our Model 1, 2D CNN model, is able to analyze only mono-temporal tensors and from them to extract features forecasting the central pixel deforestation label in the following year. We used the union of the following data pairs of tensors and labels as dataset: $[\mathbf{S^j}, \mathbf{X}^j_{2014}]$ as an input tensors and Y^j_{2015} as the set of labels to be predicted where $j \in \mathbf{J}_{2014}$. $[\mathbf{S^j}, \mathbf{X}^j_{2015}]$ as an input tensors and Y^j_{2016} as the set of labels to be predicted where $j \in \mathbf{J}_{2015}$. $[\mathbf{S^j}, \mathbf{X}^j_{2016}]$ as an input tensors and Y^j_{2017} as the set of labels to be predicted where $j \in \mathbf{J}_{2016}$. We evaluated its performance on : $[\mathbf{S^j}, \mathbf{X}^j_{2017}]$ as an input tensors and Y^j_{2018} as the set of labels to be predicted where $j \in \mathbf{J}_{2017}$.

Here we note that to choose the best trained model form each model class, Model 2, Model 3, Model 4, we used as validation and text data that has labels in 2017, and therefore this models were biased towards the more recent year. Therefore, when choosing our best trained 2D CNN model, we use all data pairs to train it, but for early stopping validation data and model selection test data we used the pair $[\mathbf{S^j}, \mathbf{X}^j_{2016}]$ - Y^j_{2017}, $j \in \mathbf{J}_{2016}$.

A.3 Model architectures

A.3.1 Rationales

Deep learning methodologies are invaluable in attempting to make sense of the rapidly growing amount of satellite data being beamed back from space. Their capabilities of analysing satellite imagery of any resolution and providing unprecedented insights of a huge social and environmental importance from successfully predicting poverty (Jean et al., 2016) to anticipating undetected weather events (Hong et al., 2017). Zhu et al. (2017) gave a summary of the recent advances in the application of deep learning methods to remote sensing data. Forecasting spatial patterns of deforestation is a temporal extension of the well studied land use/cover classification task.

Kussul et al. (2017) investigated the performance of Random Forests classifier (RF), an ensemble of fully connected Neural Networks with single hidden layer (NN) and two ensembles

of Convolutional Neural Networks (the first 1D CNN and the second - 2D CNN) for crop classification. They used time series of six spectral Landsat-8 bands and two Synthetic-aperture radar (SAR) bands as an input feature vector (of total dimension 54) for each of the models. While the first 1D CNN model considered only the target pixel's channels as an input and used convolution in the spectral domain, the 2D CNN takes as an input a patch of pixels and aims to label the central pixel. Thus convolution was made in the spatial domain. They reported the overall classification accuracy of RF, the ensemble of NN and the ensemble of 1D and 2D CNNs to be 88.7%, 92.7%, 93.5%, and 94.6% respectively. Their results agree with the common belief that 2D convolutions in the spatial domain outperform not only in the computer vision domain but in the hyperspectral imaging as well. Therefore, we set our first model to be a 2D CNN that analyzes mono-temporal multi-spectral images.

Recent computer vision developments in 3D CNNs by Tran et al. (2015), where the third dimension usually refers to the time axis, has been utilized in the architecture of some hyperspectral classification networks. However, in the hyperspectral case, the third dimension usually refers to the spectral domain. Thus, in a 3D CNN, convolution operations are performed spatial-spectrally, while in 2D CNNs, they are done only spatially. Li et al. (2017) proposed a simple 3D CNN consisting of only two convolutional layers and one fully connected layer followed by a softmax layer. They compared its classification performance against a stacked autoencoder, a deep brief network and a 2D CNN on five different hyperspectral image datasets. While for the last three models, a prior Principal Component Analysis was performed to decrease the spectral dimension of the data, their simple 3D CNN architecture was applied on the hyperspectral images directly. Their model slightly outperform in all of the five data sets. The 2D CNN method was second best overall. An important note they made was that, as 3D CNN contains far fewer parameters to tune than a 2D CNN, it converges relatively quicker. A more complicated 3D CNN was also considered by Chen et al. (2016) who paid great attention on regularization techniques such as L2 regularization and Dropout. We note that the two proposed 3D CNNs models (Chen et al., 2016; Li et al., 2017) aimed to takes full advantage of both spectral and spatial information contained within HSIs data (where a pixel has more than 100 spectral bands), but they only analyzed mono-temporal cloud-free images.

The increasing temporal capabilities of today's sensors enable the use of temporal, along with spectral and spatial features. We decided to test the ability of a 3D CNN network to extract spatio-temporal features of forest loss.

When dealing with multidimensional time series data, Recurrent Neural Networks have proven their ability to mange time dependencies. Recently, they have started to be popular tool in the remote sensing community. In Lyu and Lu (2016) a land cover change detection task was addressed using a RNN model on a small time series. In their algorithm, a long

short-term memory (LSTM) based RNN is used to learn the joint spectral-temporal feature representation from a bi-temporal image sequence. In addition, Lyu and Lu (2016) showed the flexibility of their network by applying it to detect not only binary labels, but multi-class changes and showed that their framework performs just as good for change detection in an "unseen" scene without fine-tuning. The RNN-based approach of Ienco et al. (2017) was proposed for land cover classification. Both models, these of Lyu and Lu (2016) and Ienco et al. (2017), demonstrated their superiority over standard machine learning algorithms. However, in both of them a time series of single pixel channels is considered as an input and thus the spatial features of a location remains unexplored.

Conversely, as an attempt to utilize both the temporal and the spatial domain of a video data, a commonly used architecture in the computer vision field deals with first extracting high-level visual percepts from the top-layers of pretrained 2D CNNs and second, feeding them to an RNN in order to analyze temporal variation. This approach was first considered in the remote sensing field by Mou et al. (2019) in a framework named Recurrent Convolutional Neural Network (ReCNN). It detects land cover changes using bi-temporal satellite imagery. This architecture first learns the spatio-spectral high-level features of the two satellite images (each of which has 6 bands of resolution 30m^2 taken in different years) via 2D CNN encoder and then consequently feed these two features vectors ($\mathbf{f}^{T^1}, \mathbf{f}^{T^2}$) in a RNN architecture. They argued that the conditional probability of a pixel's label can be modelled by an RRN where $p(\mathbf{f}^{T^2}|\mathbf{f}^{T^1}) = \phi(\mathbf{h}^{T^2})$ and $\mathbf{h}^{T^2}$ is the final recurrent hidden state that the RNN branch outputs. This RNN idea was taken from Lyu and Lu (2016) where instead of $\mathbf{f}$, they used the single pixel channels. Finally, they applied a fully connected layer, followed by a Softmax activation to express ϕ.

They compared the performance of such a framework for both bi- and multi-class change detection on two different scenes and reported results achieved from long short-term memory (LSTM) based ReCNN, Gated Recurrent Unit (GRU) based ReCNN, Fully Connected ReCNN and a RNN framework as in Lyu and Lu (2016). They demonstrated that architecture like ReCNN can achieve state-of-art accuracy due to its ability to learn spatio-spectro-temporal features via its 2D encoder. ReCNN with a LSTM cell was reported to the best. The authors of Interdonato et al. (2019) used a similar approach to perform land cover classification task. They also combined a CNN and ReCNN architectures as a fusion of two branches and demonstrated how the efficient leverage of these two complementary sources of information together can outperform architectures where only one of them is considered (just the CNN or just the ReCNN branch).

Ballas et al. (2015) argued that methods where an RNN is used to analyze already extracted high-level visual percepts (as ReCNN does) suffer from capturing fine motion information due

to the spatial invariance of the top-layers of CNNs. They propose an architecture that applies an RNN not only on the 2D CNN top-layer but also on the intermediate convolutional layers. This was possible due to the use of Convolutional Recurrent Neural Networks (ConvRNN), a method where the input's spatial structure is preserved by replacing all matrix multiplications with convolutions. This idea was first introduced in the Remote sensing field by SHI et al. (2015) who argued that while the commonly used fully connected - LSTM layer well handle temporal correlation, it performs too much reduction on the spatial side of the data. They showed that the proposed model, characterised as an extension of the general LSTM that has convolutional operations on the place of the matrix multiplication operations, could be used as a network for general spatio-temporal sequence forecasting problems. This network architecture was then successfully applied for predicting sea level anomaly Braakmann-Folgmann et al. (2017). Marc Rußwurm and Marco Körner used this approach to improve their simpler LSTM RNN crop classifier, (Rußwurm and Körner, 2017), by substituting the fully connected LSTM operations with convolutional ones (Rußwurm and Körner, 2018b). Moreover, they noticed that this updated architecture is able to handle images which have not been filtered from clouds and achieved state-of-art accuracy.

The above innovations inspired us to trail 2D CNN, 3D CNN and ConvRNN architectures. They have demonstrated their power in land use/cover classification on data similar to that which available to us.

A.3.2 Components

Convolutional Neural Networks

Our data set was spatially organized into a grid with pixel-level observations. A leading type of Neural Network that are specially designed for analyzing grid-structured data is Convolutional Neural Networks (CNNs). CNNs have similarities with the general Artificial Neural Networks (ANNs) : they are still built up of filters that have learnable weights and biases and each of the filter takes some inputs, perform a dot product and activate it with a non-linear function. They differ in that they assume that the input data has a specific structure that determines their filters shape. Moreover, in each layer, the filters of CNNs are only linked with a local regions of the input tensor rather than with the full set of input's entries as are filters (neurons) in the fully connected layers of ANNs. Fig. A.1 illustrates how these two networks differ. Because in ANNs the outputs of the filters are scalar values, they are usually called neurons. In an ANN, a neuron n in layer l takes some input vector $\mathbf{x}$ (or the outputs of previous layer $l-1$), performs a dot product of all its values with its learnable weights $\mathbf{w}$ and add the bias term b. Then this neuron output is activated by a non-linear activation function f. Thus, a single neuron returns a

scalar feature after processing all the input data. Any information about the structure of the input is lost. In CNNs each filter slides ("convolve") across one or more dimensions of the input tensor, performs a dot product between its entries and the input entries at these local regions and record its response at each location in an activation map. The number of dimensions across which the filter slides determines the dimensionality of the output map. With 1D-, 2D- or 3D-CNNs one usually refers to the number of dimensions across which the filters slide. In this project we used models that involve 2D and 3D convolutional filters. Below we present the key concepts of 2D and 3D CNNs that one needs to know to understand the architecture of our models.

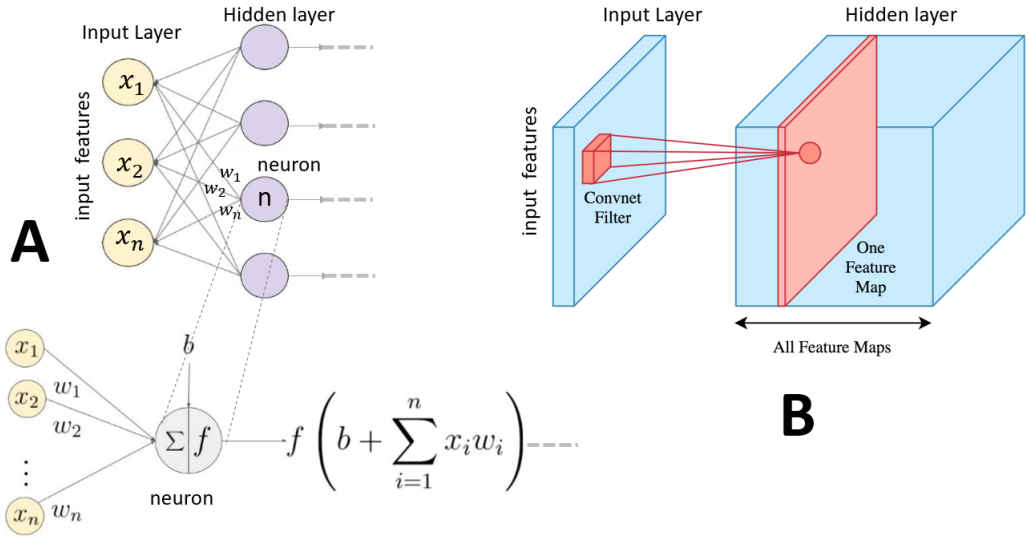


Fig. A.1 A) neuron (filter) in an Artificial Neural Network is connected to all neurons in the previous layer (input features). It preforms a dot product of its weights with the input feature vector and returns a scalar activation value. **B)** A filter in a 2D Convolutional Neural Network has 3D shape. It slides across the height and width of the input 3D tensor (features) and thus is connected to only one local region of the input at a time. It performs a dot product of its weights with the input entries at that location and map the activated value to a 2D map of activations (feature map). Adapted from: Fei-Fei Li (2017)

2D Convolutions

In this subsection we present the key concepts of 2D CNNs.

While a black and white image has a 2D rectangular shape, colour images are represented as a stack of three 2D grid maps each of which specify the intensity of Red, Green and Blue present at a particular location on the grid. The mixture of the three forms the colour. For hyper/multi-spectral images the number of channels is usually considerably larger. 2D CNNs for image processing constrain their filters to be made of set of spatially small 2D kernels, where the number of kernels always extends to the number of channels in the image. The stack

of these kernels along the channel axis forms the 3 dimensional filter. The filter slides across the height and width of the input, performs a dot product at each location and its bias term and map its activated responses to a 2D feature map (activation map). Fig. A.2 illustrates this 2D convolutional operation. In the figure the image has size $5 \times 5 \times 3$ (height, width, depth). Additionally, a zero padding of size 1 is applied which makes its size $7 \times 7 \times 3$. Padding means to add extra pixels outside the image (zero padding is when these added pixels have value 0 in all of their channels and the size of the padding defines how many pixels are added in each direction). In Fig. A.2 two filters are present in the first hidden layer of the network. Each of them has weights with dimension $3 \times 3 \times 3$ and a bias term. The image and the filters are shown as sliced across the channel domain. Take the first filter for example. It performs a dot product at each location and maps it to a 2D feature map. The equation for 2D Convolution is as follows:

$$v_{l,j}^{x,y} = f\left(\sum_{m=0}^{M-1}\sum_{h=0}^{h=H_j-1}\sum_{w=0}^{w=W_j-1} k_{h,w}^{ljm} v_{(l-1),m}^{(x+h),(y+w)} + b_{lj}\right)$$

Where l denotes the layer where the new output v is. j is the number of feature maps in this layer l and M is the number of feature maps (the depth of the input 3D tensor) in the previous $l-1$ layer. Applying j filters to an image results in output with j feature maps (3D output tensor of depth j). x and y are the spatial coordinates of v. $k_{h,w}^{ljm}$ is the h,w,m value of the j^{th} filter in layer l and b_{lj} is its bias term. f is a non-linear activation function. In Fig. A.2 the highlighted output (=-2) is the $x,y = 1,2$ entire of feature map $j = 1$ in layer $l = 2$. Here, the image is convolved at stride 2, where by stride 2 it is meant that the filter performs a dot products with local regions of the image that have their centers 2-pixels apart.

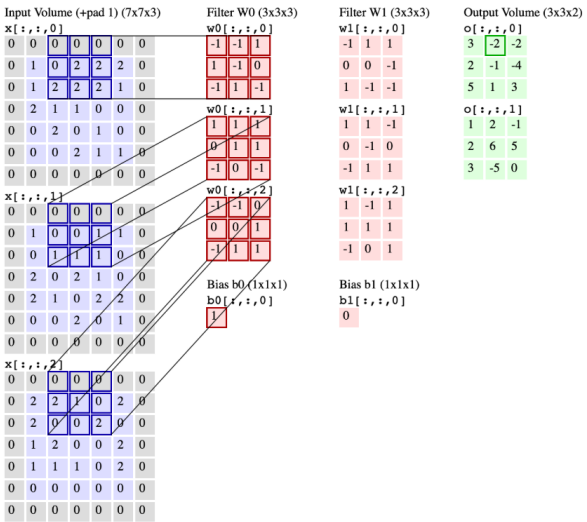


Fig. A.2 Convolution of 2 filters with an image. Sourced from (Fei-Fei Li, 2017)

3D Convolutions

In a 3D convolutional operation, filters slide across three of the dimensions of the input tensor and therefore they are made of a set of 3D kernels, where again this set size extends to the forth dimension of the input tensor. The stack of these 3D kernels makes the filter four dimensional tensor. As our data set has four dimensions: channels, time, height and width, if we slide the filters across the space and time domain, the network will be able to learn filters that get activated when they detect some type of spectral feature at particular time and location. 3D convolution is analogous to the 2D convolution with the difference being that the filter and the tensor has one more additional dimension. To visualize how 3D convolution works, consider Fig. A.3, where we present a four dimensional tensor as sliced across its forth dimension - the channel domain (the slices of the filter are denoted as 3D kernels). The filter slides along the three dimensions (height, width and time) of the input tensor, performs a dot product of its weights and the entries of the input tensor at these regions, add the bias and map the activated responses he gets at each spatio-temporal region to a 3D activation map. We present the equation for 3D convolution as Li et al. (2017) did. It is:

$$v_{l,j}^{x,y,z} = f\left(\sum_{m=0}^{M-1}\sum_{h=0}^{H_j-1}\sum_{w=0}^{W_j-1}\sum_{t=0}^{T_j-1} k_{h,w,t}^{ljm} v_{(l-1),m}^{(x+h),(y+w),(z+t)} + b_{lj}\right)$$

Where again l denotes the layer where the new output v is. j is the number of 3D feature maps in this layer l and M is the number of feature maps in the previus layer (the forth dimension of the input 4D tensor). x,y and t are the spatio-temporal coordinates of v. $k_{h,w,t,m}^{ljm}$ is the h,w,t,m weight of the j^{th} filter ($\in \mathbb{R}^{m\times H_j\times W_j\times T_j}$) and $b_{l,j}$ is its bias term. f is a non-linear activation function. Applying j filters to a 4D tensor results in output with j 3D feature maps, which when stack together, form the new four dimensional tensor of activations.

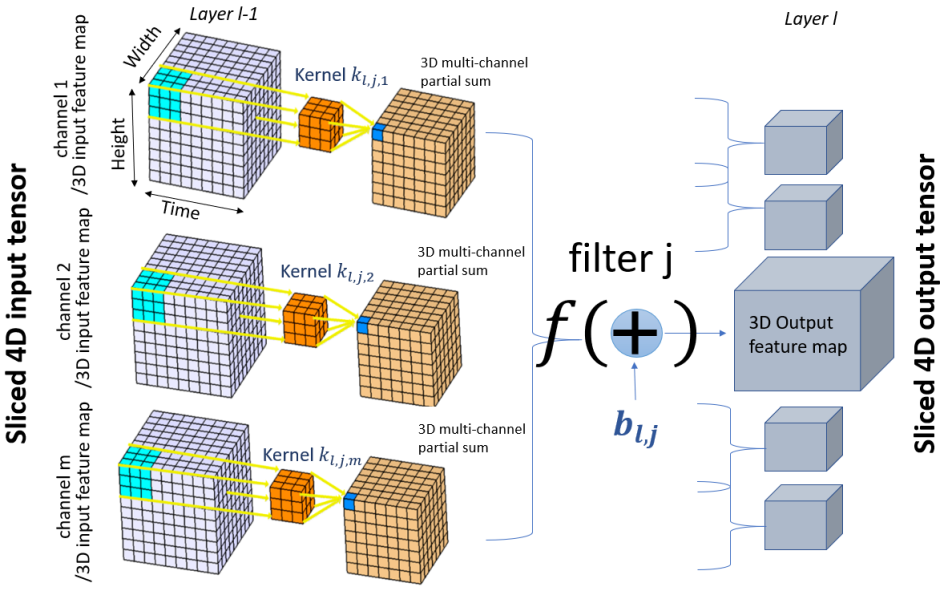


Fig. A.3 3D Convolution of 4D tensor.

Deep Convolutional Neural Networks

Networks are “deep" when they have more than one hidden layer. All layers presented in the figures above (Fig. A.1,Fig. A.2 and Fig. A.3) are the first hidden layer of the networks which take as an input the features of the input layer - the raw images. A layer is called hidden if it is not the input or the output layer. A network is called “Convolutional" if it has at least one convolutional filter in any of its layers. As we discussed above, in a 2D CNN each filter slides along 2 dimensions of the 3D input tensor and returns a 2D activation map. The stack of all filters’ activation maps in a layer is the new 3D tensor of high level features that will be propagated to the next layer. For a 3D convolution, the 3D activation maps of the filters in a layer are stacked to form the new 4D tensor. By propagating the image in such a way through the layers, the network is able to extract high level features.

Fig. A.10 illustrates the architecture of our first proposed model, which is indeed a deep 2D Convolutional Neural Network. Our 2D CNN model takes as an input a mono-temporal multispectral image, propagates it through the network and returns a softmax output that indicates the probability of observing deforestation at the location where the center of the image is. Hereby we note once again, the task of this network is classification forecasting, rather than nowcasitng. In Fig. A.10 one can see that each convolutional layer, activated with rectified linear unit (Hinton (2010) $ReLU(x) = max(x, 0)$), is followed by a layer called “2D Batch Normalization" layer. In the diagram of our model’s architecture one can also see layers

named as “Dropout" and “Spatial Pyramid Pooling Layer". In the rest of this section we explain what these layers do and discuss several issues one must consider when utilizing them.

Batch Normalization

Batch Normalization was introduced by Ioffe and Szegedy (2015). It is a layer that normalizes each filter to have a zero mean and unit variance. Ioffe and Szegedy (2015) showed that employing such layers in Neural Networks can be beneficial in several ways: the networks train faster as it enables the gradient descent algorithm to take higher learning rates; the convergence of the loss function is significantly less sensitive to how the weights are initialized; it offers some level of regularization by adding small noise to the data and sometimes can even work as well as dropout which can decrease the need of dropout layers present in the network. A 2D Batch Normalization layer in a CNN as proposed by Ioffe and Szegedy (2015) normalizes the entries of each feature map before their activation. The normalization is done by taking the mean and variance estimated across all locations and batches.

Consider a minibatch that has m 3D tensors (height, width and depth) that are convolved with the d filters of the current layer. The output of this layer is then a stack of d 2D feature maps. During training each of $(l,k)^{th}$ element $(l \in 0,2,..H_j, k \in 0,1..W_j)$ of the j^{th} $(j = 1,2,3,,..d)$ feature map that evolves from processing image under index b $(b \in 0,1,2,3...m)$ in the batch is transformed as follows:

$$\hat{y}_{b,j,l,k} = \hat{x}_{b,j,l,k} \times \gamma + \beta \qquad \hat{x}_{b,j,l,k} = \frac{x_{b,j,l,k} - \hat{E}[x_j]^{Moving}}{\sqrt{\hat{\sigma}_j^{Moving} + \varepsilon}}$$

where γ and β are learnable scale and shift parameters and ε is a constant added for numerical stability. Also:

$$\hat{E}[x_j]^{Moving} = \sum_{\mathbb{B}} \hat{E}[x_j] \qquad \hat{\sigma}_j^{Moving} = \frac{m \times H_i \times M_i}{m \times H_i \times M_i - 1} \sum_{\mathbb{B}} \hat{\sigma}[x_j]$$

Are the moving averages of the empirical mean($\hat{E}[x_j]$) and variance ($\hat{\sigma}^2[x_j]$) across batches (B is the batches index set). During inference they are kept as the constants. $\hat{E}[x_j]$ and $\hat{\sigma}^2[x_j]$ are obtained as follows:

$$\hat{E}[x_j] = \sum_{i=0}^{m} \sum_{h=0}^{H_j} \sum_{w=0}^{W_j} x_{i,j,h,w} \quad \hat{\sigma^2}_{x_j} = \frac{1}{m \times H_j \times M_j} \sum_{i=0}^{m} \sum_{h=0}^{H_j} \sum_{w=0}^{W_j} (x_{i,j,h,w} - \hat{E}[x_j])^2$$

Where H_j, W_j are the spatial size of the j^{th} feature map and $x_{i,j,h,w}$ is the $(h,w)^{th}$ entire of the j^{th} feature map of the i^{th} 3D tensor in the batch. The equations for 3D Batch Normalization is

analogous with the difference being in that feature maps are 3D tensors and hence:

$$\hat{E}[x_j] = \sum_{i=0}^{m}\sum_{t=0}^{T_j}\sum_{h=0}^{H_j}\sum_{w=0}^{W_j} x_{i,j,t,h,w} \quad \hat{\sigma}^2[x_j] = \frac{1}{m \times T_j \times H_j \times M_j}\sum_{i=0}^{m}\sum_{t=0}^{T_j}\sum_{h=0}^{H_j}\sum_{w=0}^{W_j}(x_{i,j,t,h,w} - \hat{E}[x_j])^2$$

In their original paper, Ioffe and Szegedy (2015) employed Batch Normalization by first normalizing the entries of the feature map $\hat{x}_{b,j,l,k} \rightarrow BN[x_{b,j,l,k}] = \hat{y}_{b,j,l,k}$ and then activating them by a non-linear activation function $f(\hat{y}_{b,j,l,k})$, e.g. ReLU. However, the case what should be the right order of applying f and $BN()$ to the input features is a topic of debate. Although up to our knowledge there is not a scientific body of work that address this problem, as many other experts in the field, we empirically showed that when ReLU is applied before a Batch Normalization layer : $y = BN[ReLU(x)]$, our networks perform better. The results of these experiments are shown in the Appendix.

Dropout

Dropout was proposed by Srivastava et al. (2014) as a technique for regularizing neural networks by adding noise to the entries of the hidden layer. More precisely, during training it works by multiplying the hidden activations with a Bernoulli random variable which takes value 0 with probability p or 1 - with 1-p respectively. As networks get deeper, the number of weights grows exponentially. This cause networks to overfit if no regularization measurements are employed. Employing dropout approximates an inexpensive way of training and inference of exponentially many networks. The way this is done is by randomly switching off different neuron units during each training forward pass. Without some of its neurons, the network represents a different function, or sub-network. When trained with dropout, the network cannot depend on any given neuron as it might be suddenly dropped out. This prevents it from learning features that depends on each other and also from returning an output that depends on one particular feature. During inference, we want the model to use all of its learned weights and not to drop out. When deployed, we multiply the scores of each neuron by the probability of it not being dropped $1-p$. To understand how it works, let us consider a feature vector of activations $\mathbf{x} = (x_1, x_2, ..x_d)$. If we apply dropout on this layer, the vector becomes $\mathbf{x} = (a_1 x_1, a_2 x_2, ...a_d x_d)$ where a_k are independent Bernoulli random variables. When testing, the vector becomes $\mathbf{x} = ((1-p) \times x_1, (1-p) \times x_2, ...(1-p) \times x_d)$ (See Fig. A.4). Alternatively, one can scale up the activations by multiplying them with $\frac{1}{1-p}$ during training and not modify them at inference (this is how pytorch implements it). In our models we have employed dropout in the second to last fully connected layer. The reason for that is because each of other layer of our networks is followed by a Batch Normalization layer. In their research Li et al. (2018) theoretically showed

that although both Batch Normalization layer and Dropout layer are very powerful regualization tools, their joint utilization in a neural network can lead to worse performance. The reason for this is because employing a Dropout layer has the side effect of shifting the variance of the neurons when the model is switched form training to inference state. On the other side, at inference the Batch Normalization layer maintains its statistical variance that has been learned during training. This variance dismatch, which Li et al. (2018) defined as “variance shift", makes the model unstable if a Dropout layer is applied before a Batch Normalization layer. Li et al. (2018) also confirmed their findings empirically by performing experiments on widely used networks architectures. Li et al. (2018) then suggested two methods that can prevent this variance disharmony when both regularization techniques are used: the simpler one is Dropout layer to be applied only after the last Batch Normalization layer, and the other is to use modified formula for the Dropout scaling factor 1-p. In our model we employed their first suggestion and only used dropout after the last Batch Normalization Layer. We confirmed their suggestion is valid by performing experiments on applying Dropout layer in several other layers, which were then followed by a Batch Normalization layer.

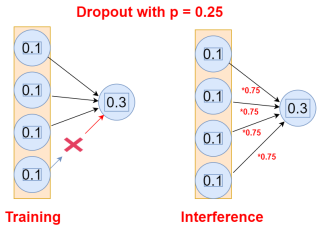


Fig. A.4 Applying dropout in a fully connected layer with probability of dropping a neuron p = 0.25. During inference each feature is multiplied by 1-p = 0.75.

Spatial Pyramid Pooling layer

The final piece of architecture present in the first two models that has not yet been discussed is the Spatial Pyramid Pooling layer. To begin with, general pooling layers are another type of layers commonly used in 2D CNNs architectures that are inserted between convolutional layers. However, their filters do not have learnable weights. Their only function is to progressively decrease the input spatial size which consequently decrease the number of network parameters. The effect of employing pooling layers is reduced number of computational operations and decreased overfitting. Pooling layers vary depending on the way they reduce the impute size: max-pooling layers, average-pooling layers, etc. The pooling filters slide along the image and independently downsample each depth slice of the input. Fig. A.5 demonstrate how a Maxpool filter of spatial size 2×2 slides across the input height and width at stride 2 and return the downsampled 3D tensor with decreased spatial size by factor of 2.

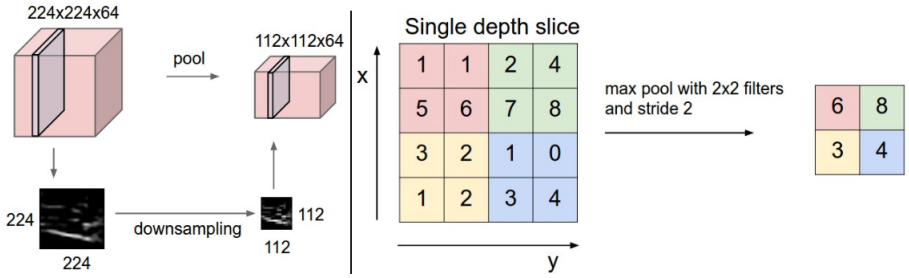


Fig. A.5 Independently downsamling each slice of an input volume with a Maxpooling filter of spatial size2 $\times$ 2 at stride 2. **Left:** Input volume of size 224x224x64 result in a volume of 112x112x64 (Note the depth of the volume is preserved). **Right:** Downsampling of a single slice with the corresponding Maxpool filter's kernel. The 2 $\times$ 2 kernel slides across the input slice and returns the maximum pixel value it receives at each location. Source: Fei-Fei Li (2017)

Standard 2D CNN architectures for image classification usually consist of several to many convolutional layers which are then followed by fully connected layers. The last high-level 3D feature map generated by the final convolutional layer is then flattened to a 1d vector and passed to the first fully connected layer. The size of this 1d feature vector is determined by the size of the final 3D feature map which also defines the number of filters the network has in the first fully connected layer. While the number of parameters in the convolutional layers depends only on the filters' sizes, the output of these layers depend on the filters and image sizes. Therefore in order to be able to design an architecture that has convolutional followed by fully connected layers, the size of the image must be predefined. He et al. (2014) proposed another pooling strategy, "Spatial Pyramid Pooling", that enables networks to be trained on images with various spatial sizes. They proposed 2D CNN with a spatial pyramid pooling layer between the last convolutional and the first fully connected layer, called "SPP-net", which has fixed number of parameters and is able to analyze images regardless of their spatial size. They demonstrated that employing the Spatial Pyramid Pooling strategy improve on many widely used CNNs architectures for image classification and object detection tasks that fit the input image to the required size by cropping or padding it. We decided to implement their strategy as we wanted to assess our model's performance when the input satellite images capture smaller or larger land regions without the need of prepossessing the images or modifying the models. This also allowed us to explore what is the optimal image size for a model with fixed number of parameters. Fig. A.6 illustrate how the Spatial Pyramid Pooling layer takes a 3D tensor with arbitrary spatial size and returns a fixed-size feature vector.

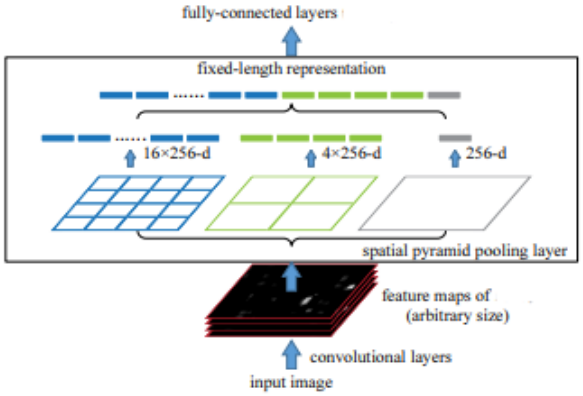


Fig. A.6 Input 3D tensor of arbitrary spatial size and depth 256 is forwarded to a spatial pyramid pooling layer that has 3 filters. The three filters slide across the spatial domain of the input and return 3D tensors of size $16 \times 16 \times 256$, $4 \times 4 \times 256$ and $1 \times 1 \times 256$ respectively. There tensors are then flattened and connected together to form a feature vector of fixed size = $256 \times (16 \times 16 + 4 \times 4 + 1 \times 1)$ that is then the input of the following fully connected layer. Source: He et al. (2014)

A Spatial Pyramid Pooling layer has predefined fixed number of pooling filters. One can chose the function with which these filters downsample: max, average, etc. However, the spatial size of these filters and the stride at which they slide across the spatial domain of the input tensor is dynamic, it adapts according to the spatial size of the input. This filters are then able to return 3D tensors of predefined spatial size and same depth as the input tensor. Having fixed number of 3D output tensors with fixed spatial size and depth, when flattened and contacted together, they form a fixed-size feature vector. In Fig. A.6 the spatial pyramid pooling layer has predefined number of filters 3 with predefined spatial sizes 16×16, 4×4 and 1×1 respectively. The input 3D tensor has arbitrary spatial size and depth of 256. After flattening and connecting the output of each pooling filter, the extracted 1d feature vector has size $256 \times (16 \times 16 + 4 \times 4 + 1 \times 1)$. The equations that govern the dynamics of the pooling filters' spatial sizes and stride is as follows:

$$s_h = floor(H_x/H_k) \qquad s_w = floor(W_x/W_k)$$
$$k_h = floor(H_x/H_k) + H_x mod(H_k) \qquad k_w = floor(W_x/W_k) + W_x mod(W_k)$$

Where s_h, s_w are the strides at which the pooling filter must slide across the height H_x and width W_x of the input 3D tensor. This filter must also have height and width k_h and k_w respectively. The output after sliding it across the input tensor has the desired height and width H_k, W_k and same depth d as the input tensor depth.

Recurrent Neural Networks

Rather than learning features that can explain the current state of a pixel, our model differs in that it tries to learn features that can identify if the target area will become deforested the following year. Our hypothesis is that those features can be extracted from past and current satellite images and moreover, that the temporal structure of these observations contains information that can be beneficial for the accuracy of our prediction - e.g if we assume that a deforestation event is not a sudden event but rather a continuous process, the model may be able to learn the target area's deforestation tendency.

There are two ways a model can learn these features - one is called feed forward and the other - recurrent. So far we presented our CNN model and 3D-CNN model as feed forward learning models - the signals flow one direction only: from the input layer of image features to the final layer producing the output (class label $\hat{Y}_{t+1}$). In such models the output of any layer does not affect the current or any of the previous layers. In our second model, Model 2: 3D CNN, we extract features $\mathbf{s}_{t+1}$ given a 4D input tensor that is formed by concating sequence of inputs images $\mathbf{x}_{t-k}, \mathbf{x}_{t-(k-1)}, .., \mathbf{x}_t$ with a fixed window size k along the time axis and propagate it forward through the network until the final layer that has softmax activation indicating our confidence of the central pixel label.

In a recurrent learning process, the signals may travel both forward and backward direction by introducing loops in the network. Fig. A.7 outline the main difference between feed-forward and recurrent learning networks. The forward propagation of a simple "vanilla" Recurrent Neural Network with only one hidden layer is governed by the following equation:

$$\mathbf{s}_t = \phi(\mathbf{W}_s \times \mathbf{s}_{t-1} + \mathbf{U}_x \times \mathbf{x}_t + \mathbf{b}) = \phi([\mathbf{W}|\mathbf{U}] \times [\mathbf{s}_{t-1}|\mathbf{x}_t]^T + \mathbf{b})$$

Where:
$\times$ is matrix multiplication operation,
$\mathbf{s}_i \in \mathbb{R}^h$, $\mathbf{x}_i \in \mathbb{R}^m$, $\mathbf{W} \in \mathbb{R}^{h \times h}$, $\mathbf{U} \in \mathbb{R}^{h \times m}$, $\mathbf{b} \in \mathbb{R}^h$,
so that $[\mathbf{W}|\mathbf{U}] \in \mathbb{R}^{h \times (h+m)}$ and $\mathbf{b} \in \mathbb{R}^{\approx}$ are the weight matrix and the bias vector of the hidden layer and $[\mathbf{s}_{i-1}|\mathbf{x}_i]^T \in \mathbb{R}^{h+m}$ is its input. ϕ is a non-linear squashing function, usually hyperbolic tangent. We set $\mathbf{s}_0 = \mathbf{0}_h$.
We see that in thus defined forward propagation, the output of a hidden layer $\mathbf{s}_{k-1}$ is looped back and contacted with the new input image $\mathbf{x}_k$. This new input ($[\mathbf{s}_{t-k}|\mathbf{x}_k]^T$) is then passed again to the same hidden layer and a new "updated" $\mathbf{s}_k$ output is returned, allowing the network to process the images sequentially. The iteration continues until all input images are learned. At each step, the output $\mathbf{s}_k$ has "learned" all the input images the network have seen so far: $\{\mathbf{x}_{k-1}, \mathbf{x}_{k-2}..\}$).

Due to this internal memory, RNN's are able to remember important information about the new inputs they received, which makes them powerful in predicting what's coming next.

This ability to "memorize" have made them very popular in tasks that deals with sequential data type. Depending on the objective, several types of architectures have been developed. One possible architecture is "many-to-many" where at each iteration the model use the most recent $\mathbf{s}_k$ to predict the future output (in our case this can be class label $\hat{Y}_{k+1} = \theta(\mathbf{s}_k)$). In this case the loss is evaluated as sum of all individual losses: $L_{tot} = \sum_k L_k = BCEloss(\hat{Y}_{k+1}, Y_{k+1})$. Other tasks may require a single prediction of a future event, given past time-series data, for which one usually use the architecture "many-to-one". In such a model, we may only use the final "most knowledgeable" feature vector $\mathbf{s}_t$ to make a prediction and may only report the final loss $L = BCEloss(\hat{Y}_{t+1}, Y_{t+1})$. This two differences are reflected in the red square of Fig Fig. A.7 **B)**. Due to their empirically proven power, many variations of RNNs have been developed.

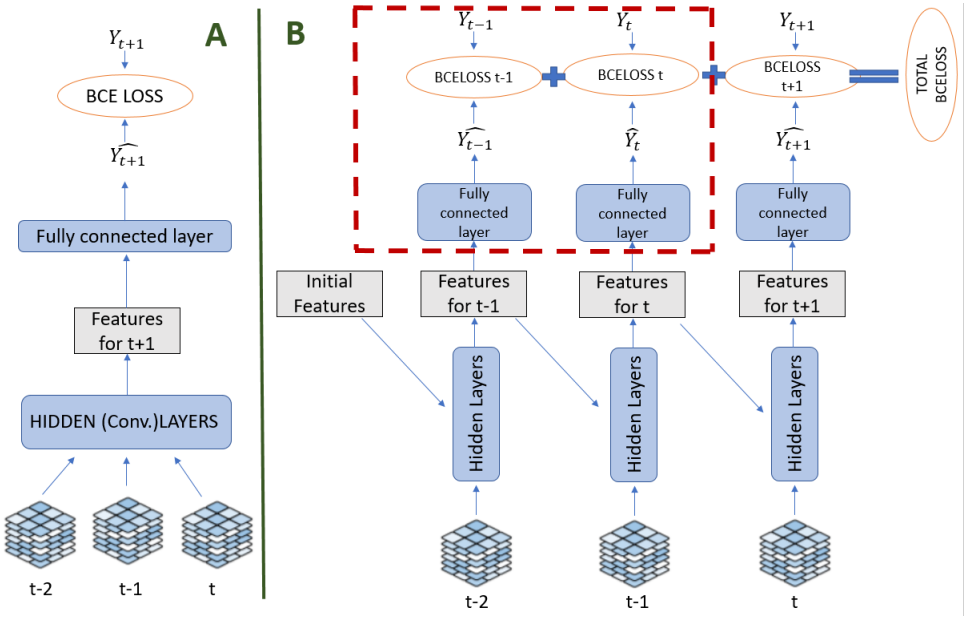


Fig. A.7 A) CNN model with feed forward architecture: The signals from the images travel one direction only,from the input layer to the output layer; **B)** Vanilla Recurrent Neural Network has recurrent learning: the output of the hidden layer is looped back, cocated with the new image and forwarded to the same layer again. When the output of the layer is also forwarded to the next layer at each time, the model has "many-to-many" architecture. When the hidden layer forwards the output only at the last time itteration, the architecture is "many-to-one".

Long Short Term Memory Cell

Nevertheless, when dealing with long sequential data, during learning, back- propagation, a "Vanilla" RNN may suffer from exploding or vanishing gradients. While for the first case, exploding gradients, truncating the gradients can solve the undesirable effect, the vanishing

gradients problem is much harder to overcome. Fortunately, Hochreiter and Schmidhuber (1997) proposed the concept of Long-Short Term Memory RNN structure which was able to solve the vanishing gradients problem. Since than, several variations of the LSTM RNNs were proposed. Below we briefly outline the concept of the LSTM cell as defined in the Deep learing book by Goodfellow et al. (2016). The governing equations in an LSTM cell are as follows:

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \mathbf{h}_{t-1} + \mathbf{U}_f \mathbf{x}_t + \mathbf{b}_f) = \sigma([\mathbf{W}_f | \mathbf{U}_f] \times [\mathbf{h}_{t-1} | \mathbf{x}_t]^T + \mathbf{b}_f)$$
$$\mathbf{i}_t = \sigma(\mathbf{W}_i \mathbf{h}_{t-1} + \mathbf{U}_i \mathbf{x}_t + \mathbf{b}_i) = \sigma([\mathbf{W}_i | \mathbf{U}_i] \times [\mathbf{h}_{t-1} | \mathbf{x}_t]^T + \mathbf{b}_i)$$
$$\mathbf{o}_t = \sigma(\mathbf{W}_o \mathbf{h}_{t-1} + \mathbf{U}_o \mathbf{x}_t + \mathbf{b}_o) = \sigma([\mathbf{W}_o | \mathbf{U}_o] \times [\mathbf{h}_{t-1} | \mathbf{x}_t]^T + \mathbf{b}_o)$$
$$\mathbf{g}_t = \tanh(\mathbf{W}_g \mathbf{h}_{t-1} + \mathbf{U}_g \mathbf{x}_t + \mathbf{b}_g) = \tanh([\mathbf{W}_g | \mathbf{U}_g] \times [\mathbf{h}_{t-1} | \mathbf{x}_t]^T + \mathbf{b}_g)$$
$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t$$
$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t)$$

$\times$: matrix multiplication operator

$\odot$: the Hadamard (element-wise) product

$\mathbf{x}_i \in \mathbb{R}^d$: input vector to the LSTM unit

$\mathbf{h}_i \in \mathbb{R}^h$: output vector of the LSTM unit (also known as hidden state)

$\mathbf{g}_i \in \mathbb{R}^h$: input activation vector

$\mathbf{c}_i \in \mathbb{R}^h$: LSTM internal state vector

$\mathbf{f}_i \in \mathbb{R}^h$: forget gate's activation vector

$\mathbf{i}_i \in \mathbb{R}^h$: input gate's activation vector

$\mathbf{o}_i \in \mathbb{R}^h$: output gate's activation vector

$\mathbf{W} \in \mathbb{R}^{h \times h}, \mathbf{U} \in \mathbb{R}^{h \times d}$ and $\mathbf{b} \in \mathbb{R}^h$: weight matrices and bias vector parameters

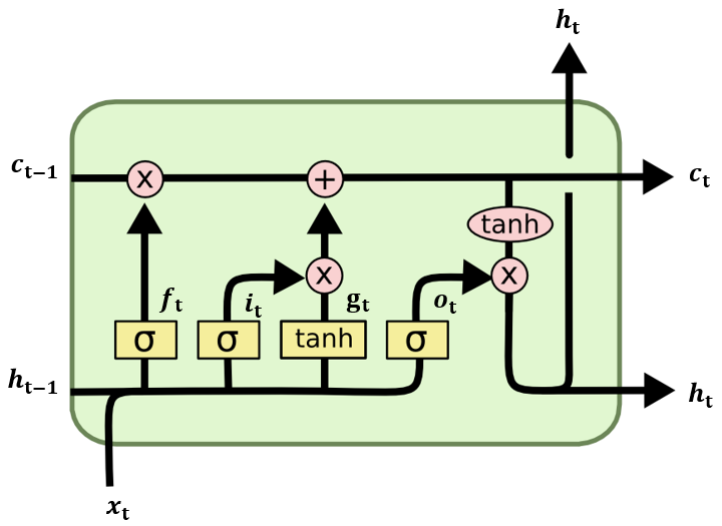


Fig. A.8 Information flow in a LSTM cell. Sourced from: Hoffman (2018)

Fig. A.8 illustrate this LSTM cell architecture. A new feature vector $\mathbf{g}_t$ is learned as in the "vanilla" RNN case - a regular neuron units (with weights matrix and bias vector $\mathbf{W_g}|\mathbf{U_g}|\mathbf{b_g}$) are fed with the connected new input vector $\mathbf{x}_t$ and the previous output vector $\mathbf{h}_{t-1}$. In the case above, we have used hyperbolic tangent function to squash the neurons' outputs but any squashing nonlinear function can be used. In a LSTM architecture, however, this combination of present input and past output is also fed to three other gate vectors, which will decide how the new learned feature vector $\eth_{\approx}$ will be handled: they block, pass or output the new information based on its importance, which they have filtered with their own set of weights. How much of $\mathbf{g}_t$ will be accumulated in the internal state cell $\mathbf{c}_t$ is controlled by the input gate vector $\mathbf{i}_t$. A forget gate vector $\mathbf{f}_t$ decides how much of the current cell state information $\mathbf{c}_{t-1}$ will be blocked. Finally, the new output, the updated memory $\mathbf{c}_t$, is squashed with a hyperbolic tangent function but before being returned, it is regulated by the output gate vector $\mathbf{o}_i$. All the gate unit vectors have a sigmoid nonlinear activation, or also called squashing, function. Some variations of the LSTM architecture use the state cell as an extra input to the gate units. By temporally stacking several LSTMs layers above one another, a deeper network can be achieved.

Convolutional Long Short Term Memory Cell

We now extend this idea of LSTM cell with the recently developed Convolutional LSTM structure. The modification of this LSTM RNN architecture as Convectional was first proposed by SHI et al. (2015). In the same year Ballas et al. (2015) proposed Convolutional RNN with Gated Recurrent unit (unit similar to the LSTM unit). The goal of SHI et al. (2015) work was to develop an algorithm that gives precise prediction of rainfall intensity in a local region over

a short period of time, a problem which they defined as "spatio-temporal sequence forecasting". They used a sequence of past radar maps as an input and a sequence of a fixed number future radar maps as an output, task very similar to ours. Furthermore, they evaluated the performance of their model, which they named ConvLSTM, on the Moving-MNIST dataset and showed that ConvLSTM is able to achieve state-of-art accuracy on any task that deals with spatio-temporal sequence forecasting. Their design differs from the general LSTM cell in that all the inputs $\mathbf{x_1}, \mathbf{x_2}, ...\mathbf{x_t}$, hidden states $\mathbf{h_1}, \mathbf{h_2}, ...\mathbf{h_t}$, internal state cells $\mathbf{c_1}, \mathbf{c_2}, ...\mathbf{c_t}$ and gates of ConvLSTM are 3D tensors whose last two dimensions define their spatial dimensions. One can imagine them as vectors standing on a spatial grid. For each cell in that grid the ConvLSTM determines its future state and output by analyzing the input features and the past states of its local neighbours. This is achieved by substituting the matrix multiplication operations with convolutions. Below we provide the key equations of the ConvLSTM cell that we emplyed in our model. We adopted the ConvLSTM equations from Rußwurm and Körner (2018b) and make note that they differ form those proposed by in that SHI et al. (2015) LSTM cell structure uses the state cell as an extra input to the gate units.

$$
\begin{aligned}
\mathbf{f}_t &= \sigma(\mathbf{W}_f * [\mathbf{h}_{t-1}|\mathbf{x}_t] + \mathbf{b}_f) \\
\mathbf{i}_t &= \sigma(\mathbf{W}_i * [\mathbf{h}_{t-1}|\mathbf{x}_t] + \mathbf{b}_i) \\
\mathbf{o}_t &= \sigma(\mathbf{W}_o * [\mathbf{h}_{t-1}|\mathbf{x}_t] + \mathbf{b}_o) \\
\mathbf{g}_t &= \tanh(\mathbf{W}_f * [\mathbf{h}_{t-1}|\mathbf{x}_t] + \mathbf{b}_g) \\
\mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t \\
\mathbf{h}_t &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t)
\end{aligned}
$$

$*$ denotes the convolution operator
$\odot$ denotes the Hadamard (element-wise) product
d, w, h: the depth, width and height of the input tensor
r: number of channels of the hidden state, internal state cell and all the gates tensors
$k_1 \times k_2$: kernel size of the feature maps
$\mathbf{x}_i \in \mathbb{R}^{d \times w \times h}$: input tensor to the LSTM unit
$\mathbf{h}_i \in \mathbb{R}^{r \times w \times h}$: output tensor of the LSTM unit (hidden state)
$[\mathbf{h}_{i-1}|\mathbf{x}_i] \in \mathbb{R}^{(r+d) \times w \times h}$: connected input tensor
$\mathbf{g}_i \in \mathbb{R}^{r \times w \times h}$: input activation tensor
$\mathbf{c}_i \in \mathbb{R}^{r \times w \times h}$: LSTM internal state tensor
$\mathbf{f}_i \in \mathbb{R}^{r \times w \times h}$: forget gate's activation tensor
$\mathbf{i}_i \in \mathbb{R}^{r \times w \times h}$: input gate's activation tensor

$\mathbf{o}_i \in \mathbb{R}^{r\times w\times h}$: output gate's activation tensor
$\mathbf{W} \in \mathbb{R}^{r\times(r+d)\times k_1\times k_2}$ are the weights of the r stacked convolutional featues maps (each $\in \mathbb{R}^{(r+d)\times k_1\times k_2}$) and $\mathbf{b} \in \mathbb{R}^r$ is a vector with the bayes each of the r features maps has. The initial state cell $\mathbf{c_0}$ and $\mathbf{h_0}$ are initialized as zero value tensors. To preserve the spatial size of the convolutional outputs we used zero padding of size $[\frac{k_1}{2}]\times[\frac{k_2}{2}]$ where we used kernels with odd sizes k_1, k_2.

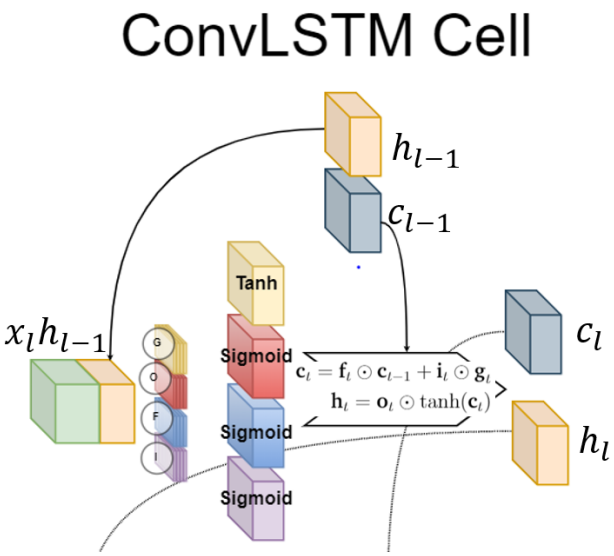


Fig. A.9 Convolutional LSTM Cell

A.3.3 Model Architecture 1: 2D Convolution Neural Network

Our first proposed model, which we consider as our base model due to its simple architecture, is a 2D Convolution Neural Network with four 2D convolutional layers (2D Conv), one Spatial Pyramid Pooling layer (SPP) and 2 fully connected layers (FC) with a Dropout layer(DO) in between them. The final fully connected layer takes as input feature vector of size 100 and returns a scalar value squashed by a sigmoid non-linearity function, σ. The output is therefore in range [0,1] and indicates the model confidence of observing deforestation at the location corresponding to the spatially central pixel (with spatial coordinates $(r+1, r+1)$) of the input tensor in the following year. The filters of each layer are activated with ReLU and this activations are then normalized by a Batch Normalization layer(BN).

For a datapoint j, the input of this model is a 3D tensor, $\mathbf{SX}_t^j$, that dimension $\mathbb{R}^{7\times(2r+1)\times(2r+1)}$. This input results from stacking its static 3D tensor, $\mathbf{S^j} \in \mathbb{R}^{2\times(2r+1)\times(2r+1)}$, and one 3D tensor, $\mathbf{X_t^j} \in \mathbb{R}^{2\times(2r+1)\times(2r+1)}$, of j^s time series of 3D tensors $\{\mathbf{X_k^j}\}_{k=t-2}^{t}$. They are stacked along the channel axis so that the input of the network becomes $\mathbf{SX}_t^j$. The predicted label is $\hat{Y}_{t+1}^j$ where

$j \in \mathbf{J}_t$ as defined above. Due to the SPP layer our model is capable of analyzing 3D tensor of any spatial size, which is regulated by r.

We have constructed this model such that we can vary the spatial size, the stride and the padding of the filters of any of the 2D convolutional layers (2DConv). The number of filters per convolutional layer is also set as free model parameter. The Spatial Pyramid Pooling layer (SPP) parameters are also allowed to vary, the number of pooling filters ,n, and the spatial size of this filters ,$\mathbf{k} \in \mathbb{R}^{2,n}$. By varying them one changes the size of the first fully connected layer (FC), as its size depend on the Spatial Pyramid Pooling layer output. Finally, the last free model parameter is the Dropout rate, p, with which the Dropout layer(DO) switches-off some of the connections between the two fully connected layers while training. As suggested by Li et al. (2018) we only employed a dropout layer after the last Batch Normalization Layer (BN). In this model architecture each convolutional layer is activated by ReLU, then these activations are normalized with a 2D Batch Normalization layer. The normalized features are then propagated to the next convolutional layer ($2DConv \rightarrow ReLU \rightarrow 2DBN \rightarrow 2DConv$). The normalized activations of the last 2D convolutional layer are then downsampled by the SPP layer and passed to the first FC layer. Its output of size 100 is activated again with ReLU and normalized with 1D BN layer. Only then dropping is implemented. We summarize the propagation of this model as follows: $\mathbf{SX}_t^j \in \mathbb{R}^{7\times(2r+1)\times(2r+1)} \rightarrow 4\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow$ SPP($n,\mathbf{k}$) $\rightarrow$ FC(spp,100) $\rightarrow$ ReLU $\rightarrow$ 1DBN $\rightarrow$ DO(p) $\rightarrow$ FC(100,1) $\rightarrow \sigma \rightarrow \hat{Y}_{t+1}^j$

Fig. A.10 illustrates these architecture.

A.3.4 Model Architecture 2: 3D Convolution Neural Network

Our first proposed model is able to take advantage of the spatial and spectral domain of the input due to its 2D convolutional layers. However, it is not able to utilize the time domain of a datapoint j. It analyzes a 3D tensor $\mathbf{SX_t}^j$ that arses from stacking the Static features of j, $\mathbf{S}^j$, and only one element of j^{th} time series $\mathbf{X}_t^j$.

Our second model, 3D CNN, takes on the other side takes full advantage of the time domain and thus is able to analyze j^{th} spectral,spatial and time domain simultaneously. This is achieved via 3D convolutional layers present in its architecture, that consists of two branches which meet together to return the final prediction. More precisely, for the j^{th} data point the model takes as an input the j^s static tensor $\mathbf{S}^j \in \mathbb{R}^{2\times(2r+1)\times(2r+1)}$ and its full time series of non static tensors $\{\mathbf{X}_{t-2}^j, \mathbf{X}_{t-1}^j, \mathbf{X}_t^j\}$, each $\in \mathbb{R}^{5\times(2r+1)\times(2r+1)}$. It then propagate these two parts of the input to two different branches. The static tensor $\mathbf{S}^j$ is passed to a 2D Convolutional Branch and the time series of tensors $\{\mathbf{X}_k^j\}_{k=t-3}^t$ to a 3D Convolutional Branch. Each branch extract high-level features which are then propagated together in the rest of the network. Finally, the

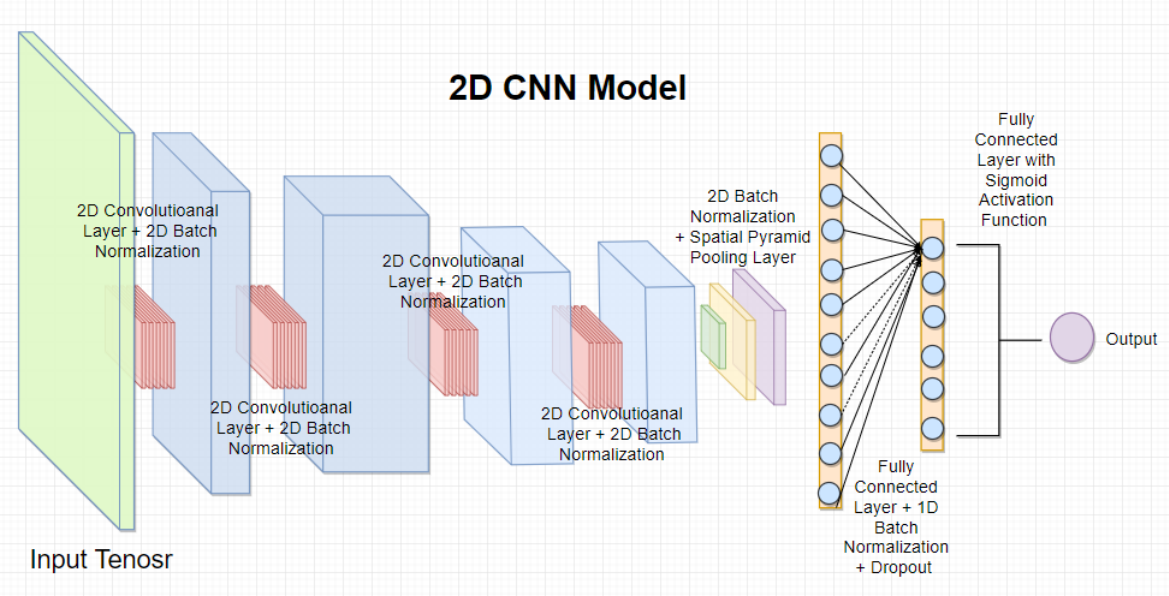


Fig. A.10 Model 1: 2D CNN. Architecture summary:
Input tensor $\mathbf{SX}_t^j \in \mathbb{R}^{7\times(2r+1)\times(2r+1)} \rightarrow 4\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow$ SPP($n,\mathbf{k}$) $\rightarrow$ FC(spp,100) $\rightarrow$ ReLU $\rightarrow$ 1DBN $\rightarrow$ DO(p) $\rightarrow$ FC(100,1) $\rightarrow \sigma \rightarrow \hat{Y}_{t+1}^j$

model returns an output $\hat{Y}_{t+1}^j$ indicating the confidence of the model to observe deforestation at the target location in the following year t+1 where $j \in \mathbf{J}_t$.

Here we give more detailed explanation what each branch does. $\{\mathbf{X}_k^j\}_{k=t-3}^t$ is passed to a 3D Convolutional Branch in the form of 4D tensor obtained by stacking the sequence by the time domain. Thus the input tensor $\mathbf{X}^j \in \mathbb{R}^{5\times3\times(2r+1)\times(2r+1)}$ has its first domain defined by the channels, the second by the time and the last two, by the space. The 3D Convolutional Branch "convolve" with $\mathbf{X}^j$ across its last three dimensions, time, height and width in two sequential 3D convolutional layers. Due to our limited time domain, of size 3, the 4D filters have shape $\in R^{c,2,k_h,k_w}$, where the first dimension extends to the number of the input channels and the last three define the shape of its 3D kernels. While setting different values to their spatial sizes is possible, the size of its time domain could only be 2. The model slides its filters' 3D kernels along thee time domain at stride 1 and no padding is applied on the input 4D tensor. Therefore, after the two 3D convolutional layers the output of the 3D Convolutional Branch was a 3D tensor of high-level features with no time domain, $\mathbf{Z}_x^j$. We summarize this propagation as follows: $\mathbf{X}^j \in \mathbb{R}^{5\times3\times(2r+1)\times(2r+1)} \rightarrow 2\times$(3DConv$\rightarrow$ReLU$\rightarrow$3DBN)$\rightarrow \mathbf{Z}_x^j \in \mathbb{R}^{c_1,h,w}$

$\mathbf{S}^j$ is passed to a 2D Convolutional branch that "convolve" with the input along the spatial domain in two sequential 2D convolutional layers (conv. layers) and return a 3D tensor of high level features, $\mathbf{Z}_s^j$. We set the model to have convolutional filters in both branches of the same spatial size so that both 3D tensors of high-level filters evolving from the two branches to have

the spatial size,$(h\ timesw)$. We summarise this branch as follows: $\mathbf{S}^j \in \mathbb{R}^{2\times(2r+1)\times(2r+1)} \rightarrow 2\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN)$\rightarrow \mathbf{Z}_s^j \in \mathbb{R}^{c_1,h,w}$

This two 3D tensors of high-level features,$\mathbf{Z}_x^j, \mathbf{Z}_s^j$, returned form each branch are then stacked along their third domain to form the 3D tensor $\mathbf{Z}^j$, and propagated it to the rest of the network. The final part of the network has another two 2D convolutional layers and after the last convolutional operation, the output is propagated to a SPP layer, two FC layers with DO in between and a sigmoid squashing function as in our CNN model. We summarise the final part of the network as follows: $\mathbf{Z}^j \in \mathbb{R}^{(c_1+c_2),h,w} \rightarrow 2\times$(2DConv $\rightarrow$ ReLU $\rightarrow$ 2DBN) $\rightarrow$ SPP$(n,\mathbf{k}) \rightarrow$ FC(spp, 100) $\rightarrow$ ReLU $\rightarrow$ 1DBN $\rightarrow$ DO$(p) \rightarrow$ FC$(100,1) \rightarrow \sigma \rightarrow \hat{Y}_{t+1}^j$

In this model architecture, again, we utilized 2DBN and 3DBN between each 2D and 3D convolutional layers after applying ReLU activation function. The number of filters in each of the 2D and 3D conv layers was set as free parameter, as well as the spatial size of the filters. Parameters of SPP and DO layers are allowed to vary too. The model is able to analyze tensors of any spatial size.This model flexibility allowed us to experiment with its architecture. Fig. A.11 illustrate our second model architecture.

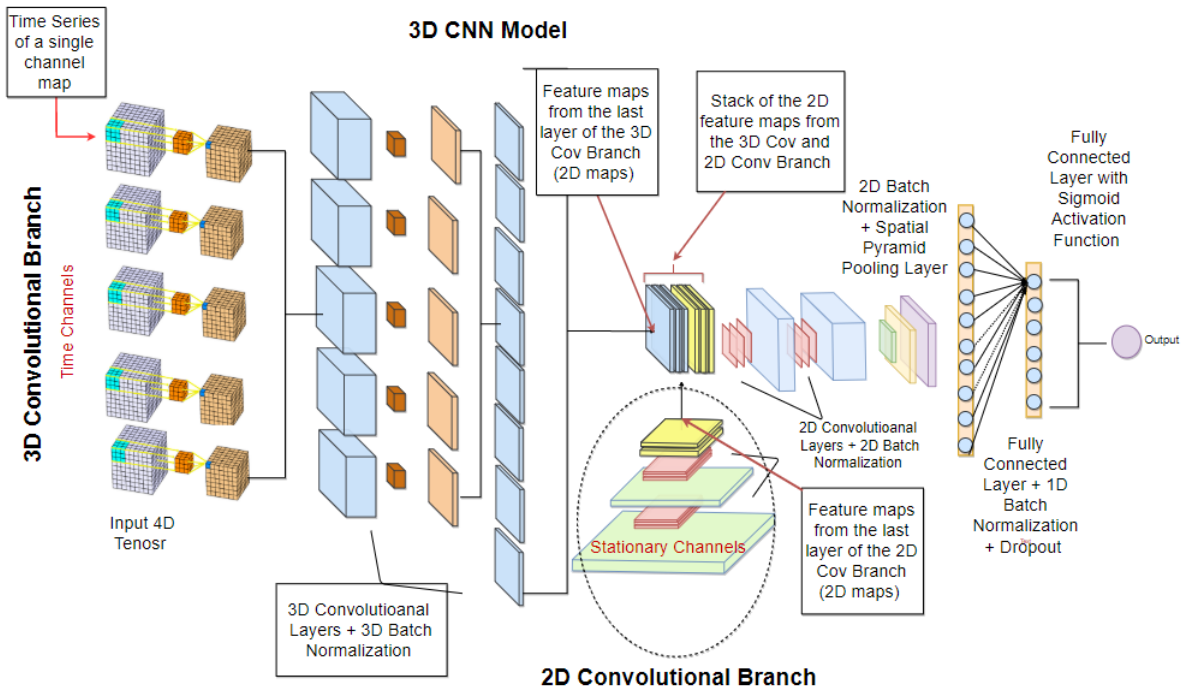


Fig. A.11 Model 3: 3D CNN Model. Architecture summary:
$\mathbf{X}^j \in \mathbb{R}^{5\times3\times(2r+1)\times(2r+1)} \rightarrow 2\times$(3DConv$\rightarrow$ReLU$\rightarrow$3DBN)$\rightarrow \mathbf{Z}_x^j \in \mathbb{R}^{c_1,h,w}$
$\mathbf{S}^j \in \mathbb{R}^{2\times(2r+1)\times(2r+1)} \rightarrow 2\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN)$\rightarrow \mathbf{Z}_s^j \in \mathbb{R}^{c_2,h,w}$
$\mathbf{Z}^j \in \mathbb{R}^{(c_1+c_2),h,w} \rightarrow 2\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow$ SPP$(n,\mathbf{k}) \rightarrow$ FC(spp,100) $\rightarrow$ ReLU $\rightarrow$ 1DBN $\rightarrow$ DO(p) $\rightarrow$ FC(100,1) $\rightarrow \sigma \rightarrow \hat{Y}_{t+1}^j$

A.3.5 Model Architectures 3 & 4: Convolutional and Deep Convolutional Long Short Term Memory Recurrent Neural Network

While our second, 3D CNN, model was able to utilize all three dimensions of the input data, it has feed-forward learning. When one has sequential data, recurrent learning is usually very powerful learning process. This inspire us to design our next two models so that they learn the temporal structure of the input data recurrently. More precisely, they both have Convolutional Long Short Term Memory cell, or cells, in their architecture.

Both our Model 3 and Model 4 have architecture similar to that of our 3D CNN model. For a data pint j, they take as an input its static tensor $\mathbf{S}^j \in \mathbb{R}^{2\times(2r+1)\times(2r+1)}$ and its time series $\{\mathbf{X}^j_{t-2}, \mathbf{X}^j_{t-1}, \mathbf{X}^j_t\}$, each $\in \mathbb{R}^{5\times(2r+1)\times(2r+1)}$. Again the static tensor $\mathbf{S}^j$ and the time series of tensors $\{\mathbf{X}^j_k\}^t_{k=t-3}$ are separately propagated through two separate branches, **(Encoder + Conv LSTM RNN)** and **2D CNN Static**, each of which return 3D tensors of high-level features ,$\mathbf{Z}^j_x, \mathbf{Z}^j_s$. $\mathbf{Z}^j_x, \mathbf{Z}^j_s$ are then stacked together along their channel axis to form $\mathbf{Z}^j$. This joint 3D tensor is propagated through the rest of the network. Finally, the model return $\hat{Y}^j_{t+1}$ where $j \in \mathbf{J}_t$.

Here we give detailed explanation what these two branches does. The branch that takes $\{\mathbf{X}^j_{t-2}, \mathbf{X}^j_{t-1}, \mathbf{X}^j_t\}$ of both Model 3 and Model 4 differs form this in our second model,3D CNN, in two ways, it has Encoder sub-branch and a Convolutional Long Short Term sub-branch, **ConvLSTM**.

The **Encoder** encodes the time series of 3D tensors $\{\mathbf{X}^j_k\}^t_{k=t-3}$ to another time series of 3D high-level features $\{\tilde{\mathbf{X}}^j_k\}^t_{k=t-3}$. It does so by feeding each $X^j_{t_j}$ through the same three 2D conv.layers. We summarise this sub-branch as follows:**Encoder** : $\{\mathbf{X}^j_k\}^t_{k=t-3} \rightarrow \{\tilde{\mathbf{X}}^j_k\}^t_{k=t-3}$:
$\mathbf{X}_t \in \mathbb{R}^{5\times(2r+1)\times(2r+1)} \rightarrow 3\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow \tilde{\mathbf{X}}_\mathbf{t} \in \mathbb{R}^{c_1\times h_1\times w_1}$
$\mathbf{X}_{t-1} \in \mathbb{R}^{5\times(2r+1)\times(2r+1)} \rightarrow 3\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow \tilde{X}_{t-1} \in \mathbb{R}^{c_1\times h_1\times w_1}$
$\mathbf{X}_{t-2} \in \mathbb{R}^{5\times(2r+1)\times(2r+1)} \rightarrow 3\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow \tilde{\mathbf{X}}_{\mathbf{t-2}} \in \mathbb{R}^{c_1\times h_1\times w_1}$

This time series of high-level 3D features, $\{\tilde{\mathbf{X}}^j_k\}^t_{k=t-3}$, is then handled recurrently buy the **ConvLSTM** sub-branch.**ConvLSTM** of Model 3 pass this new time series to one 2D Convolutional Long Short Term Memory cell and after the last iteration of this cell, this branch of the model returns a 3D feature tensor, $\mathbf{Z}^\mathbf{j}_\mathbf{x}$, that stores the essential spatio-temporal "memory" of the model input. Model 4 differs from Model 3 in that its **ConvLSTM** sub-branch is "deep". More precisely, the time series of high-level features, $\{\tilde{\mathbf{X}}^j_k\}^t_{k=t-3}$, is passed to a stack of several, say d, ConvLSTM cells. The latest in time output of the deepest cell is returned as a 3D feature tensor that stores the essential spatio-temporal "memory", $\mathbf{Z}^\mathbf{j}_\mathbf{x}$. We sumarise this sub-branch of Model 4 as follows: **Conv LSTM RNN:** $\{\tilde{\mathbf{X}}^j_k\}^t_{k=t-3} \rightarrow d\times$(ConvLSTM) $\rightarrow \mathbf{Z}^j_x \in \mathbb{R}^{c_2,h_2,w_2}$

The branch of Model 3 and 4 taking $\mathbf{S}^j$ is identical and also very similar to that of Model 2. $\mathbf{S}^j$ is passed to three sequential 2D convolutional layers (conv. layers) and return a 3D tensor of high level features, $\mathbf{Z}_s^j$. Here again we set the model to have convolutional filters in this branch such that the output of this branch has the same spatial size as the output of the **ConvLSTM** sub-branch. Thus, this branch, which we denote as **2D CNN Static**, returns $\mathbf{Z}_s^j \in \mathbb{R}^{c_3,h_2,w_2}$. We summarise this branch as follows: **2D CNN Static:** $\mathbf{S}^j \in \mathbb{R}^{2\times(2r+1)\times(2r+1)}$ $\rightarrow$ 3$\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN)$\rightarrow$ $\mathbf{Z}_s^j \in \mathbb{R}^{c_3,h_2,w_2}$

Finally, the outputs of the two branches,$\mathbf{Z}_x^j$,$\mathbf{Z}_s^j$, are stacked together along the channel axis to form $\mathbf{Z}^j \in \mathbb{R}^{(c_2+c_3),h_2,w_2}$ which is then propagated through the rest of the network. We denote this part of the network as **2D CNN Joint**. It consists of three 2D conv. layers, SPP layer and two 2 FC layers with DO layer in between. The output of the last FC layer is squashed with sigmoid non-linearity, σ, and has value $\hat{Y}_{t+1}^j \in [0,1]$.We summarise this propagation as follows: **2D CNN Joint** $\mathbf{Z}^j \in \mathbb{R}^{(c_2+c_3),h_2,w_2}$ $\rightarrow$ 3$\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow$ SPP(n,$\mathbf{k}$) $\rightarrow$ FC(spp,100) $\rightarrow$ ReLU $\rightarrow$ 1DBN $\rightarrow$ DO(p) $\rightarrow$ FC(100,1) $\rightarrow$ σ $\rightarrow$ $\hat{Y}_{t+1}$

Fig. A.12 illustrate the model architecture of Model 4, that has $d = 2$ ConvLSTM cells one above another in its **ConvLSTM** sub-branch. Here again, we have designed our models to be flexible with respect to their hyper-parameters. For Model 4 one can explore how the model performance change when the number of filters and their spatial size of each layer present in the network take different values, by changing the depth, d, of the ConvLSTM network section, by controlling the overfitting effect with different dropout ratio, p, and the downsampling rate of the parameters of the SPP layer. Finally, as with our other models, the input tensors can be of any spatial size ($(2r+1)\times(2r+1)$).

A.4 Methodological notes

Data within the training periods was split into training, validation and test sets by 3:1:1.

A.4.1 Training

For Model Architectures 2, 3 and 4 models were first trained on 2014-2017 input data matched with 2018 labels. They were then trained on 2015-2018 input data on 2019 labels.

Model Architecture 1 was trained on 2014 input with 2015 labels, 2015 input with 2016 labels, 2016 input with 2017 labels, 2017 input with 2018 and 2018 input with 2019 labels all at once.

5-fold cross validation took place during the training routines to track how well the models were learning.

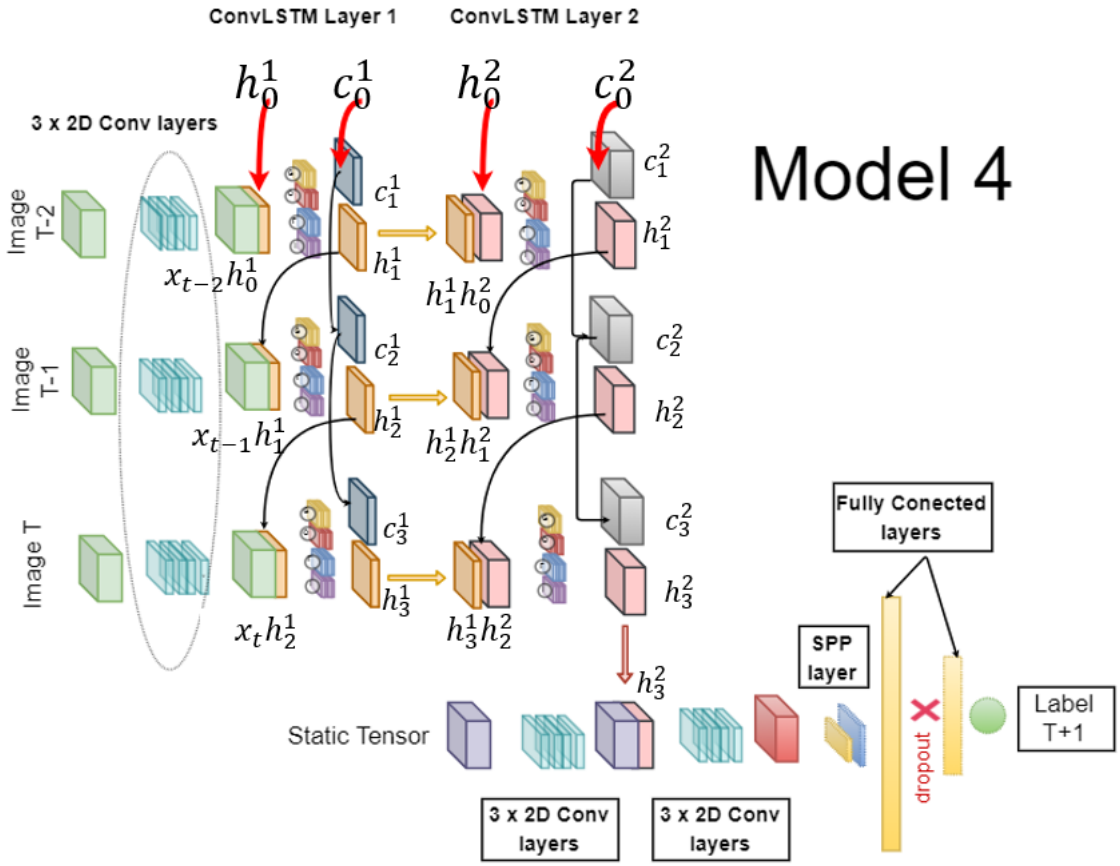


Fig. A.12 Model 4: Deep ConvLSTM Neural Network with 2 ConvLSTM cells. Architecture summary:
Encoder: $\{\mathbf{X}_k^j\}_{k=t-3}^{t} \rightarrow \{\tilde{\mathbf{X}}_k^j\}_{k=t-3}^{t}$
$\mathbf{X}_t^j \in \mathbb{R}^{5\times(2r+1)\times(2r+1)} \rightarrow 3\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow \tilde{\mathbf{X}}_t^j \in \mathbb{R}^{c_1\times h_1\times w_1}$
$\mathbf{X}_{t-1}^j \in \mathbb{R}^{5\times(2r+1)\times(2r+1)} \rightarrow 3\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow \tilde{\mathbf{X}}_{t-1}^j \in \mathbb{R}^{c_1\times h_1\times w_1}$
$\mathbf{X}_{t-2}^j \in \mathbb{R}^{5\times(2r+1)\times(2r+1)} \rightarrow 3\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow \tilde{\mathbf{X}}_{\mathbf{t-2}}^{\mathbf{j}} \in \mathbb{R}^{c_1\times h_1\times w_1}$
Conv LSTM RNN: $\{\tilde{\mathbf{X}}_k^j\}_{k=t-3}^{t} \rightarrow 2\times$(ConvLSTM) $\rightarrow \mathbf{Z}_x^j \in \mathbb{R}^{c_2,h_2,w_2}$
2D CNN Static:
$\mathbf{S}^j \in \mathbb{R}^{2\times(2r+1)\times(2r+1)} \rightarrow 3\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN)$\rightarrow \mathbf{Z}_s^j \in \mathbb{R}^{c_3,h_2,w_2}$
2D CNN Joint:
$\mathbf{Z}^j \in \mathbb{R}^{(c_2+c_3),h_2,w_2} \rightarrow 3\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow$ SPP$(n,\mathbf{k})$ $\rightarrow$ FC(spp,100) $\rightarrow$ ReLU $\rightarrow$ 1DBN $\rightarrow$ DO(p) $\rightarrow$ FC(100,1) $\rightarrow \sigma \rightarrow \hat{Y}_{t+1}^j$

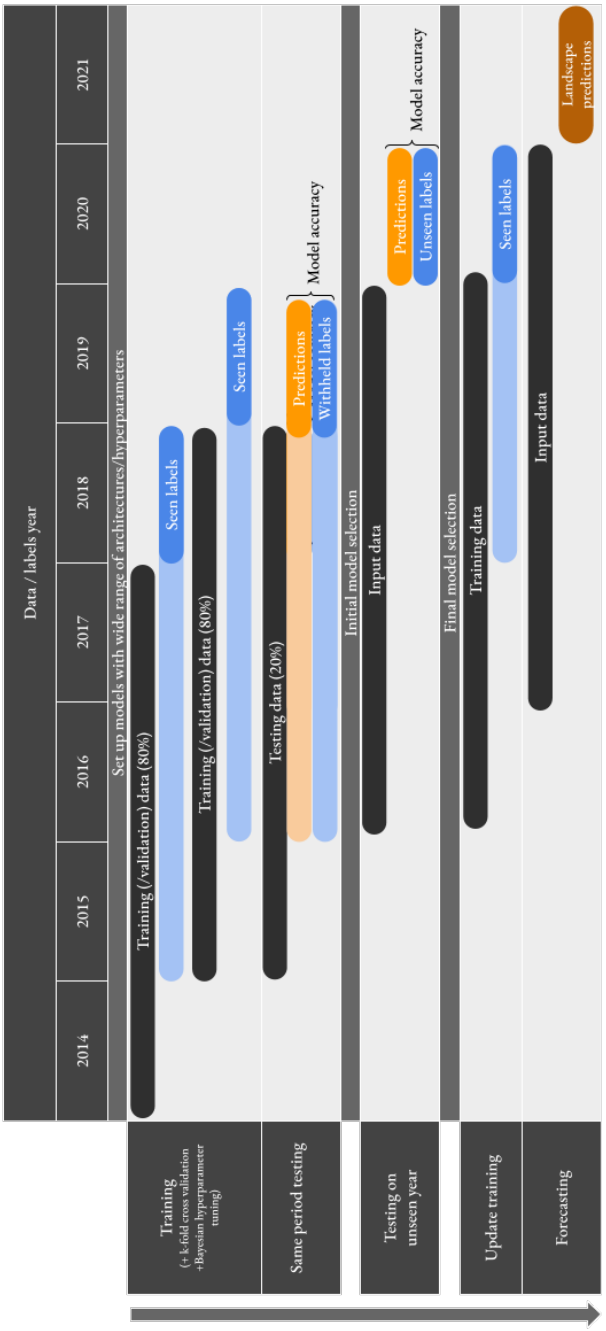


Fig. A.13 Arrangement of training, testing and forecasting processes

A.4.2 Model tuning

For the initial broad model inter-comparison we conducted a grid search to find effective model parameters. For the development of the most promising models we used a Bayesian hyperparameter sweep Snoek et al. (2012) as implemented on WandB.

A.4.3 Testing

To understand how well models would forecast they had to be tested on a time period that was outside of the training period. To do this we see how well they can predict deforestation in 2020 (an unseen year for the model). The models that can predict deforestation that happened 2020 with the best accuracy are identified and their parameters are recorded.

A.4.4 Forecasting

The models with the best parameters were updated with training data from 2015 to 2019 (with 2020 labels). The model weights were retained to keep the learning that had taken place on the 2014-2018 data.

The updated model is used to forecast deforestation for 2021. It produced a risk map across all forested land in the area of interest. A threshold based on anticipated overall levels of deforestation could be applied to the map at a required level (e.g. from 0 to 1) to identify pixels that are likely to undergo deforestation.

A.4.5 Class Imbalance Problem

The number of pixels that become deforested in Madre de Dios 2017 and 2018 accounted for around 0.30%, 0.26% of the total number of pixels covering the forested study area on which our models were trained. This ratio of $99.7 : 0.3$ made our dataset extremely imbalanced. Therefore, by simply prediction all pixels as forested one would get misleading accuracy of 99.7%, but such model will useless model. To address this issue we took several steps. Our objective was to be able to identify the top 20% most susceptible areas with high accuracy so that attention could be. In reality, a perfect model predicting up to 20% deforestation labels would be able to achieve no more than 80.3% accuracy as 19.7% non-deforested pixels would be labelled as suspected. Therefore, when testing and training our models we needed to make class-aware sampling. Recently, this class imbalance problem was addressed by Buda et al. (2018) in the context of Deep Learning models. They examined the impact such an imbalance can have on CNNs classifiers via several experiments and empirically concluded that the effect of it is detrimental. They also investigated the efficiency of several algorithms dealing with class

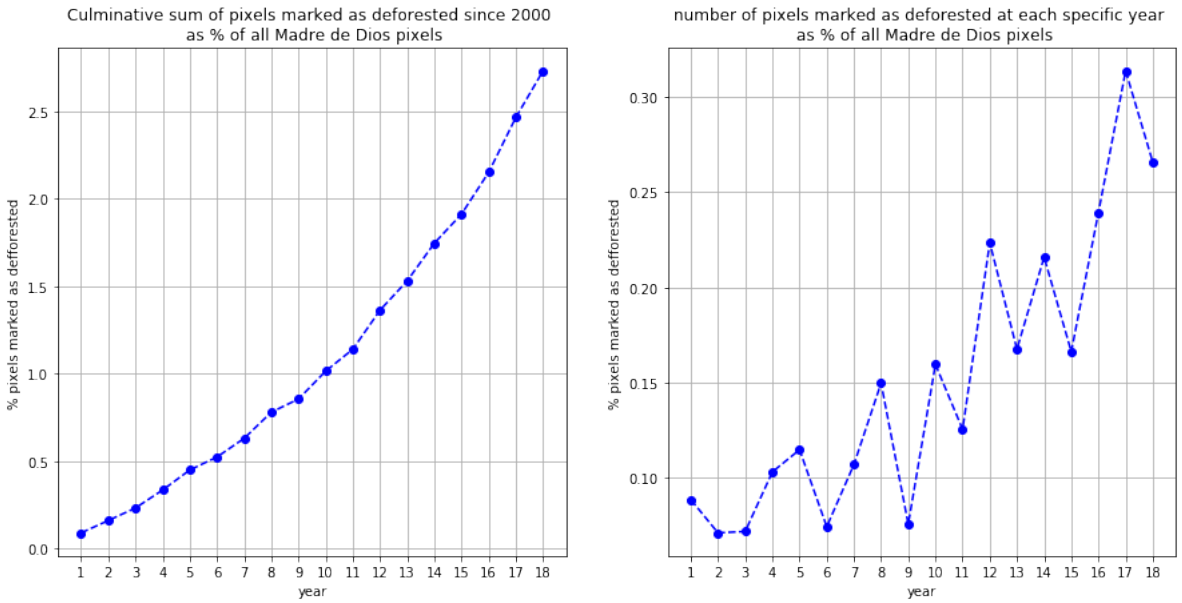


Fig. A.14 Percentage deforested pixels of all pixels covering Madre De Dios area.

imbalance, commonly used in the Machine Learning field, when applied on Deep Learning methods. Their results shows that the best approaches in the DL field are oversampling, or if training time is a problem, undersampling is an alternative, where sampling must be done to the extend where the imbalance is completely removed.

While we assumed (Buda et al., 2018) conclusion that undersampling is the right approach when dealing with imbalanced data, the extend to which this should be done we set as a "free",tuning, parameter. Our reasoning for this is that the cost of missing to detect future deforestration event is much higher than raising a wrong deforestation alarm.

To explore what should be the ratio of forested to deforested pixels in our train, validation and test data, we developed a sampling algorithm, which when given a dataset $\mathbf{D}_t$ and a parameter θ, returns a new undersampled dataset $\tilde{\mathbf{D}}_t$. This new dataset $\tilde{\mathbf{D}}_t$ has all data-points of the the minority class, pixels labelled as deforested ($Y^j_{t+1} = 1,\ \ j \in \mathbf{J}_t$), but the number of observations form the majority class, forested pixels($Y_{t+1} = 0\ \ j \in \mathbf{J}_t$), is set to be such that the ratio of the two classes in $\tilde{\mathbf{D}}$ is θ. The observations from the majority class are randomly selected and at each call of this algorithm, the set of observation drawn from the majority class was updated with new, distinct, set of randomly selected forested observations. The following section explains how we utilized this algorithm in our models training process.

A.4.6 Optimizer and loss function

In our training process we used Adam optimizer (Kingma and Ba, 2014) and Binary Cross Entropy Loss.

A.4.7 Early Stopping regularization

Early stopping is a regularization technique that is used when there is an iterative learning process. Several variations of this technique exist, but the one we utilize aims to make the model a better fit to data outside of the training set, validation dataset.

Here we describe the way we employed early stopping regularization. After each epoch, the model under training was set to interference state and its performance was evaluated on a separate validation data set. The ratio of training data size to validation data size was kept to 60:20. Each data set has equal ratio of 0 and 1 labels, the ratio of forested to deforested pixels labels, which was set to θ. At each epoch iteration, if the performance of the model on this validation data set was better than its performance in the previous iteration according to a certain criterion, the leering process continue to iterate through the batches one more time until new epoch is completed. If however, the performance on this iteration is worse, a counter variable starts counting how many times the updated model after each epoch is worse than the previously best one. That is, if the model gives tree times in a row, i+1, i+2, i+3, worse performance than the one recorded at iterations i, the counter has value 3. The counter counts up to pre-defined "waiting time" and is reset to 0 each time a new, better performing model is found. In our training strategy we set the "weighting time" to be 3. Which means that if our model performs three times in a row worse than it did at iteration i, then we stop the learning process, and chose the model with weights evaluated at iteration i. If the model however performs worse at iteration i+1 and i+2 than it did at i, but at i+3 it is better according to the criteria, the counting process restart and the model. The model at iteration i+4 is now compared with the previously best, i+3.

The above described algorithm is general, and well known within DL community. The way we modified this algorithm is by re-sampling our taring and validation data sets after each epoch by calling our sampling algorithm with input parameter θ , which preserves the ratio of 0 to 1 labels in both train and validation data set, but update the under-sampled class of 0 labels, with new, distinct, randomly sampled 0 labels. We also investigated the effect of employing this Early Stopping technique with different criteria, namely the Area Under The Receiver Operating Characteristics (AUC), Weighted Binary Cross entropy Loss and the Cost from Cost matrix. As training time was an issue, we only run our algorithm up to 5 epochs, with "waiting time" 3 and the best performing model was returned. Fig. A.15 illustrates the training process

of three different models, where for each different criterion was selected and demonstrate the advantages of utilizing Early Stopping regularization.

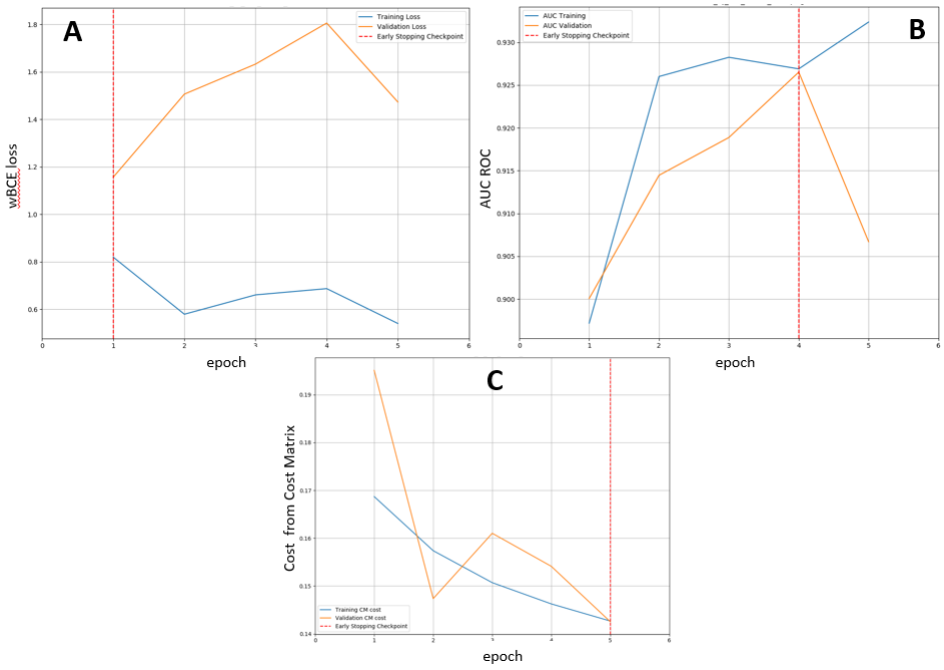


Fig. A.15 Training with an Early Stopping regularization.
A The selected criterion is Weighted Binary Cross entropy Loss. The selected model is the one learning its weights up to epoch 1. The model overfits afterwards. While the loss evaluated on the train data decrease (blue line), on validation data the loss it increase (yellow line) with minimum validation loss recorded at epoch 1.
B The selected criterion is AUC ROC. The selected model is the one learning its weights up to epoch 4. The model does not overfit up to epoch 4. However, the AUC on validation data decrease at the next epoch 5. The learning process is ended due to time limitations.
C The selected criterion is Cost from Cost matrix. The selected model is the one learning its weights up to last epoch 5 when both train and validation cost was at its lowest figure. The can not be said to overfit. However, while a smooth decrease of the cost was observed on train data, the validation cost has drop at epoch 2 and spike at epoch three.

A.4.8 Training models with mono-temporal and multi-temporal inputs

Our four models' ability to correctly forecast deforestation events is compared using labels corresponding to the year 2018, $\hat{Y}^j_{2018}$, for $j \in \mathbf{J}_{2017}$. The input data for our last three models, Model 2, Model 3, and Model 4, is $\{(S^j, X^j_{2015}, X^j_{2016}, X^j_{2017}) | j \in \mathbf{J}_{2017}\}$, whereas Model 1 takes as input $\{(S^j, X^j_{2017}) | j \in \mathbf{J}_{2017}\}$.

When training, the only possible training data for Model 2, Model 3, and Model 4 is $\{(S^j, X^j_{2014}, X^j_{2015}, X^j_{2016}) | j \in \mathbf{J}_{2016}\}$ with corresponding labels $\hat{Y}^j_{2017}$, for $j \in \mathbf{J}_{2016}$. This data,

before being undersampled (with a ratio of 0:1 labels = 99.7:0.3), was divided into training, validation, and test data subsets with a ratio of 6:2:2. The training and validation datasets were used for training with Early Stopping regularization. We used the test data to select the best 'tuned' model of each class. The ratio of 0 to 1 labels was kept the same in all three subsets: training, validation, and test. The sampling algorithm with parameter θ was then used on each of the data subsets when we examined the effect of different imbalance ratios. However, when training our CNN model, different setups for training data are possible. In the subsection below, we discuss how we addressed that issue.

A.4.9 Training and testing CNN models with different time set up

Our first, 2D CNN, model takes a singe 3D-tensor as an input $\mathbf{SX}_t^j \in R^{7\times 2r+1\times 2r+1}$ and output a integer score $\hat{x}_t^j \in [0,1]$ that can be interpreted as our confidence level of pixel j to be deforested in year t+1. Then, for specified threshold tr we assign $Y_{t+1}^j = I_{x_t > tr}$. Considering the above notations, the data set for a CNN model is then defined as $D_t = \{(\mathbf{SX}_t^j], Y_{t+1}^j) | j \in \mathbf{J}_t\}$ if only one year data is used. However, one can also use the more than one year data by disregarding the time index of the data but only using features with one year lag from their predicted labels. To be more precise, a set up as $\mathbf{D}_{t_k} = D_{t-k} \cup D_{t-(k-1)} \cup .... \cup D_{t-1} \cup D_t$ is possible, in which case k+1 years of data is utilized. This possibilities made us ask several questions. Firstly, what is the optimal number of year, k? Secondly, if one year is used, is our first model time invariant? That is, if we train it on $D_{2015} = \{(\mathbf{SX}_{2015}^j], Y_{2016}^j) | j \in \mathbf{J}_{2015}\}$, how it performs on the two data-sets $D_{16} = \{(\mathbf{SX}_{16}^j], Y_{17}^j) | j \in \mathbf{J}_{2016}\}$ and $D_{2015} = \{(\mathbf{SX}_{2017}^j], Y_{2018}^j) | j \in \mathbf{J}_{2017}\}$?. Finally, to be compared with the rest of our models, that have taring and validation and test data sets with labels in 2017, how should we train our CNN model so as to be equally biased towards labels in 2017 while simultaneously utilizing all years data.

To answer the last, we did as follows, while the training data year was allowed to vary. However, for validation and test data we used 20% and 20% of the original D_{2016} data (where the ratio of 0:1 labels was 99.7 : 0.3) respectively. The rest 60% were considered as part of the training data. We then applied the sampling algorithm with parameter θ on each of the data subsets to examine the effect of different imbalance ratios θ.

After we selected our best 2D CNN model, we used its hyper parameters to train a model on $D_{2015} = \{(\mathbf{SX}_{2015}^j], Y_{2016}^j) | j \in \mathbf{J}_{2015}\}$ and examined how it performs on the two data-sets $D_{2016} = \{(\mathbf{SX}_{2016}^j], Y_{2017}^j) | j \in \mathbf{J}_{2016}\}$ and $D_{2017} = \{(\mathbf{SX}_{2017}^j], Y_{2018}^j) | j \in \mathbf{J}_{2017}\}$.

A.5 Computation

A.5.1 Software

The development of the neural networks and processes to train and test them were done using the PyTorch library Version 1.9.0. An Anaconda virtual environment was used to manage package dependencies. The environment to perform all tasks relevant to this project (*forecast*) is saved on the project GitHub repository as a *.yml* file.

A.5.2 Computing resources / hardware

A High Performance Computing cluster was used to train, test and forecast. Typically, tasks ran with four Tesla P100-PCIE-16GB GPUs and twelve 5980MB CPUs up to twelve hours although this could vary depending on resource availability and task requirements. The operating system was Linux-3.10.0-1160.66.1.el7.x86_64-x86_64-with-glibc2.10 and Python version was 3.8.2.

A.6 Initial trails: experimental set up

A.6.1 Model 1: 2D CNN model

To tune our first model we selected a base set of hyperparameters and performed a grid search.

Our base model was trained on the full set of data available:

$$\mathbf{D}_{2014-2016} = D_{2014} \cup D_{2015} \cup D_{2016} =$$

$$\{(\mathbf{SX}^j_{2014}, Y^j_{2015}) | j \in \mathbf{J}_{2014}\} \cup \{(\mathbf{SX}^j_{2015}, Y^j_{2016}) | j \in \mathbf{J}_{2015}\} \cup \{(\mathbf{SX}^j_{2016}, Y^j_{2017}) | j \in \mathbf{J}_{2016}\}$$

However, to ensure equal bias towards labels in the year 2018 as with the rest of our models, we used only 60% of $\{(\mathbf{SX}^j_{2016}, Y^j_{2017}) | j \in \mathbf{J}_{2016}\}$ for training and allocated the remaining 20% for validation and 20% for test data. We iteratively cycled the training and validation data with k-fold cross-validation. When splitting the data, no undersampling was performed, and the split was done so that the original ratio of 0 to 1 labels was preserved in all three data subsets (99.688:0.312). Thus, 64,803,129, 21,601,044, and 21,601,046 pixels were allocated for the training, validation, and test datasets respectively, with the total size of the 2016 data being $|\mathbf{J}_{2016}| = 108,005,219$. For training and validation undersampling of our base model, we used $\theta = 1$, meaning the ratio of 0 to 1 labels was modified to 1:1. As a result, after undersampling, the sizes of the training and validation datasets were reduced to 404,978 and 134,994, respectively. Unless explicitly stated in the model experiment setup (case 12),

after each epoch, we changed the training dataset to one year older data, which was again undersampled to have a ratio $\theta = 1$. Therefore, at epoch 1, 60% of the undersampled 2016 data, $\tilde{D}^{60\%}_{2016}$, was used for training, and 20%, $\tilde{D}^{20\%}_{2016}$, was allocated for validation. At epoch 2, new undersampled data was added for training, namely $\tilde{D}_{2015}$, while the validation data was resampled with new 2016 zero values via our sampling algorithm with θ again set to 1. For epoch 3, the same procedure was performed with the new training data being $\tilde{D}_{2014}$. If the model was allowed to continue iterating to epochs 4 and 5, the training data would be newly undersampled $\tilde{D}^{20\%}_{2016}$ and $\tilde{D}_{2015}$.

The default Early Stopping Criterion was AUC, the weighting time 3 and the maximum number of epoch iterations 5. Weighted BCE with positive weights 5 was used as loss to be optimize. We used Adam optimizer (Kingma and Ba, 2014) with default learning rate of 0.0001. The batch size of all training experiments in this report was set to 80.

The default model architecture was as follows:

Input tensor $\mathbf{SX}^j_t \in \mathbb{R}^{7\times35\times35} \rightarrow 4\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow$ SPP(2, [13, 5]) $\rightarrow$ FC(1552,100) $\rightarrow$ ReLU $\rightarrow$ 1DBN $\rightarrow$ DO(0.2) $\rightarrow$ FC(100,1) $\rightarrow \sigma \rightarrow \hat{Y}^j_{t+1}$

That is, the default input spatial dimensions was (35×35) and it was propagated to 4 2D conv. layers that decreased its spatial size form 35 to 25. Conv. layer 1 had 8 filters each with spatial size (5×5), conv. layers 2 and 3 had 16 filters each with spatial size (3×3),and finally conv. layer 4 had 8 filters each with spatial size (3×3). All slided at stride 1 with no padding applied. Then a SPP layer downsampled it to a 1d vector of size 1552 via 2 max pooling filters returning 3D tensors of spatial size 13×13 and 5×5 respectively. This was fed to the two fully conceded layers which had dropot ratio 0.2. We then let examined how the network will perform if the the following hyperparameters were changed: shown: dropout ratio: 0.5, dropout ratio: 0, $\theta : |Y = 0| : |Y = 1|$ = 1:2 = 0.5, Input spatial size: 45×45, SPP with n = 2 and $\mathbf{k} = [15, 5]$, Number of filters in conv. layers 1 to 4: 16, Early Stopping criterion: BCE loss, BCE loss as loss to be optimized, Weighted BCE loss with positive weight 2 as loss to be optimized, Learning rate of Adam: 0.001, use only year D_{2016}. Results of these experiments are shown in Table A.3.

A.6.2 Model 2: 3D CNN model

For the other three models the full time series of data was utilised:

$$\{(S^j, X^j_{2014}, X^j_{2015}, X^j_{2016}) | j \in \mathbf{J}_{2016}\}$$

It was split to train, validation and test data sets with ratio 6:2:2 in such a manner that the original ratio of 0 to 1 labels was preserved (99.69 : 0.31). Thus 64803129, 21601044 and

21601046 pixels were allocated for the train, validation and test data sets respectively with the total size of the 2016 data being $|\mathbf{J}_{2016}|$ = 108005219. For train and validation undersampling of our base model we used $\theta = 1$, or said differently, the ratio of 0 to 1 labels was modified to 1:1. Thus, after undersampling, the training and validation data sizes were decreased to 404978 and 134994 respectively.

Here again for default training set up we used weighted BCE loss with positive weight 3, ADAM optimizer with learning rate 0.0001 and used AUC as an early stopping criterion.

The model architecture was as follows:
Branch 1 : $\mathbf{X}^j \in \mathbb{R}^{5\times3\times35\times35} \rightarrow 2\times$(3DConv$\rightarrow$ReLU$\rightarrow$3DBN)$\rightarrow \mathbf{Z}_x^j \in \mathbb{R}^{c_1,h,w}$
Branch 2: $\mathbf{S}^j \in \mathbb{R}^{2\times35\times35} \rightarrow 2\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN)$\rightarrow \mathbf{Z}_s^j \in \mathbb{R}^{c_2,h,w}$
Branch 3: $\mathbf{Z}^j \in \mathbb{R}^{(c_1+c_2),h,w} \rightarrow 2\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow$ SPP$(1,15) \rightarrow$ FC(spp,100) $\rightarrow$ ReLU $\rightarrow$ 1DBN $\rightarrow$ DO(0.3) $\rightarrow$ FC(100,1) $\rightarrow \sigma \rightarrow \hat{Y}_{t+1}^j$

That is, each of the 2 3D conv. layers of the branch taking $\mathbf{X}^j \in \mathbb{R}^{5\times3\times35\times35}$ have 16 4D filters of dimension $\mathbb{R}^{5,2,3,3}$ which slide at stride 1 and no padding was applied. Each of the 2 2D conv. layers of the branch taking $\mathbf{S}^j \in \mathbb{R}^{2\times35\times35}$ have 8 3D filters of of dimension $\mathbb{R}^{2,3,3}$ which slide at stride 1 and no padding is applied. Finally, the branch taking $\mathbf{Z}^j \in \mathbb{R}^{32,31,31}$ has 4 3D filters of dimension $\mathbb{R}^{32,3,3}$ which slide at stride 1 in each of its 2 2D conv. layers. No padding is applied. The SPP layer has one max pooling filter that returns a 3D tensor of dimension $\mathbb{R}^{4,13,13}$ which is flattened and passed to the first fully connected layer. A dropout ratio before the final FC layer of rate 0.3 is applied.

We tested this base set up against: dropout = 0.5, larger spatial sizes of the filters in the first two branches - filters of branch 1 and branch 2 $\in \mathbb{R}^{5,2,5,5}$ and $\in \mathbb{R}^{2,5,5}$ respectively, increased number of filters in each of the three branches - 8, 32, 8 respectively, SPP with n=2 and $\mathbf{k} = [12,5]$, Cost of Confusion Matrix as an early stopping criterion, input image size 31 and 45, $\theta = 0.5$, weighted BCE loss with weight 10, L2 regularization with parameter 0.6 and finally, learning rate of ADAM 0.001. Table A.4 summarize the results of this experiment.

A.6.3 Model 3 and 4: ConvLSTM RNN model & Deep ConvLSTM RNN model

For our last two models, Model 3: ConvLSTM RNN and Model 4: Deep ConvLSTM RNN, the training, validation and test data set up was identical as this of Model 2. Base Model 3 differed form base Model 4 only in that Model 4 has two identical ConvLSTM cells one after another, which defined it as “deep".

The base set up was weighted BCE loss with positive weight w = 5, AUC as an early stopping criterion, Adam optimizer with learning rate 0.0001, $\theta = 1$.

The base models architecture is as follows:

Encoder:

$\{\mathbf{X}_k^j\}_{k=t-3}^t \in \mathbb{R}^{5\times25\times25} \rightarrow 3\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow \{\tilde{\mathbf{X}}_k^j\}_{k=t-3}^t \in \mathbb{R}^{c_1\times h_1\times w_1}$

Conv LSTM RNN: $\{\tilde{\mathbf{X}}_k^j\}_{k=t-3}^t \rightarrow 2\times$(ConvLSTM) $\rightarrow \mathbf{Z}_x^j \in \mathbb{R}^{c_2,h_2,w_2}$

2D CNN Static:

$\mathbf{S}^j \in \mathbb{R}^{2\times25\times25} \rightarrow 3\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN)$\rightarrow \mathbf{Z}_s^j \in \mathbb{R}^{c_3,h_2,w_2}$

2D CNN Joint:

$\mathbf{Z}^j \in \mathbb{R}^{(c_2+c_3),h_2,w_2} \rightarrow 3\times$(2DConv$\rightarrow$ReLU$\rightarrow$2DBN) $\rightarrow$ SPP$(1,12) \rightarrow$ FC(spp,100) $\rightarrow$ ReLU $\rightarrow$ 1DBN $\rightarrow$ DO(0.3) $\rightarrow$ FC(100,1) $\rightarrow \sigma \rightarrow \hat{Y}_{t+1}^j$

That is, the input spatial size is 25x25. The **Encoder** branch have 3 2D conv. layers with 8 filters each, of dimension $\mathbf{R}^{5,3,3}$ that slides at stride 1 and no padding is applied.

The **Conv LSTM RNN** than take $\{\tilde{\mathbf{X}}_k^j\}_{k=t-3}^t \in \mathbb{R}^{8\times19\times19}$ pass it to $2\times$(ConvLSTM) ($1\times$(ConvLSTM) for Model 3) and return $\mathbf{Z}_x^j \in \mathbb{R}^{8,19,19}$. Both ConvLSTM cells have 8 filters for each of the gates, with each filter $\in \mathbb{R}^{8,3,3}$. However, a padding of size 1 is applied in order for the cell to be recurrent. Each filter slides at stride 1.

Both **2D CNN Joint** and **2D CNN Static** have 8 filters in each of their 3 2D conv. layers, each of spatial dimension 3x3, sliding at stride 1 and no padding is applied. The SSP layer has one maxpooling filter that returns 3D tensor $\in \mathbb{R}^{8,12,12}$. Finally, dropout of rate 0.3 is applied between the two FC layers.

We compared these models, against this that has the following changes: dropout = 0.5, for Model 3 the number of filters in each of the for branches was changed from (8,8,8,8) to (8,8,4,4), (16,8,4,4), (16,16,4,4), and for Model 4 from (8,(8,8),8,8) to (16,(8,8),8,8), (16,(16,16),8,8), (8,(8,8),4,4), Where the first input of the notation (.,.,.,.) is the number of filters in the **Encoder**, the second for $2\times$(ConvLSTM) / $1\times$(ConvLSTM) for Model 3 and so on. We also compared models when for early stopping criterion the confusion matrix was evaluated, when the input have size 31, 35, wen $\theta = 0.5$, L2 regularization with parameter 0.2 and finally, ADAM with learning rate 0.001. The results of this experiments are displayed in Table A.5 and Table A.6 for Model 3 and 4 respectively.

A.7 Experimental Results

A.7.1 Broad model intercomparison (Madre de Dios)

Model performance within training period

Table A.3 Model 1, Experimental results

hyperparameter	epochs	Total time	AUC on Train	AUC on Valid.	AUC on Test
base	3	3.2h	0.924	0.921	0.921
dropout ratio: 0.5	5	6h	0.935	0.918	0.917
dropout ratio: 0	1	2.5h	0.911	0.922	0.921
θ : 0.5	4	1.7h	0.930	0.934	0.933
In size: 45×45	4	5h	0.943	0.950	**0.951**
SPP: n = 2,$\mathbf{k} = [15,5]$	4	4h	0.936	0.918	0.919
hidden dim = 16	4	4.4h	0.940	0.942	0.942
WBCEL as ES Crit.	1	4h	0.912	0.878	0.877
BCE loss	5	5h	0.944	0.908	0.907
WBCE loss with w = 2	5	5h	0.944	0.914	0.913
learining rate 0.001	4	5.25	0.939	0.930	0.931
use only year D_{2016}	2	2h	0.914	0.927	0.926

Table A.4 Model 2, Experimental results

hyperparameter	epochs	Total time	AUC on Train	AUC on Valid.	AUC on Test
base	3	6h	0.92862	0.94201	0.9428
dropout = 0.5	2	10h(max)	0.93471	0.94981	0.9510
filters sizes: (5,2,5,5),(2,5,5),(32,3,3)	2	12h (max)	0.93588	0.94741	0.9456
num of filters: 8,32,8	2	10h(max)	0.94540	0.95599	0.9563
SPP with n=2 and $\mathbf{k} = [12,5]$	4	10h(max)	0.94666	0.95731	**0.9569**
Cost as ES	3	10h(max)	0.94196	0.93736	0.9371
In size 31	5	10h(max)	0.94359	0.95100	0.9509
In size 45	3	10h(max)	0.94132	0.95189	0.9514
θ=0.5	4	5h	0.93257	0.94898	0.9479
wBCE loss, w= 10	3	10h(max)	0.93109	0.93944	0.9393
L2 reg. with 0.6	2	10h(max)	0.92067	0.92135	0.9203
ADAM lr = 0.001	4	10h(max)	0.94794	0.93456	0.9348

Table A.5 Model 3, Experimental results

hyperparameter	epochs	Total time	AUC on Train	AUC on Valid.	AUC on Test
base	3	6h	0.92862	0.94201	0.9428
dropout = 0.5	5	10h(max)	0.93047	0.94497	0.9457
hidden dim=(8,8,4,4)	5	10h(max)	0.92581	0.94429	0.9429
hidden dim=(16,8,4,4)	3	10h(max)	0.92859	0.94707	0.9474
hidden dim=(16,16,4,4)	3	8h	0.92962	0.94573	0.9464
Cost Matrix as ES	5	5h	0.93768	0.93787	0.9378
size = 31	4	10h(max)	0.93143	0.94165	0.9417
size = 35	4	9h	0.93248	0.94808	**0.9484**
$\theta = 0.5$	3	2h	0.90706	0.92173	0.9361
wBCE loss with w = 10	4	6h	0.91791	0.94352	0.9423
L2 regularization with param 0.6	2	10h(max)	0.92067	0.92135	0.9203
ADAM lr = 0.001	4	10h(max)	0.94794	0.93456	0.934

Table A.6 Model 4, Experimental results

hyperparameter	epochs	Total time	AUC on Train	AUC on Valid.	AUC on Test
base	4	5h	0.92831	0.94184	0.9420
dropout = 0.5	5	8h	0.93149	0.94147	0.9413
hidden dim= (16,(8,8),8,8)	2	10h(max)	0.92767	0.93447	0.9350
hidden dim=(16,(16,16),8,8)	2	10h(max)	0.92766	0.94244	0.9433
hidden dim= (8,(8,8),4,4)	5	5h	0.92678	0.94099	0.9408
Cost Matrix as ES	5	5h	0.93565	0.94194	0.9417
$\theta = 0.5$	3	3h	0.91713	0.93774	0.9363
wBCE loss with w = 10	3	5h	0.91369	0.92903	0.9295
L2 regularization with param 0.6	1	10h(max)	0.88856	0.91088	0.9102
ADAM lr = 0.001	5	7	0.94709	0.95324	**0.9540**

A.7.2 Testing on 2018

Our four best models of each class are compared on labels corresponding to year 2018, $\hat{Y}^j_{2018}, j \in \mathbf{J}_{2017}$. The input data for our last three models, Model 2, Model 3, Model 4, is $\{(S^j, X^j_{2015}, X^j_{2016}, X^j_{2017}) | j \in \mathbf{J}_{2017}\}$, whereas Model 1 takes as an input $\{(S^j, X^j_{2017}) | j \in \mathbf{J}_{2017}\}$.

Scenario 1: Ratio of forested to deforested pixels in 2018 (0:1 labels) is 1:1 and our model predicts the top 50% of the data it receives as deforested. The data size after undersampling with ratio 1:1 is 572978.

Scenario 2: Ratio of forested to deforested pixels in 2018 (0:1 labels) is 1:4 and our model predicts the top 20% of the data it receives as deforested. The data size after undersampling with ratio 4:1 1432445

Scenario 3: Ratio of forested to deforested pixels in 2018 (0:1 labels) is 9:1 and our model predicts the top 20% of the data it receives as deforested. The data size after undersampling with ratio 9:1 2864890

We ran this test on the three best models, form class Model 1: 2D CNN, Model 2: 3D CNN, and Model 4: Deep ConvLSTM RNN model. Fig. A.16, Fig. A.17 and Fig. A.18 summarize the our final results. Overall, all model preserve their accuracy when tested on the subsequent year and achieve similar accuracy. Nevertheless, Model 2: 3D CNN achives highest score on Scenario 3, which approximate the actual imbalance of forested to deforested pixels the most.

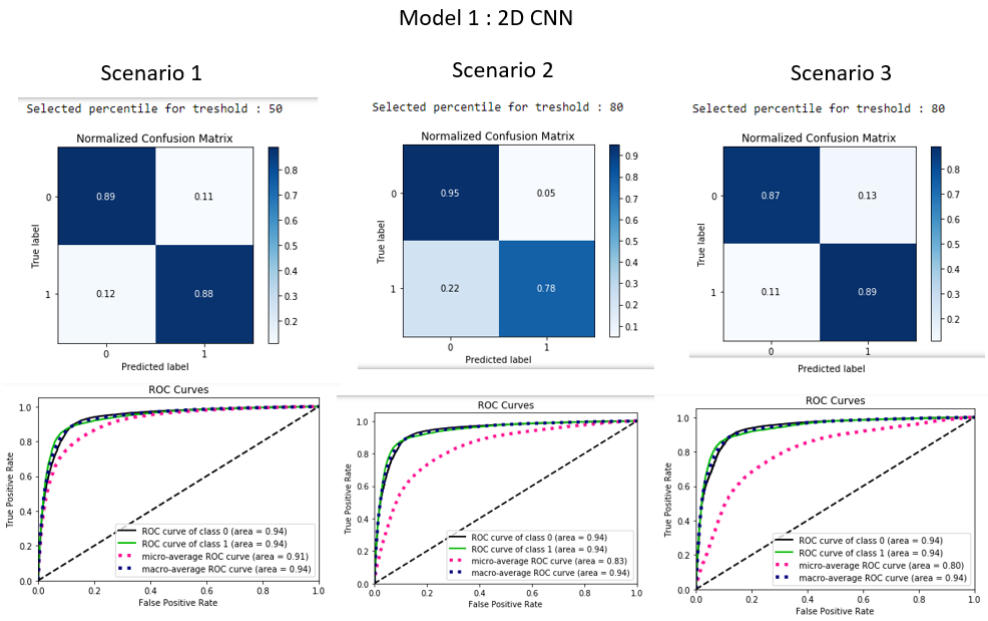


Fig. A.16 Model 1: 2D CNN

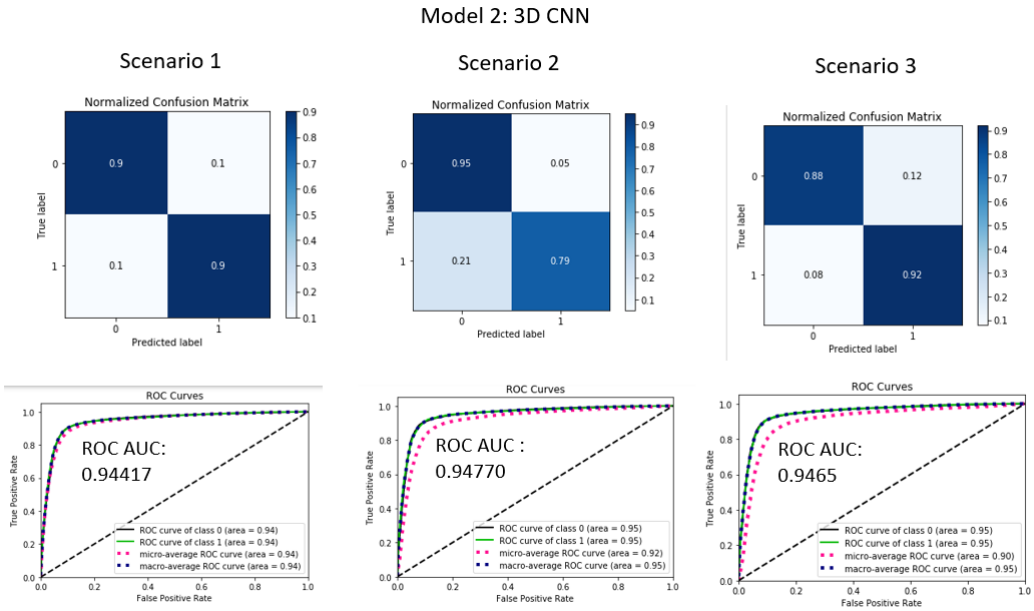


Fig. A.17 Model 2: 3D CNN

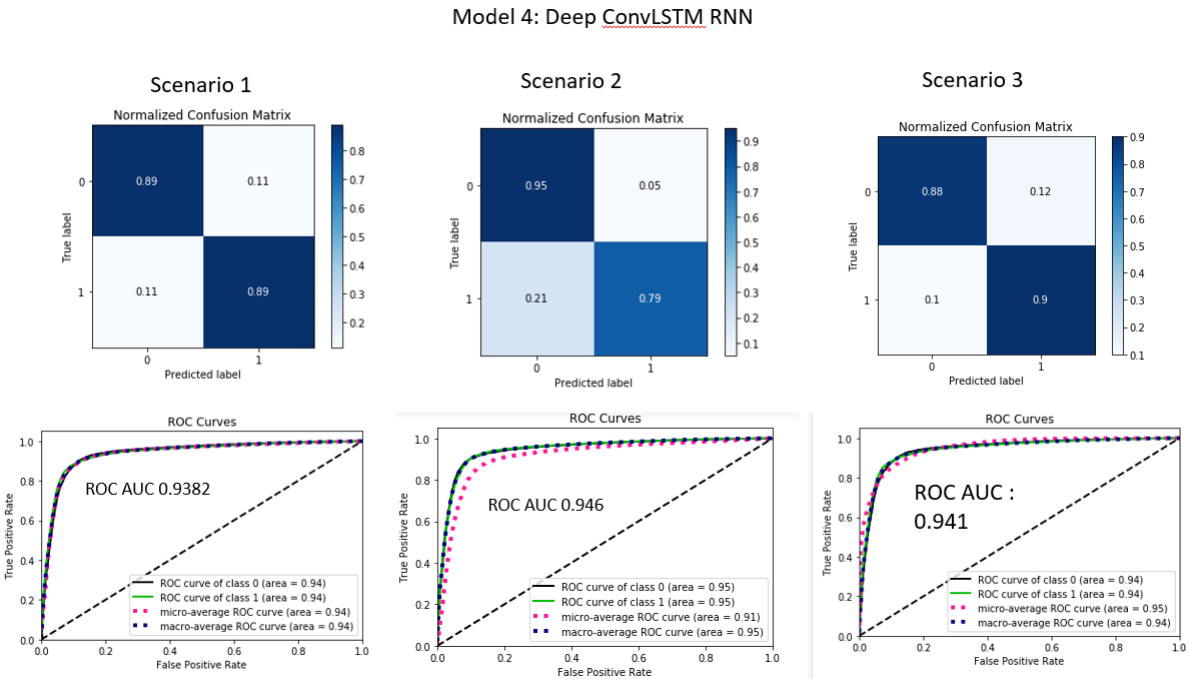


Fig. A.18 Model 4: Deep ConvLSTM RNN

Notes on initial experimentation

The spatial size of input matters in 2D CNNs. For almost all models increasing the number of filters in their layer is beneficial. Setting weighted BCE loss with high weight penalty to wrongly predicted deforested labels, or sampling the model in favour for the deforested pixels ($\theta = 0.5$ - ratio 0:1 labels) does not show any noticeable difference in accuracy. The best performing model is Model 2: 3D CNN, followed by Model 4 and 3. Model 1 is ranked as last.

For our first Model 1, 2D CNN, that the size of the input image is crucial. Secondly, increasing the networks' depth also resulted in better performance on test data. The way we employed dropout did not have noticeable influence, moreover, when it was completely switched off, the accuracy on test and validation data was higher, and the opposite trend was observed when dropout was set to 0.5. Therefore, we can be confident that the way we utilize it is efficiently. Finally, training with smaller learning rate results in a comparable accuracy for the benefit of the faster learning process. We believe this is possible due to the Bach Normalization Layer after each activation.

Our second Model, 3D CNN, noticeably outperformed Model 1 in all experimental scenarios and we confidently conclude that the utilization of the time domain is of huge importance. For this model, the size of the input image is not noted to make huge difference when set to 31, 35 or 45. However, increasing the number of filters in the 3D CNN leads to better accuracy. Furthermore, if the downsampling rate is at a lower rate, performance improves. The wast two facts give us the intuiting that the model has much more information to learn and a deeper network may even result in a higher accuracy.

The third model, ConvLSTM RNN performs almost as well as model 2. All figures about AUC ROC on test data are very similar and clear conclusions can not be made. The only clear conclusion we can make is that including weight decay, also known as L2 penalty resulted in the worst performing model in each model class, although applied with the small parameter 0.6.

Our best model of class Model 4, Deep ConvLSTM model, resulted by a training process when the learinig rate was set to 0.001, and is the second best model overall. The inclusion of additional ConvLSTM cell did not show any significant improvement.

Overall, we conclude that our second proposed model, 3D CNN, achieves the best AUC measurement on the 2017 test data set. It is also simpler then our Model 3 and 4. Furthermore, due to its overall highest performance in all test scenarios, we believe the convergence is not by "chance".

A.7.3 Developed models: Madre de Dios and Junin region

2D CNN results

Training and testing plots and accuracy statistics are available on WandB [1]

3D CNN results

Training and testing plots and accuracy statistics are available on WandB [2]

[1] https://wandb.ai/patball/forecasting2D
[2] https://wandb.ai/patball/forecasting

Appendix B

Chapter 3 Supplementary Materials

B.1 Study sites and remote sensing data collection

Orthomosaics were generated from the raw aerial photographs in AgiSoft Metashape. This software uses structure from motion to generate a 3D elevation surface from the raw aerial photographs (Westoby et al., 2012) which is then projected back into 2D to create a landscape scale image (orthomosaic). This gave georeferenced photographs, largely corrected for distortions. Some distortion remained at the edges of the mosaic, so the analysis focused on the core areas where the imagery was not distorted.

B.1.1 Sepilok (East & West) and Danum

Sepilok Forest Reserve (5° 50' N 117° 56 ' E) is a region of lowland tropical rainforest in Sabah, a state in Malaysia. The Reserve is one of the oldest protected areas of forest in Asia, founded by the Sabah Forest Department in 1931. The Reserve spans nearly 4500 ha, with ground elevation varying between 50 and 250 metres above sea level. Three distinct forest types are present: alluvial dipterocarp, sandstone dipterocarp and heath forest. Danum Valley contains lowland tropical rain forests dominated by dipterocarps and are among the tallest forests on the planet (Shenkin et al., 2019). The three Malaysian sites experience a similar climate with approximately 2300 mm rainfall per year with the wettest months between November and February (Nilus et al., 2011).

The 2014 lidar data of Sepilok and Danum were collected using a Leica ALS50-II ALS flown at an altitude of nearly 2000 metres, attached to the belly of a Dornier 228-201. The ALS sensor emitted pulses at around 80 Hz with a field of view of 14.0°, and a footprint of about 40 cm diameter. The average pulse density was 11 pulses m^{-2}. The ALS data were preprocessed by NERC's Data Analysis Node and delivered in LAS format.

The 2020 lidar data in Sepilok was collected using a RIEGL LMS-Q560 mounted on a helicopter flying at 200 metres altitude at a ground speed of approximately 100 km/hr. This resulted in an average pulse density of 38 m^{-2}. Further processing used LAStools (http://rapidlasso.com/lastools/). Points were randomly resampled to match the point density across dates. Points were split into two groups, ground and non-ground, and a digital elevation model (DEM) was fitted to the ground points, producing a raster of 1 m resolution. The DEM elevations were subtracted from elevations of all non-ground returns, known as the digital surface model (DSM) to create a canopy height model (CHM) raster of resolution 1 m.

The 2014 RGB imagery was collected using a Leica RCD105 Digital Camera, attached to a plane flown at an altitude of 796 metres, with a ground resolution of 10 cm. Photographs have been orthorectified and collated into homogenous mosaics using Agisoft Metashape. This software used the Structure from Motion method to calculate the elevation of the observed surface, of which the photographs are mapped to, This results in georeferenced photographs that are corrected for distortions. The files were delivered in the GeoTIFF format.

B.1.2 Paracou

The Paracou field station is situated in a lowland tropical forest near Sinnamary, French Guiana (5°16'N 52 °55'W). It is a forest similar to Sepilok Forest Reserve, which is also a lowland tropical forest.

ALS data were acquired in September 2016, and November 2019 by ALTOA, operating a RIEGL LMS-Q780 sensor attached to an aircraft flying at 800m. On all dates, the scan frequency was 400 kHz and the final point density was above 50 m^{-2}. The creation of a CHM was carried out using the same process as used for Sepilok.

On the same flights as the ALS scans, RGB images were collected using an IXA180 Phase One camera with an 8 cm ground sampling distance.

Hyperspectral imagery was also collected over Paracou, with a Hyspex VNIR-1600 (Hyspex NEO, Skedsmokorset, Norway) sensor-mounted alongside the RIEGL scanner. Its bands covered a spectral range of 414–994 nm with a spectral sampling distance of 3.64 nm. Images were orthorectified and georeferenced to 1 m spatial resolution with the PARGE software using the DSM from the lidar data.

Fig. B.1 An original RGB image is displayed on the left, while a contrasted enhanced (stretched) RGB image is displayed on the right of this figure. Stretching the colours of the image to the values allows for easier identification of individual trees when carrying out manual tree crown delineations. The effect is particularly noticeable in the lower left corner of the images.

B.2 Tree crown data

B.2.1 Manual delineation

Manual tree crown labelling was carried out at each study site, to create training and test sets for our analysis. Labelling was carried out in open-source graphical geospatial software, QGIS, using a combination of RGB, lidar, and, in the case of Paracou, hyperspectral imagery. To increase the contrast between individual trees the colours of the RGB imagery were stretched. Polygons were carefully drawn around the perimeter of all distinguishable trees for a number of areas in each region. JGCB carried out the manual labelling for Paracou. A visualisation of the difference between the original and the stretched image is given in Fig. B.1.

B.2.2 Training and validation data

Examples of the training data are given in Fig. B.4.

B.3 Data preparation and processing

To apply Detectree2 to remote sensing RGB images of tropical forests, the geospatial raster image must be tiled and converted into a png format, and the manual crown segmentations used for training must also be converted from geospatial shapefiles into JSON files that align with

the pngs. These training tiles are ingested into Detectree2, which then learns how to identify a tree, and then can make predictions on new tiles.

Training tiles were separated into 5-folds so that cross validation could be performed during the hyperparameter tuning phase. Testing tiles were spatially separated from the training tiles and only seen by the network at the evaluation phase.

B.4 Model architecture, tuning and training

B.4.1 Model architecture

Mask R-CNN (He et al., 2017) is a framework for instance segmentation. It builds on Faster R-CNN (Ren et al., 2015), which carries out object detection using a Region Proposal Network (RPN). An RPN is a fully convolutional network, trained end-to-end, that generates Regions of Interest (RoIs) for each image. These RoIs have object bounds and objectness scores attached to them, giving the bounds of the RoI, and the likelihood of the RoI containing an object. These RoIs are then passed through fully connected convolutional networks to determine the class of image contained, and the exact mask of each object, within the respective bounding box. For full details on the structure of Mask R-CNN, please refer to the GitHub repository for this work (https://github.com/PatBall1/Detectree2). The schematic of the architecture of the model architecture is given in Fig. B.2.

We selected the R101-FPN configuration - this architecture combines a 101 layer deep ResNet (He et al., 2016) module with a Feature Pyramid Network module. The configuration sets the *backbone* of the network which is the part that views and extracts features from the scene as a whole. The R101-FPN backbone consists of a 101 layer deep ResNet (He et al., 2016) module with a Feature Pyramid Network (Lin et al., 2017) module. The initial model weights were generated from pre-training of the network on the ImageNet dataset[1]. It is possible to "freeze" the backbone to different depths depending on the amount of flexibility the user wants to introduce in moving away from the pre-trained model weights. An example of the different predictions is given in Fig. B.3.

B.4.2 Data augmentation

The training data was augmented by submitting the training data to a variety of transformations including vertical and horizontal flips, rotation, and varying the saturation and contrast of the

[1] https://www.image-net.org/

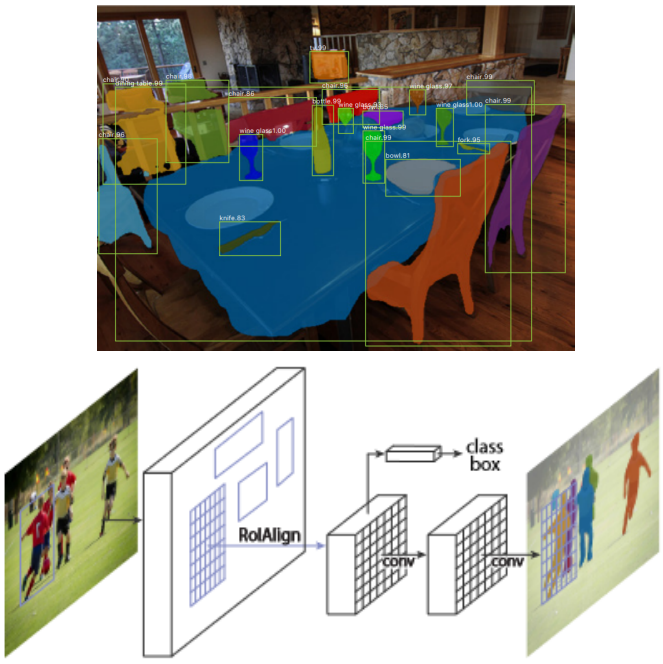


Fig. B.2 Illustration of the Mask R-CNN predictions and architecture from He et al. (2017).

images. These augmentations are designed to increase the variety of training data seen by the model, to allow it to generalise more readily to new images.

B.4.3 Training and hyperparameter tuning

We selected the *AP50* of the segmentation predictions on the (randomly selected) validation fold as our metric for optimisation. AP50 is the average precision of predictions when a correct match is granted for $IoU > 0.5$.

AP50 was also used as the metric for *early stopping* whereby if the model performance failed to improve for a set number of training iterations (the "patience"), training would be terminated and the best model up to that point would be saved. This is a technique for regularisation and prevents wasted computation. As with all deep networks, there were several of hyperparameters that controlled the way the networks handled data handling and were trained. The first of these relates to the architecture of the algorithm, namely the number of hidden layers in the ResNet backbone (He et al., 2016) of the network. ResNets are used as the backbone of Mask R-CNN as very deep neural networks are particularly difficult to train, due to the vanishing gradients problem (Glorot and Bengio, 2010). ResNets use skip connections, whereby certain layers in the network are skipped during backpropagation to avoid this problem.

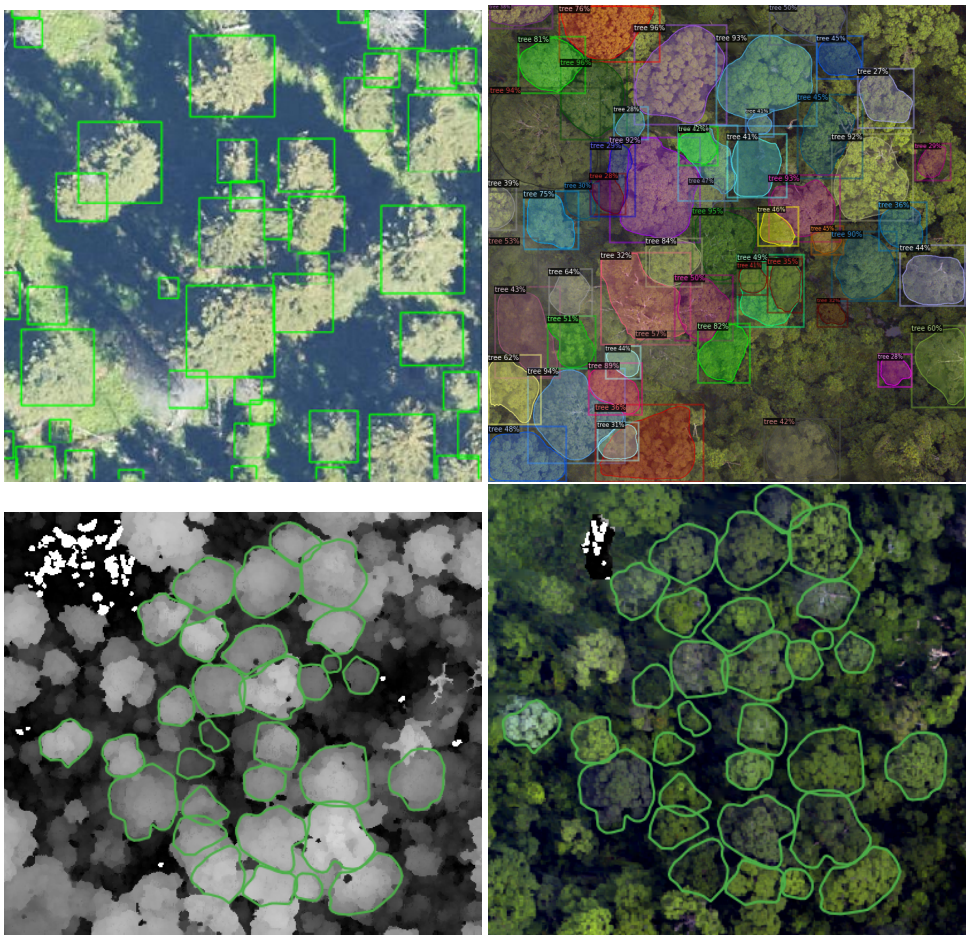


Fig. B.3 (A) Crown bounding boxes predicted by DeepForest (Weinstein et al., 2019), and (B) crowns predicted by Detectree2. The colours in plot B merely distinguish predicted trees. A comparison of manually delineated crowns, overlaid on lidar (C) and RGB (D).

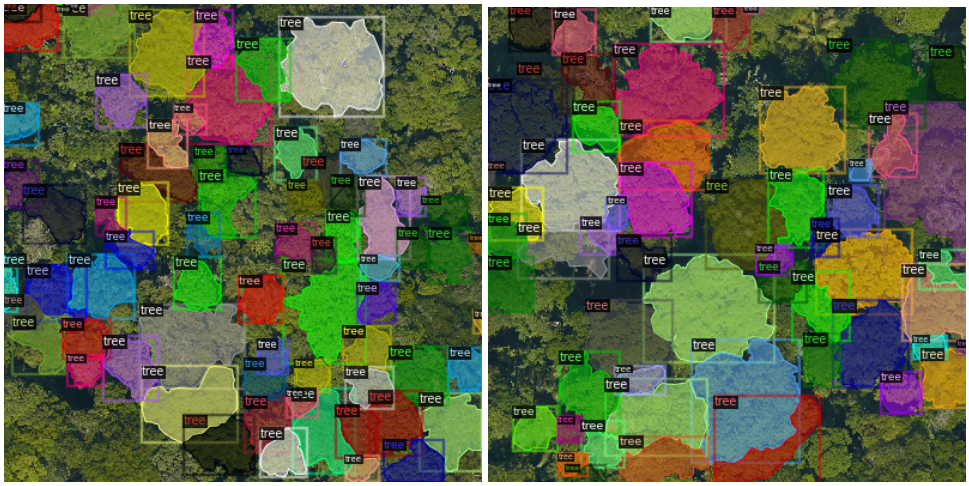


Fig. B.4 Examples of training data provided to Mask R-CNN. The different colours help to distinguish between trees.

Table B.1 Tunable hyperparameters (with their optimised value) and a description of their purpose.

Hyperparameter (=val)	Description
dl_nums_workers (=2)	To speed up the training process, we make use of the num_workers optional attribute of the DataLoader class. The num_workers attribute tells the data loader instance how many sub-processes to use for data loading. By default, the num_workers value is set to zero, and a value of zero tells the loader to load the data inside the main process. This means that the training process will work sequentially inside the main process. After a batch is used during the training process and another one is needed, we read the batch data from the disk. Now, if we have a worker process, we can make use of the fact that our machine has multiple cores. This means that the next batch can already be loaded and ready to go by the time the main process is ready for another batch. This is where the speed up comes from. The batches are loaded using additional worker processes and are queued up in memory.
ims_per_batch (=2)	If we use 16 GPUs and IMS_PER_BATCH = 32, each GPU will see 2 images per batch.
gamma (=0.1)	The iteration number to decrease the learning rate by GAMMA.
backbone_freeze_at (=3)	Freeze the first several stages so they are not trained. There are 5 stages in ResNet. The first is a convolution, and the following stages are each group of residual blocks. Freezing a layer prevents its weights from being modified. This technique is often used in transfer learning, where the base model(trained on some other dataset is frozen.
warmup_iters (=120)	Warm-up steps are just a few updates with low learning rate before training. After this warm-up, the regular learning rate is use, (schedule) to train the model to convergence. The idea is that this helps the network to slowly adapt to the data intuitively. If the data set is highly differentiated, it can suffer from a sort of "early over-fitting". If your shuffled data happens to include a cluster of related, strongly-featured observations, your model's initial training can skew badly toward those features - or worse, toward incidental features that are not truly related to the topic at all. Warm-up is a way to reduce the primacy effect of the early training examples. Without it, one may need to run a few extra epochs to get the convergence desired, as the model un-trains those early superstitions. Many models afford this as a command-line option. The learning rate is increased linearly over the warm-up period. If the target learning rate is p and the warm-up period is n, then the first batch iteration uses $1p/n$ for its learning rate; the second uses $2p/n$, and so on: iteration i uses $i*p/n$, until we hit the nominal rate at iteration n. This means that the first iteration gets only $1/n$ of the primacy effect. This does a reasonable job of balancing that influence. Note that the ramp-up is commonly on the order of one epoch – but is occasionally longer for particularly skewed data, or shorter for more homogeneous distributions. One may want to adjust, depending on how functionally extreme ones batches can become when the shuffling algorithm is applied to the training set.
momentum (=0.9)	Momentum in neural networks is a variant of the stochastic gradient descent. It replaces the gradient with a momentum which is an aggregate of gradients. Momentum can increase speed when the cost surface is highly non-spherical because it damps the size of the steps along with directions of high curvature thus yielding a larger effective learning rate along with the directions of low curvature.
batch_size_per_image (=1024)	The batch size defines the number of samples that will be propagated through the network. For instance, let's say you have 1050 training samples and you want to set up a batch_size equal to 100. The algorithm takes the first 100 samples (from 1st to 100th) from the training dataset and trains the network. Next, it takes the second 100 samples (from 101st to 200th) and trains the network again. We can keep doing this procedure until we have propagated all samples through of the network. Problem might happen with the last set of samples. In our example, we have used 1050 which is not divisible by 100 without remainder. The simplest solution is just to get the final 50 samples and train the network. Advantages of using a batch size < number of all samples: It requires less memory. Since you train the network using fewer samples, the overall training procedure requires less memory. That is especially important if you are not able to fit the whole dataset in your machine's memory. Typically networks train faster with mini-batches. That is because we update the weights after each propagation. In our example we have propagated 11 batches (10 of them had 100 samples and 1 had 50 samples) and after each of them we have updated our network's parameters. If we used all samples during propagation we would make only 1 update for the network's parameter. Disadvantages of using a batch size < number of all samples: The smaller the batch the less accurate the estimate of the gradient will be.
weight_decay (=0.001)	Weight decay is a regularization technique by adding a small penalty, usually the L2 norm of the weights (all the weights of the model), to the loss function. loss = loss + weight decay parameter * L2 norm of the weights. Why do we use weight decay? To prevent overfitting. To keep the weights small and avoid exploding gradient. Because the L2 norm of the weights are added to the loss, each iteration of your network will try to optimize/minimize the model weights in addition to the loss.
base_lr (=0.0003389)	In machine learning and statistics, the learning rate is a tuning parameter in an optimization algorithm that determines the step size at each iteration while moving toward a minimum of a loss function. Since it influences to what extent newly acquired information overrides old information, it metaphorically represents the speed at which a machine learning model "learns". In the adaptive control literature, the learning rate is commonly referred to as gain. In setting a learning rate, there is a trade-off between the rate of convergence and overshooting. While the descent direction is usually determined from the gradient of the loss function, the learning rate determines how big a step is taken in that direction. A too high learning rate will make the learning jump over minima but a too low learning rate will either take too long to converge or get stuck in an undesirable local minimum. In order to achieve faster convergence, prevent oscillations and getting stuck in undesirable local minima the learning rate is often varied during training either in accordance with a learning rate schedule or by using an adaptive learning rate.
max_iter (=462)	An iteration describes the number of times a batch of data passed through the algorithm. In the case of neural networks, that means the forward pass and backward pass. So, every time you pass a batch of data through the neural network, you completed an iteration. When you increase the number of iterations, you get closer to the minima and the set of optimal parameters, and these are what improve performance.

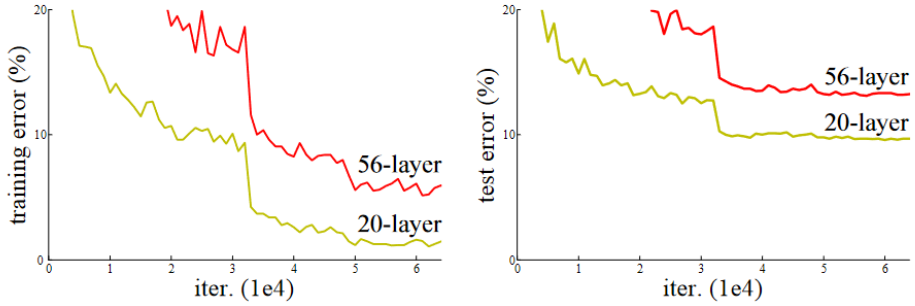


Fig. B.5 An example comparison of the training of deep and shallow networks. These plots are taken from He et al. (2016) and they illustrate that deeper neural networks do not necessarily learn as well as shallower neural networks.

Other hyperparameters that can be optimised in Mask R-CNN relate to the training of the network. Namely the learning rate, the number of iterations and the batch size. The optimisation of these hyperparameters was carried out using Weights and Biases, which uses grid search to determine the optimal hyperparameters. The hyperparameters selected are given in Table S4.1.

We found that while increasing the depth of the ResNet increased the training time of the algorithm, it improved accuracy scores. The batch size was limited by computing resources available. The number of iterations was the key metric to optimise, as we found that our model began to overfit the training data if trained for too many iterations. To determine the number of iterations, we plotted a graph (Fig. B.6) of both the total training loss and the total validation loss to determine the minimum of the total validation loss, and hence the optimal number of iterations to train the model. As Mask R-CNN consists of three predictions, namely the class, bounding box and mask, the total loss is the sum of these three losses.

We see that after around 800 iterations, the model started to overfit the training data, at the expense of performance of the validation dataset. As such we determined to stop the training of Mask R-CNN at 800 iterations, and then used the saved weights after 800 iterations to predict tree crowns in Sepilok.

B.5 Evaluation metrics

Two key metrics are used in this work to evaluate the performance of the model: the AP50 score, and the F_1 Score. AP50 relates to the area under the precision-recall curve (AUC-PR) evaluated at a particular threshold for the Intersection over Union, in our case 0.5. A high value for AP50 means that we are seeing both high precision and recall of the model at our particular IoU cutoff. The IoU cutoff was selected as 0.5, on the basis of previous studies evaluating the accuracy of individual tree delineation methods (e.g. Aubry-Kientz et al., 2019; Hao et al.,

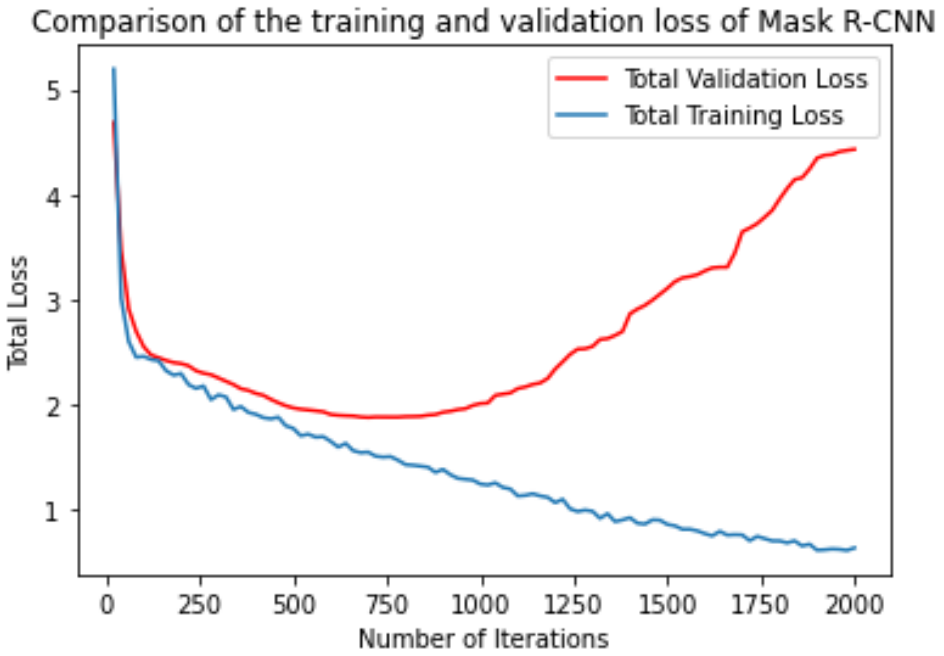


Fig. B.6 The total training and validation loss of Mask R-CNN as the model trained. Both total training and validation loss were calculated every 20 iterations.

2021). In rare cases where more than one automatic delineation had an IoU of greater than 0.5 with a manual crown, then the automatic delineation with the greatest IoU was considered a true positive and the others were classed as false positives.

B.6 Model accuracies and parameters

Full model parameters and accuracies are given in Table B.2.

B.7 Maps of predictions

Maped predictions are given in Fig. B.7 to Fig. B.10.

Table B.2 Model accuracies and parameters across sites

Train1	Train2	Model name	Evaluation	Conf_threshold	Total_tests	Area_thresh	Border_filter	Precision	Recall	F1
Paracou2016		221026_Paracou2016/model_32.pth	Paracou2016	0.3	381	16	FALSE	0.529	0.612	0.568
Danum		221026_Danum/model_33.pth	Danum	0.7	278	16	FALSE	0.673	0.629	0.651
SepilokE		221026_SepilokE/model_23.pth	SepilokE	0.6	167	16	FALSE	0.531	0.563	0.547
SepilokW		221026_SepilokW/model_16.pth	SepilokW	0.5	704	16	FALSE	0.584	0.638	0.61
Combined		221026_Paracou2016DanumSepilokESepilokW/model_22.pth	Paracou2016	0.5	381	16	FALSE	0.569	0.507	0.536
Combined		221026_Paracou2016DanumSepilokESepilokW/model_22.pth	Danum	0.6	278	16	FALSE	0.657	0.709	0.682
Combined		221026_Paracou2016DanumSepilokESepilokW/model_22.pth	SepilokE	0.3	167	16	FALSE	0.537	0.647	0.587
Combined		221026_Paracou2016DanumSepilokESepilokW/model_22.pth	SepilokW	0.6	704	16	FALSE	0.647	0.616	0.631
Combined	Paracou2016	221027combined_Paracou2016/model_12.pt	Paracou2016	0.4	381	16	FALSE	0.612	0.538	**0.573**
Combined	Danum	221027combined_Danum/model_9.pth	Danum	0.7	278	16	FALSE	0.713	0.662	**0.687**
Combined	SepilokE	221027combined_SepilokE/model_10.pth	SepilokE	0.5	167	16	FALSE	0.612	0.653	**0.632**
Combined	SepilokW	221027combined_SepilokW/model_11.pth	SepilokW	0.4	704	16	FALSE	0.64	0.656	**0.648**
Paracou2016		221026_Paracou2016/model_32.pth	Danum	0.7	278	16	FALSE	0.613	0.536	0.572
Paracou2016		221026_Paracou2016/model_32.pth	Sepilok East	0.3	167	16	FALSE	0.62	0.508	0.559
Paracou2016		221026_Paracou2016/model_32.pth	Sepilok West	0.3	704	16	FALSE	0.539	0.49	0.513
Danum		221026_Danum/model_33.pth	Paracou	0.3	381	16	FALSE	0.465	0.318	0.378
Danum		221026_Danum/model_33.pth	SepilokE	0.3	167	16	FALSE	0.595	0.395	0.475
Danum		221026_Danum/model_33.pth	SepilokW	0.3	704	16	FALSE	0.55	0.582	0.566
SepilokE		221026_SepilokE/model_23.pth	Paracou	0.3	381	16	FALSE	0.518	0.42	0.464
SepilokE		221026_SepilokE/model_23.pth	Danum	0.7	278	16	FALSE	0.487	0.618	0.545
SepilokE		221026_SepilokE/model_23.pth	SepilokW	0.6	704	16	FALSE	0.386	0.561	0.457
SepilokW		221026_SepilokW/model_16.pth	Paracou	0.3	381	16	FALSE	0.479	0.378	0.422
SepilokW		221026_SepilokW/model_16.pth	Danum	0.7	278	16	FALSE	0.715	0.597	0.651
SepilokW		221026_SepilokW/model_16.pth	SepilokE	0.3	167	16	FALSE	0.551	0.521	0.535

Table B.3 A comparison of the contribution of over/undersegmentation to the accuracies across sites

Site	# test trees	Precision	Recall	F1 score	Oversegmentation*	Undersegmentation*
Paracou	381	0.595	0.543	0.568	0.206	0.336
Danum	278	0.713	0.662	0.687	0.128	0.257
Sepilok East	167	0.612	0.653	0.632	0.13	0.448
Sepilok West	704	0.604	0.656	0.648	0.226	0.234
TOTAL/AV.	*1530*	*0.631*	*0.6285*	*0.63375*	*0.1725*	*0.31875*

* As described in Clinton et al. (2010)

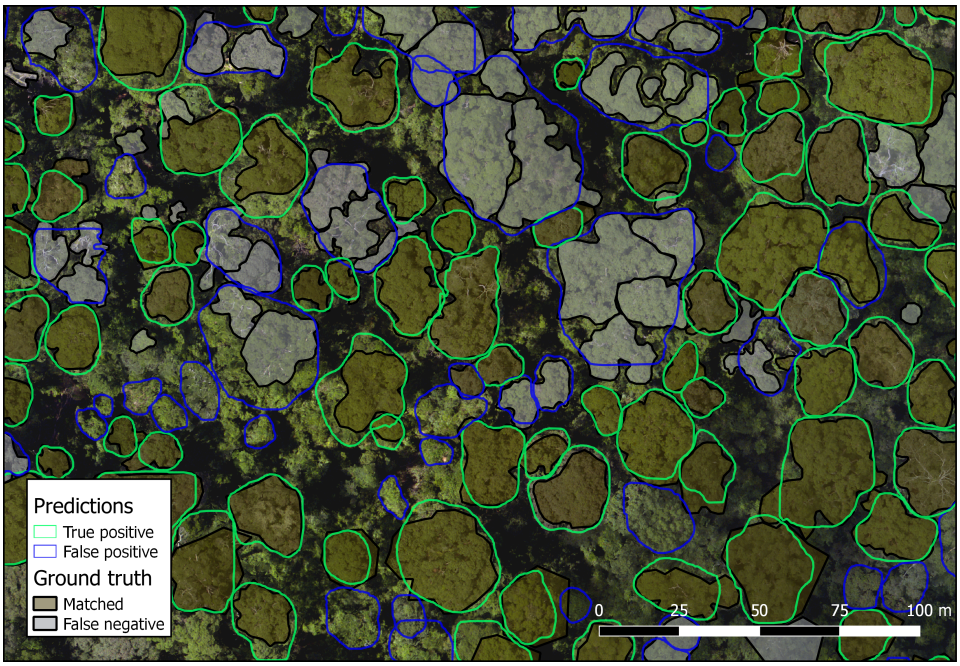


Fig. B.7 Example delineation results at Danum

Table B.4 The accuracy of predictions by tree height

Site	Conf. thresh.	Height bin (m)	Median Height	# indiv.	F1 score											
						Height bins (m)	**0-25**	**25-30**	**30-35**	**35+**						
Paracou	0.6	0-25	22.97	17	0.14	F1 score	0.14	0.42	0.614	0.719						
		25-30	28.07	103	0.42	Median height (m)	22.97	28.07	32.33	37						
		30-35	32.33	189	0.614	# individuals	17	103	189	72						
		35+	37	72	0.719											
Danum	0.7	0-25	21.04	15	0.316	**Height bins (m)**	**0-25**	**25-30**	**30-35**	**35-40**	**40-45**	**45-50**	**50-55**	**55-60**	**60+**	
		25-30	28.58	29	0.565	F1 score	0.316	0.565	0.464	0.515	0.688	0.684	0.706	0.619	0.842	
		30-35	33.125	34	0.464	Median height (m)	21.04	28.58	33.125	37.73	42.73	47.13	52.85	56.47	62.21	
		35-40	37.73	35	0.515	# individuals	15	29	34	35	46	43	47	20	9	
		40-45	42.73	46	0.688											
		45-50	47.13	43	0.684											
		50-55	52.85	47	0.706											
		55-60	56.47	20	0.619											
		60+	62.21	9	0.842											
Sepilok East	0.5	0-20	17.4	16	0.439	**Height bins (m)**	**0-20**	**20-25**	**25-30**	**30-35**	**35+**					
		20-25	22.76	61	0.595	F1 score	0.439	0.595	0.434	0.649	0.696					
		25-30	27.14	48	0.434	Median height (m)	17.4	22.76	27.14	32	37.79					
		30-35	32	33	0.649	# individuals	16	61	48	33	9					
		35+	37.79	9	0.696											
Sepilok West	0.4	0-25	21.66	36	0.231	**Height bins (m)**	**0-25**	**25-30**	**30-35**	**35-40**	**40-45**	**45-50**	**50-55**	**55-60**	**60+**	
		25-30	27.9	52	0.44	F1 score	0.231	0.44	0.471	0.604	0.609	0.571	0.674	0.621	0.692	
		30-35	32.79	119	0.471	Median height (m)	21.66	27.9	32.79	36.9	42.68	47.25	52.13	56.69	62.94	
		35-40	36.9	144	0.604	# individuals	36	52	119	144	127	107	82	26	11	
		40-45	42.68	127	0.609											
		45-50	47.25	107	0.571											
		50-55	52.13	82	0.674											
		55-60	56.69	26	0.621											
		60+	62.94	11	0.692											

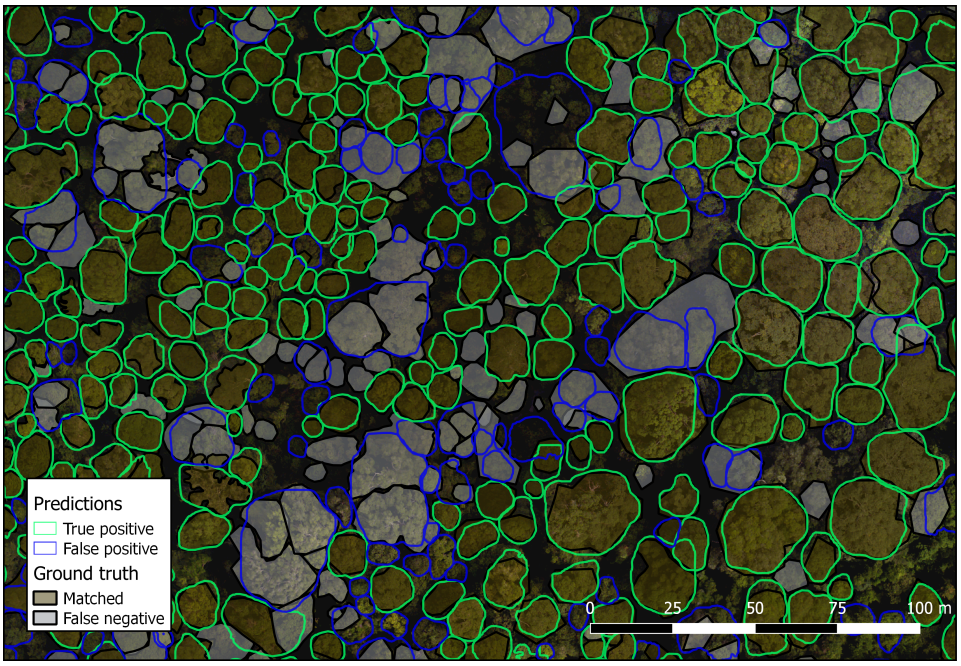


Fig. B.8 Example delineation results at Sepilok West

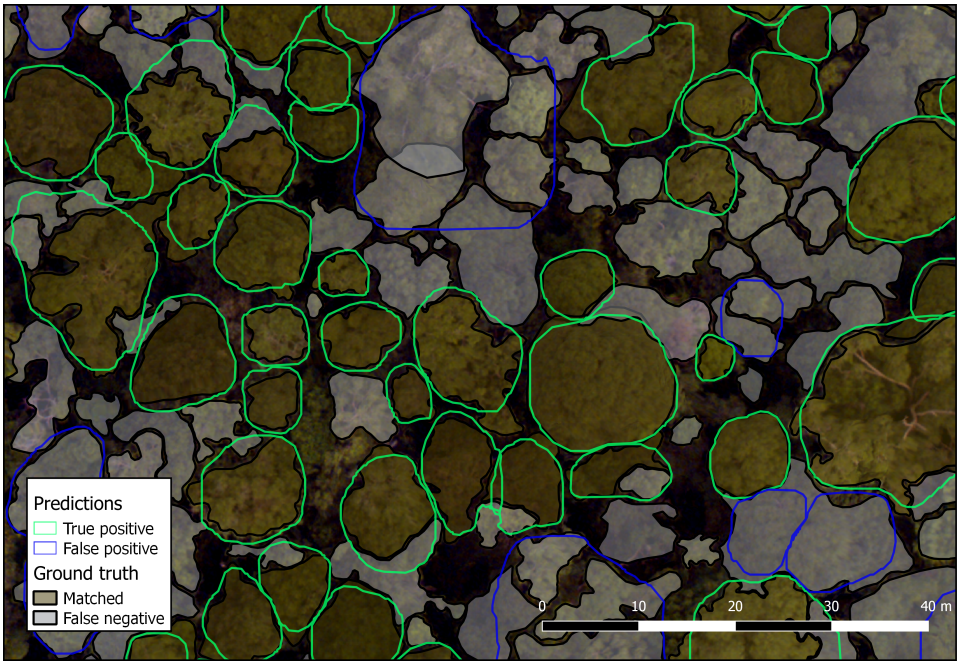


Fig. B.9 Example delineation results at Sepilok West

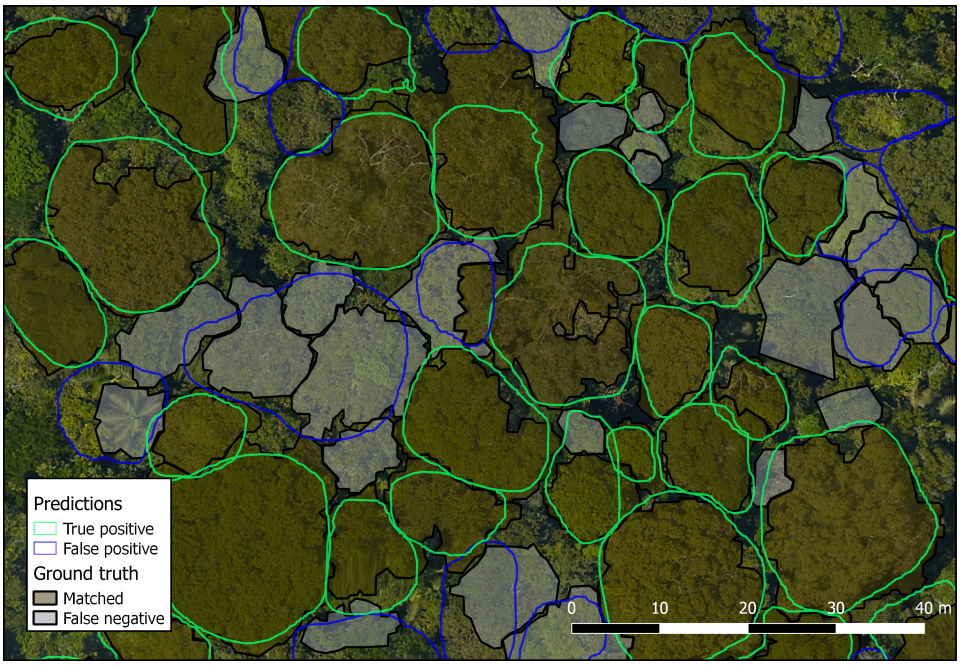


Fig. B.10 Example delineation results at Paracou

B.8 Sensitivity to image resolution

We performed a preliminary analysis on how sensitive the delineation accuracy is to changes in the image resolution. We then carried out another test, where we trained our model on the highest resolutions available to us (8 and 10 cm), before testing it on lower resolutions to determine the model's sensitivity to resolution. The lower resolutions were chosen as they are typical resolutions of modern, high-resolution satellite imagery. The final test was to train and test the model on low resolutions to determine if this workflow would increase the skill of the model on lower resolutions. As seen in Fig. B.11, illustrate that our model trained on high resolution (0.1 m) imagery is most successful when predicting on similar resolutions, and its performance degrades for resolutions an order of magnitude greater (1 m and 2 m). However, we illustrate that when the model is trained on similar resolutions to those it is tested on, performance is largely maintained. It is only on resolutions greater than 1 m that the performance degrades significantly.

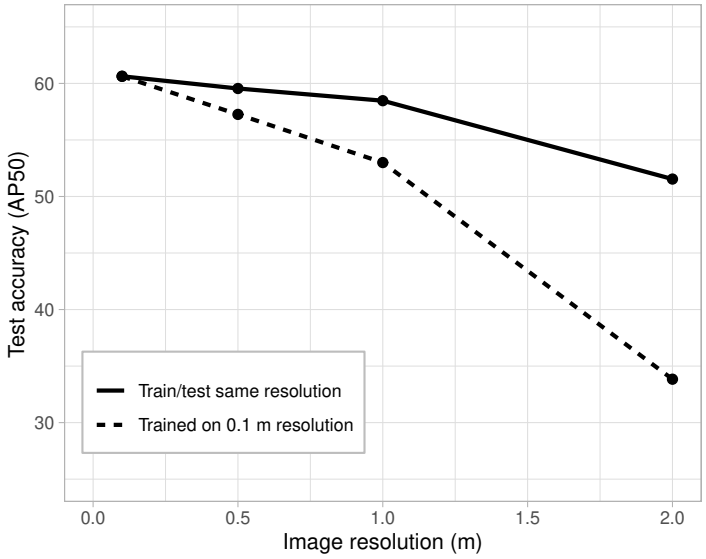


Fig. B.11 The sensitivity of the accuracy of the segmentations to the resolution of images used in training and testing.

B.9 Growth and mortality details

Growth and mortality rates were estimated from the CHMs of the repeat lidar data. Tree height was defined by the median value of the CHM within the overlayed crown polygon. To calculate growth rate (in m / yr), we evaluated the difference in height of individual trees in the two years of measurements, and divided by the time between the measurements (assuming an approximately constant growth rate).

Mortality estimates were derived by fitting a robust least squares regression to the change in tree height against the original tree height. Mortality events were defined as a drop in height of 3 standard deviations or more from this fit. The rate of mortality (%/yr) was calculated as the percentage of detected trees that died divided by the time between scans. The tree mortality rates were grouped into height bins (depending on the original height of each tree). A drop of three standard deviations is strict but it is possible that drops could result from non-mortality events such as snapping or other wind damage. Uncertainty estimates were determined by bootstrapping (repeatedly sampling from the complete set of predicted crowns).

Table B.5 The coefficients and intercepts for the robust least squares fit between original tree height and the change in tree height.

Site	Slope	Intercept
Danum	-0.110 ± 0.006	1.86 ± 0.26
Paracou	-0.028 ± 0.0003	1.06 ± 0.01
Sepilok East	-0.060 ± 0.004	1.67 ± 0.10
Sepilok West	-0.072 ± 0.003	2.72 ± 0.12

Plots for the other sites not shown in the main paper are given in Fig. B.12. These plots illustrate the different characteristics of each forest site.

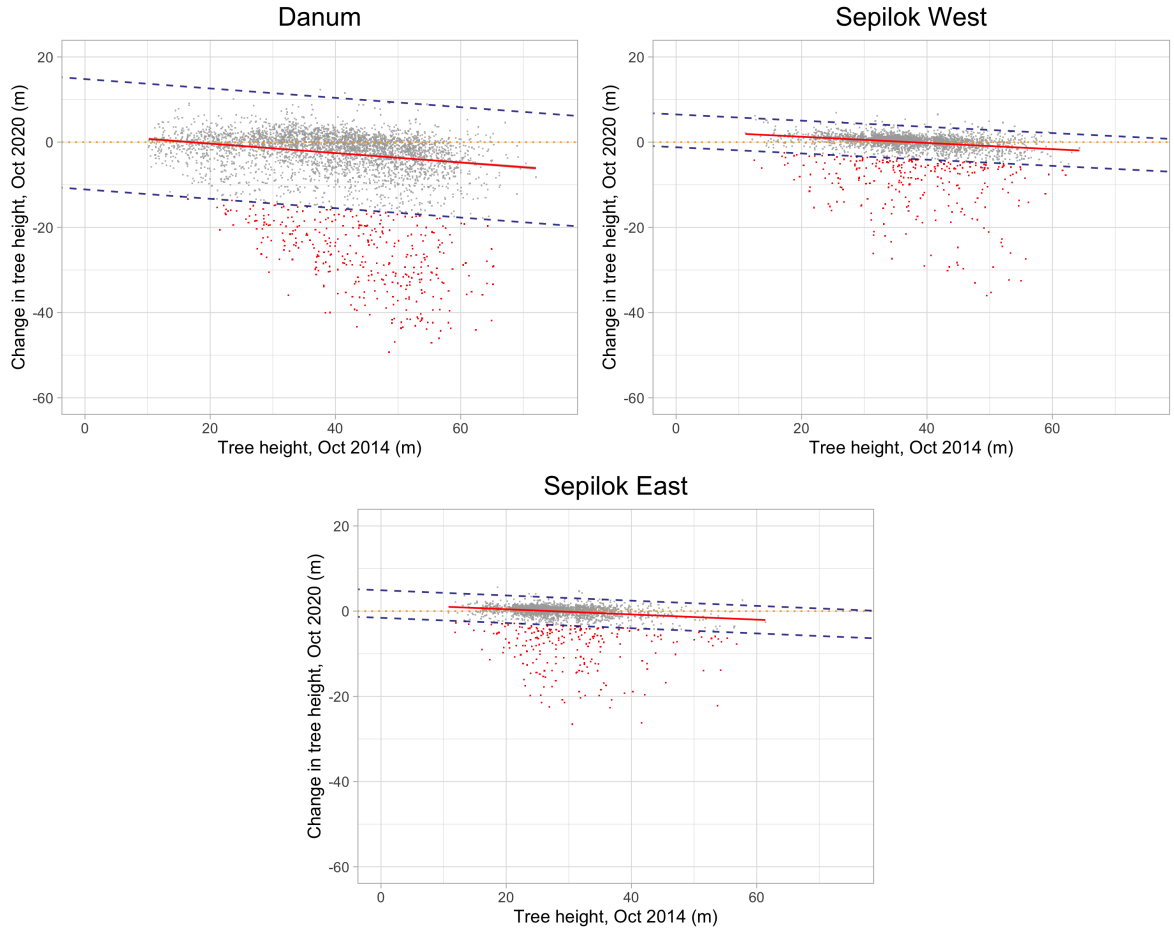


Fig. B.12 The robust least squares fit for change in height and tree height for Sabah (Danum, Sepilok West and Sepilok East). The dashed lines indicate three standard deviations either side of the best fit and red points below the lower bound indicate likely mortality events.

Table B.6 Extended results of growth an mortality.

Site	Height class	Median height	Total trees	# deaths	Mortality (%/yr)	SE mortality	Growth (m/yr)	SE growth
Paracou	0-25	22.635	12050	596	1.66004642	0.0905	0.296113478	7.00E-04
	25-35	29.5025	37302	1750	1.632460635	0.0011	0.298936236	4.00E-04
	35-45	37.03	6244	392	1.790424383	0.0275	0.295832509	0.001
	45+	45.78	51	8	2.393584517	0.2675	0.267323163	0.0164
Danum	0-25	19.74	447	13	1.220346295	0.0659	0.181524783	0.0015
	25-35	31.08	785	64	1.479070473	0.0332	0.171324629	0.0018
	35-45	40.38	1167	106	1.509182779	0.0268	0.169589407	0.0015
	45-55	49.63	1106	131	1.585789391	0.0231	0.164438566	0.0018
	55-65	57.6975	347	50	1.644841127	0.0417	0.159654501	0.0033
	65+	66.245	20	3	1.657215864	0.357	0.158552509	0.0164
Sepilok West	0-25	22.17	198	19	1.530423123	0.0847	0.182479671	0.0039
	25-35	32.105	1035	60	1.391946925	0.0329	0.18454579	0.0013
	35-45	38.78	1281	80	1.411583578	0.0809	0.184215792	0.0012
	45-55	48.65	471	51	1.565610468	0.0758	0.178992798	0.0027
	55+	56.73	57	7	1.603143402	0.138	0.178277836	0.0084
Sepilok East	0-25	22.59	769	51	1.427640015	0.035	0.175701939	0.0017
	25-35	28.7025	1253	95	1.464055351	0.041	0.173914558	0.0026
	35-45	37.44	277	30	1.565673028	0.12	0.167801421	0.0114
	45-55	50.1	54	15	1.869296162	0.276	0.135909338	0.0162
	55+	56.45	7	2	1.879232094	0.662	0.134415829	0.0341

Appendix C

Chapter 4 Supplementary materials

C.1 Three-way human manual segmentation comparison

To test how well different human interpreters agree when identifying tree crowns from imagery provided to them, two plots at Paracou were selected to be delineated by three expert human analysts. Three expert humans analysts (familiar with remote sensing data and segmentation methods), Analysts A, B and C were asked to segment the tree crowns of two plots. The segmentation was performed in QGIS with the following data layers available to the human analysts: 2015 and 2016 RGB and lidar CHM, 2016 hyperspectral layers.

The three sets of delineations for each of the two plots were compared against each other to determine the degree of congruence. A match was granted when polygons from one set had a Jaccard/IoU > 0.5 with one from another set.

Plot 1 results
Comparison A-B: 28 congruent segments (Jaccard>0.5): 90% of A's crowns and 56% of B's
Comparison A-C: 30 congruent segments (Jaccard>0.5): 97% of A's crowns and 60% of C's
Comparison B-C: 42 congruent segments (Jaccard>0.5): 84% of B's crowns and 84% of C's.

Plot 2 results
Comparison A-B: 19 congruent segments (Jaccard>0.5): 65% of 's crowns and 48% of B's
Comparison A-C: 22 congruent segments (Jaccard>0.5): 76% of A's crowns and 46% of B's
Comparison B-C: 28 congruent segments (Jaccard>0.5): 72% of B's crowns and 58% of C's.

The results show that humans can interpret images quite differently highlighting the challenge of achieving high accuracy with automated methods of tree crown delineation.

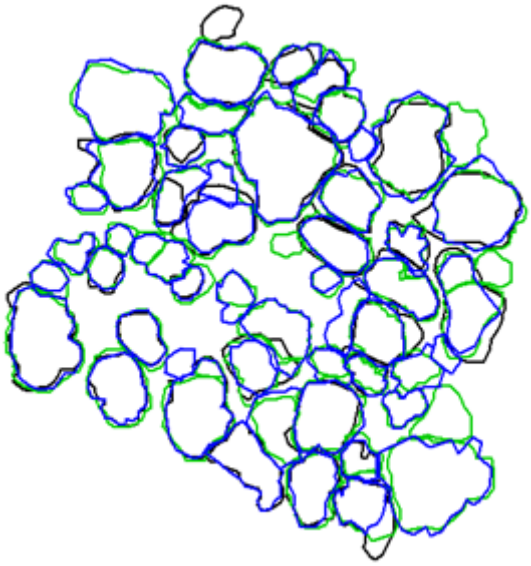


Fig. C.1 A comparison of manual human tree crown delineations (Plot 1). Set A: black, Set B: green, Set C: blue.

C.2 Tree crown database

The dataset and data description is available in this GitHub repository: https://github.com/umr-amap/ParacouTrees.

To train machine learning algorithms and evaluate automatic tree crown delineation and species identification from remote sensing data it is necessary to have an extensive ground truth map of tree crowns. Generating this takes time and attention to the specific attributes of each crown. Careful ground validation is necessary to have confidence in correct individual/species assignment and delineation (avoiding over/under segmentation).

C.2.1 Premises of the dataset

The growth of crowns is relatively slow, meaning an undisturbed crown will not change its shape significantly between scans/field missions. Creating crown polygons is a time consuming process as it requires a careful comparison/contemplation of the different modalities (and time steps thereof) of scans against field inventory data.

Significant changes to the crown are due to:

- Tree death
- Branch fall

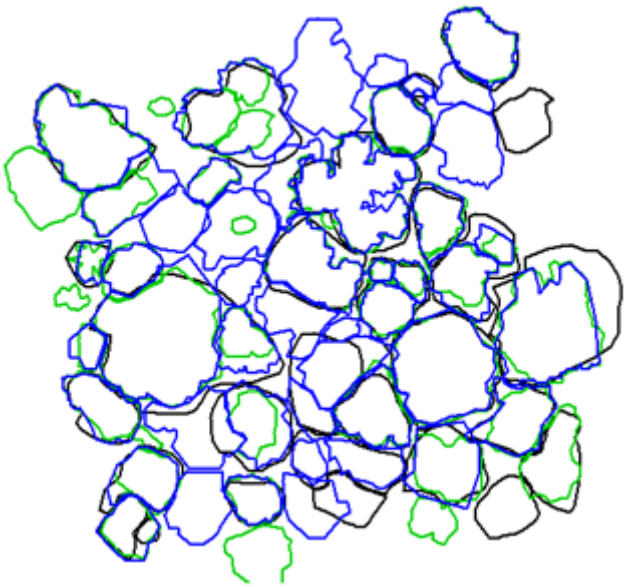


Fig. C.2 A comparison of manual human tree crown delineations (Plot 2). Set A: black, Set B: green, Set C: blue.

It is not feasible or time efficient to produce a new set of crowns for each new scan. Instead, the crowns are updated (by hand) when a significant change is detected. The fields `StartDate` and `EndDate` are used to track the validity of a crown. `fid_1` is a unique identifier for an individual tree - multiple, temporally distinct, crown polygons may be associated with an `fid_1`. `StartDate` and `EndDate` are set to `NULL` when a crown is created. If a crown is seen to no longer be valid (e.g. due to a branch fall or mortality), `EndDate` is set to the date of the scan that shows this. A new crown may be created if an existing crown changes due to branch fall, if a significant portion of an existing crown is revealed by a branch fall or the mortality of an occluding tree, or if a new tree is discovered. Full details of the fields are given below.

C.2.2 Fields of dataset

A series of fields are used to describe the crown polygons:

- `fid (int)`: unique identifier for each crown polygon
- `fid_1 (int)`: a unique identifier for individual trees (not polygons). This can be useful to track individuals if a crown has changed significantly through time (see `StartDate`, `EndDate`).
- `Site (str)`: Paracou, Nouragues etc.

- `PlotOrg (str)`: Necessary at Paracou (CIRAD, CNES or INRA). This helps in linking the polygons to the inventory datasets.
- `PlotNum (int)`: plot number
- `SubPlot (int)`: some plots have subplots contained within them
- `LocalID (int)`: the tree number as recorded on the tree's tag
- `TrunkMatch (int)`: $\{1,2,3,4\}$ These integers describe how well the crown polygon (as delineated from the remote sensing data) has been matched to a trunk in the field.
- `CrownIntegrity (int)`: $\{1,2,3,4\}$ These integers describe how sure we are that a delineated polygon is that of a single, complete crown.
- `Lianas (bool)`: as to whether lianas as present in the crown of the tree delineated
- `StartDate (date)`: Date at which the crown becomes visible or has changed shape
- `EndDate (date)`: Date at which the crown becomes absent or has changed shape
- `Dead (bool)`: a crown might be present but belong to a dead tree
- `GroundValid (bool)`: has the crown been checked in the field?
- `Creator (str)`: name of the person to have made the polygon
- `Comments (str)`: for any comments before or in the field
- `BaseLayer (str)`: which remote sensing datasource has been used as the "anchored" location of the crown

C.3 Additional plots

Fig. C.3 shows the frequency distribution of pixels by species. Fig. C.4 shows the variation in spectra between two individuals of a single species. Fig. C.5 shows the relative variation of spectra between and within species.

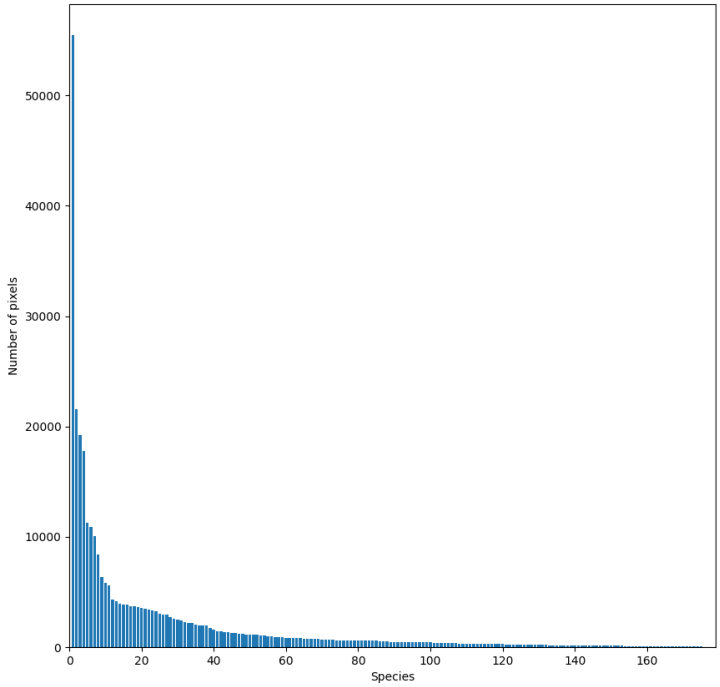


Fig. C.3 Distribution of pixels available for training by species.

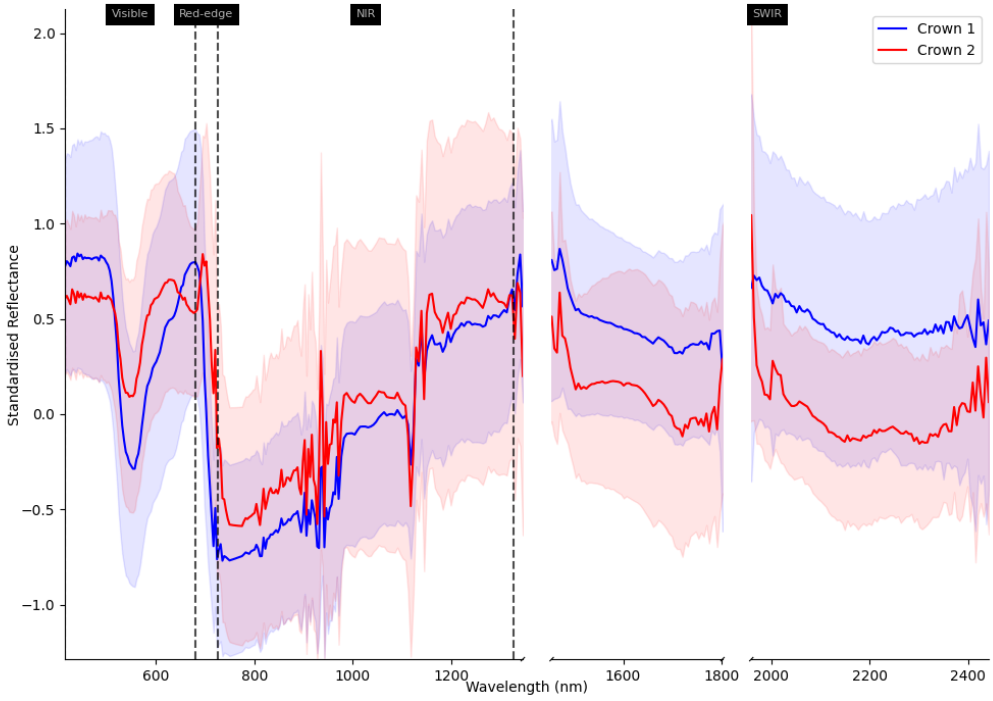


Fig. C.4 Spectral values distribution for pixels in two separate *Pradosia cochlearia* crowns. Shaded areas indicate standard deviation of spectral values for each crown.

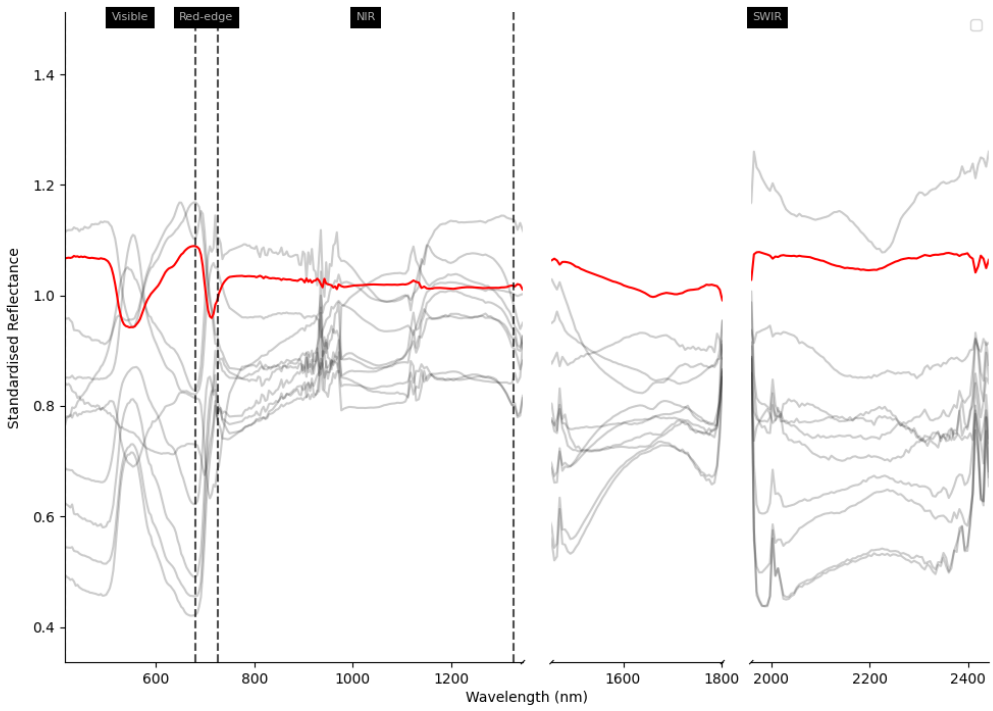


Fig. C.5 Intraspecies variation vs interspecies variation across the top ten most common tree species by pixel frequency. Each grey line is the standard deviation of spectral values for a given species at that spectral band. Low values indicate consistency for a species at a spectral band. The red line is the standard deviation across the species.

Appendix D

Chapter 5 Supplementary materials

D.1 Modelling flux tower footprint

To estimate the scanning area that would most effectively isolate the trees responsible for the signals received at the flux tower sensors, we modelled the *flux tower footprint*. A flux tower footprint refers to the geographical area that is responsible for the fluxes of carbon dioxide, water vapor, and other gases that are measured at the sensor atop of the tower. It therefore delineates the patch of forest for which the calculated fluxes, exchanges and productivity (GPP, NEE etc.) are valid. The footprint is influenced by a range of factors including the height of the tower, the roughness of the surface terrain, and atmospheric conditions, and varies through time as conditions change, which can make it difficult to calculate.

To estimate the footprint area, we used the scaling approach of Kljun et al. (2004, 2015) which maps the footprint region from highest contribution to lowest in two dimensions onto the forest canopy. This Flux Footprint Prediction (FFP) method, designed to work with data obtainable from flux tower measurements, has been shown to correspond well more complex, computationally intensive dispersion models. It gives not just the extent but also the width and shape of footprint estimates,. From the meteorological data recorded at the flux tower (including wind speed and direction) and physical parameters of the tower/sensor and canopy, we calculated regions from highest contribution of fluxes to lowest contribution over three month periods from 2014 to 2019 .

Based on these footprints it was possible to define two scanning areas: (1) a *large* area that covers the top 60% contribution throughout the year and (2) a smaller *dense* area that covers the top 30% contribution throughout the year. Based on the observed temporal variability of the footprint we were confident that our scanning area would be the primary source of the signals received by the eddy covariance sensor.

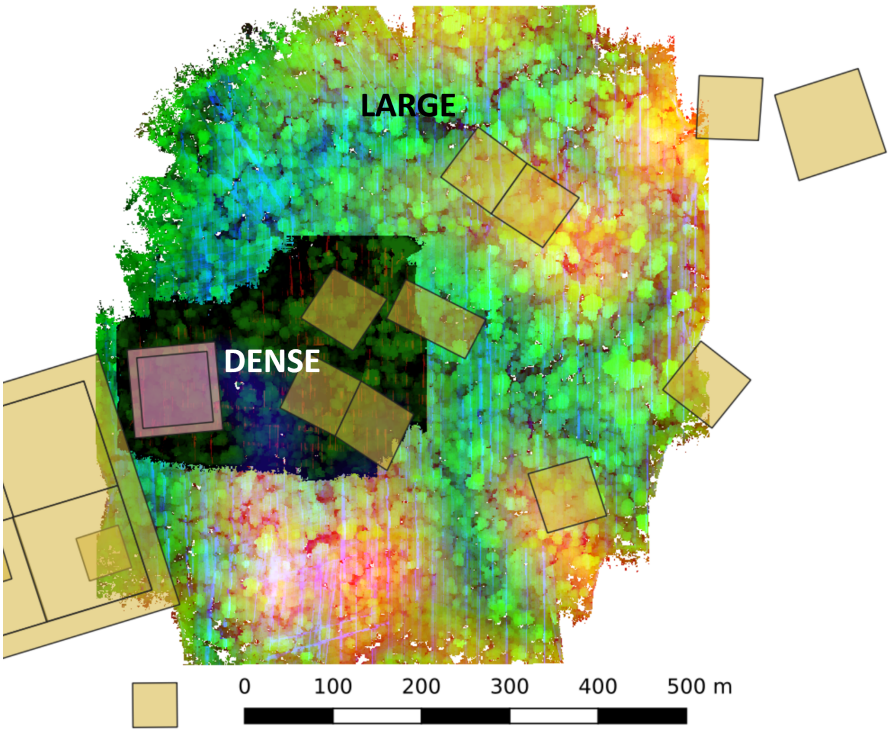


Fig. D.1 Dense and large scan area

D.2 Rainfall

To understand whether our dry central point (27-Sept) was representative over the years of study we plotted the monthly rainfall (Fig. D.2).

D.3 Tree crown dataset

As described in Chapter 4 we used automated methods to extrapolate our manual crown dataset across the entire landscape in the flux tower region. For the automatic crowns with selected those for which we had good confidence in the segmentation and the species (based on the confidence scores of the methods). This allowed us to increase our sample from 1,243 manual crowns to 5,116 manual and automatic crowns. This was particularly useful for those species with limited representation in the manual dataset (but not the rarest species which were dropped). The 100 most dominant species were retained for further analyses as these had a sample size large enough to have some confidence in the species level metrics described below. These metrics require a comparison between trees of the same species.

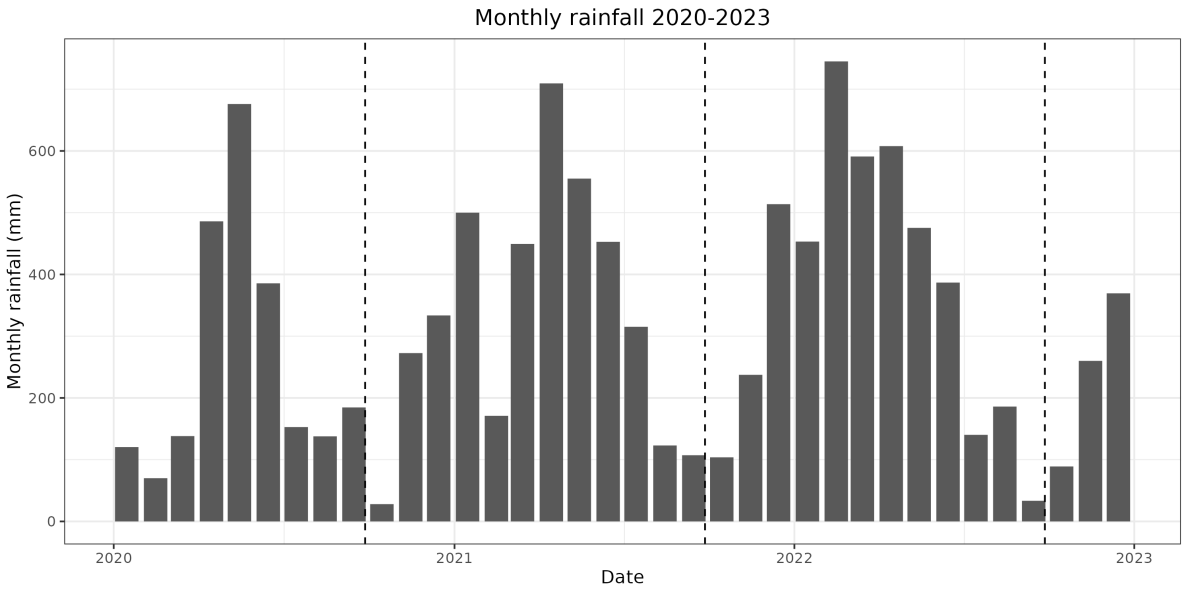


Fig. D.2 Monthly rainfall at Paracou across the period of the phenological study. The dashed line represents the historic dry point of the year (27-Sept).

D.4 Processing UAV data

D.4.1 GLI crown level extraction

GLI extraction was normalised with a 25 m buffer around the crown to account for local and inter-temporal perturbations (see Fig. D.3).

D.4.2 Pointcloud to PAD (AMAPvox)

The processing of the multitemporal lidar data required several processing steps before it was ready to analysis including:

- PosPac: Trajectometry optimization, using DGPS information (local base or KROU or KOUG)
- BayesMap Jalobeanu and Gonçalves, 2014 and ICP (CloudCompare) Besl and McKay, 1992; Rusinkiewicz and Levoy, 2001 were used to improve flight line and point clouds matching (impoved point cloud and trajectometry)
- Cloud Station: Point cloud and trajectometry export
- LasTools: Clipping and DTM extraction

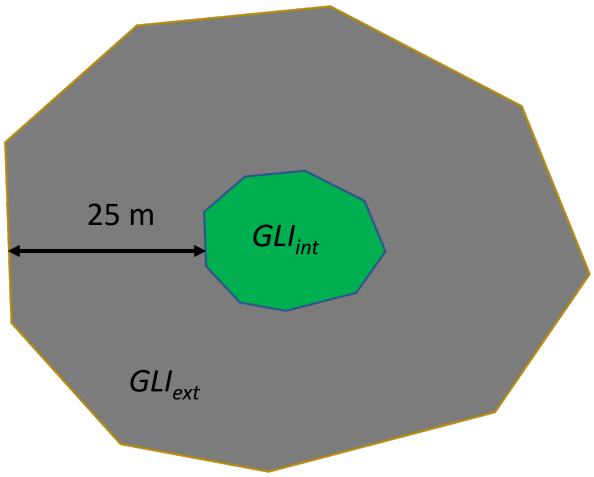


Fig. D.3 The crown (green) GLI was normalised by the GLI in a region of 25m buffer around the crown to reduce noise from perturbations across the orthomosaic.

- AmapVox: PAD computation Vincent et al., 2017
- Tree crown level extraction of PAD

The PAD readings were validated with LAI2200 readings and PAR sensors:

- On top of the flux tower :
  - 2 sensors, acquisition frequency : one measurement per 15 minutes
- In CNES plots:
  - Spacing/coordinates : 25 sensors in the middle of the plot, evenly spaced on a 50*50m grid
  - Acquisition frequency : one measurement per 15 minutes
- In INRA plots
  - Spacing/coordinates : 13 sensors in total, located in the P1 and P9 corners (6 sensors) and middle of the plots' sides (7 sensors)
  - Acquisition frequency : one measurement per 15 minutes

BayesMap was used to help align point clouds through time Jalobeanu and Gonçalves, 2014

AMAPVox is an open-source software for analysing lidar-vegetation interactions that we used to estimate plant area density (PAD) in a 3D voxelised representation of the canopy.

AMAPvox tracks every laser pulse through a 3D grid, calculating the effective sampling area and local attenuation in each voxel based on laser beam characteristics and pulse interactions within the voxel. Utilizing different return weighting options and estimation methods, it allows for the computation of canopy attenuation profiles by integrating the 3D local attenuation data horizontally while considering ground elevation.

PAD data was produced with voxel size 1m and 2m and beam angle filtering 20 °and 45 °. For details on UAV bias/correction see Vincent et al. (2023). For AMAPvox parameterisation see Vincent et al. (2021).

D.5 Additional plots

D.5.1 Phenograms

To visually inspect phenological patterns within each species we created heatmap "phenograms" (Fig. D.4).

D.5.2 Periods of trees

We assessed the periodicity of all trees. Fig. D.7 gives the distributions for ten dominant families.

D.5.3 Amplitude of signals

Fig. D.8 shows the amplitudes of the phenological signals.

D.5.4 Shedding seasonality

Fig. D.9 shows the seasonality of minimum GLI.

D.5.5 Pairwise correlations

To check for correlations between the phenological metrics we plotted the pairwise correlations (Fig. D.10).

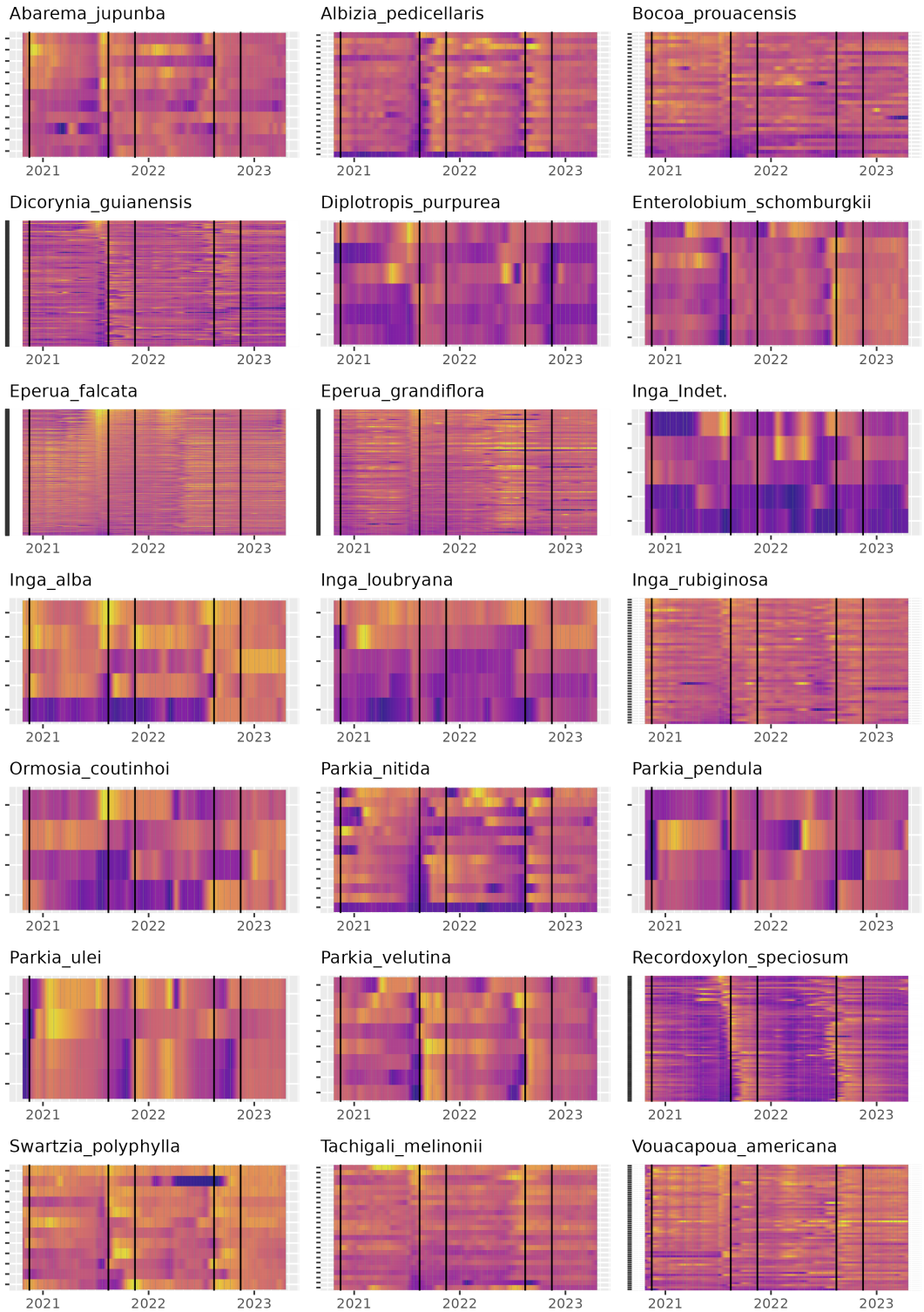


Fig. D.4 Phenograms for Fabaceae. The temporal variation in relative Green Leaf Index for each species. The y-axis (with ticks) shows each individual of the species and the x-axis shows time. Bright colours show fresh leaves where as dark colours show senescence/shedding. The black vertical lines show the beginning and end of the dry season.

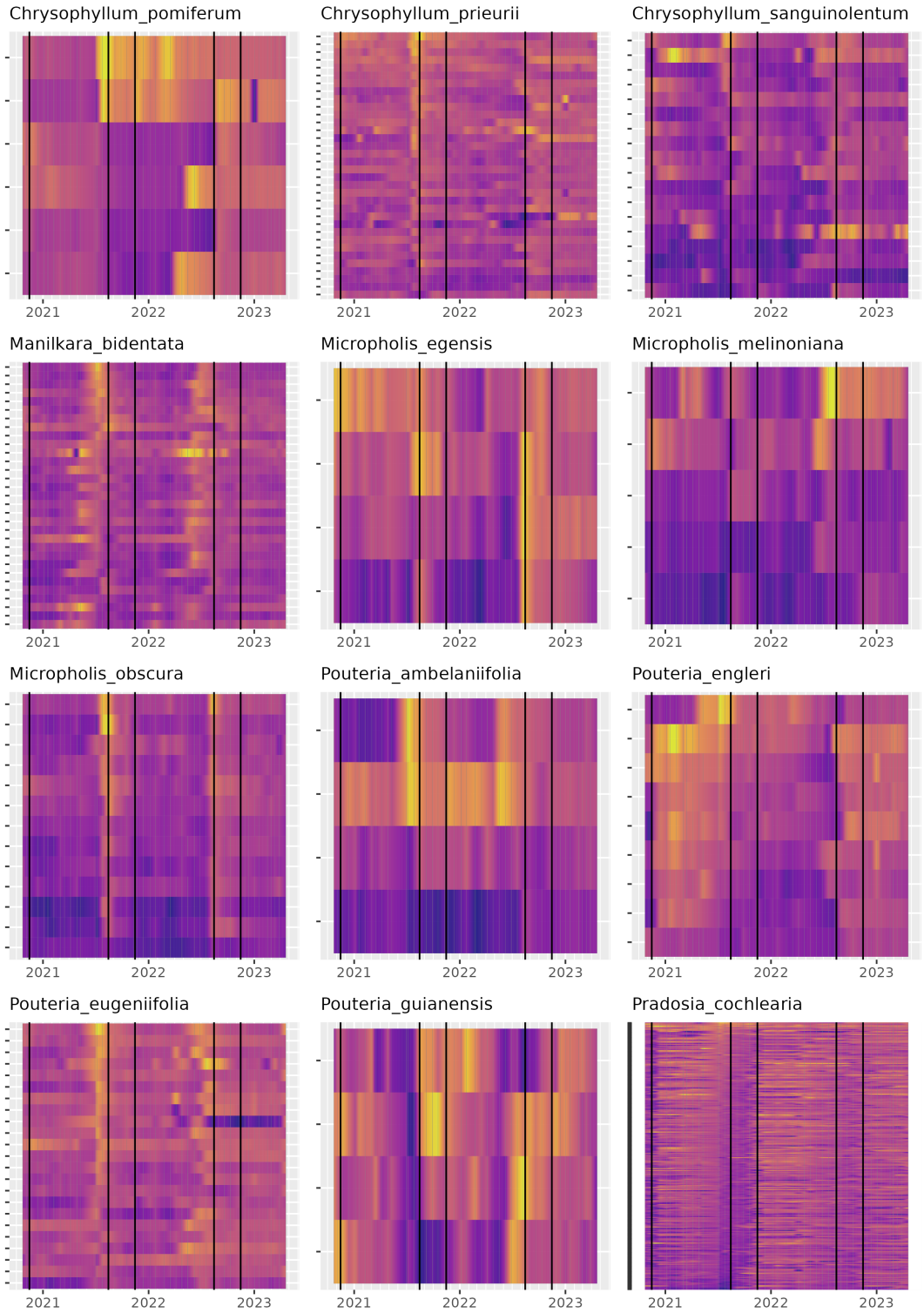


Fig. D.5 Phenograms for Sapotaceae. The temporal variation in relative Green Leaf Index for each species. The y-axis (with ticks) shows each individual of the species and the x-axis shows time. Bright colours show fresh leaves where as dark colours show senescence/shedding. The black vertical lines show the beginning and end of the dry season.

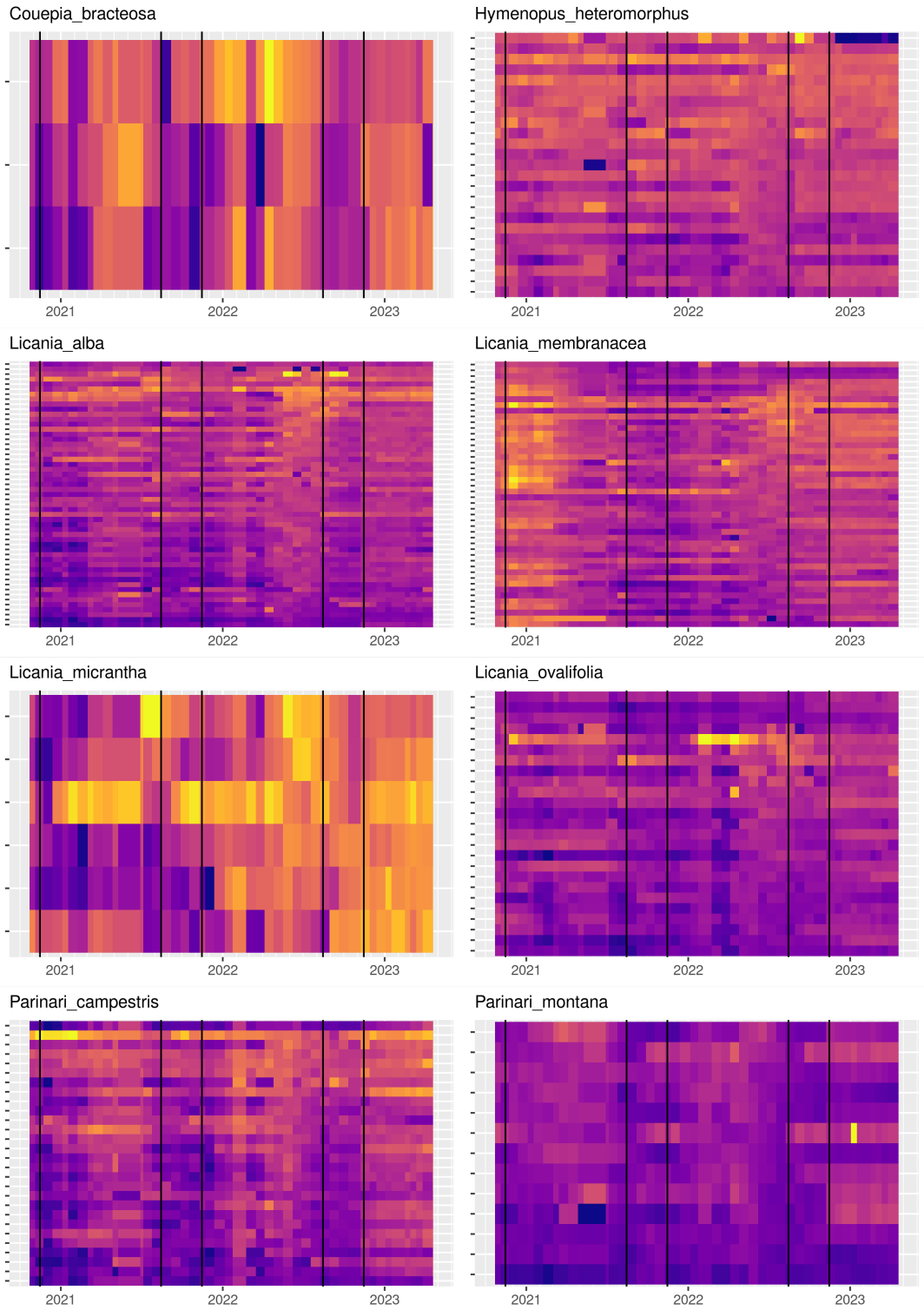


Fig. D.6 Phenograms for Chrysobalanaceae. The temporal variation in relative Green Leaf Index for each species. The y-axis (with ticks) shows each individual of the species and the x-axis shows time. Bright colours show fresh leaves where as dark colours show senescence/shedding. The black vertical lines show the beginning and end of the dry season.

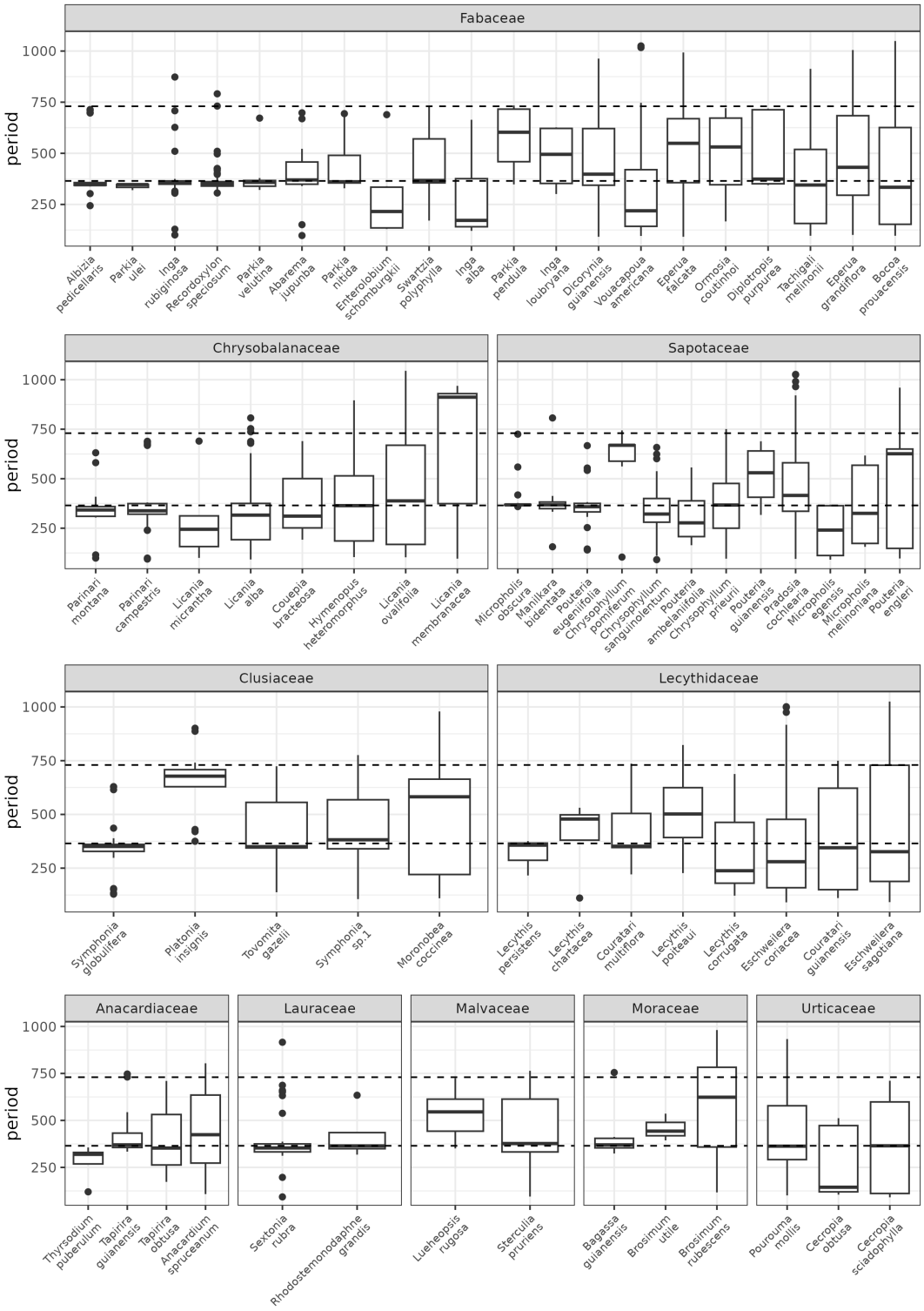


Fig. D.7 Measured periods of the species based on auto-correlation.

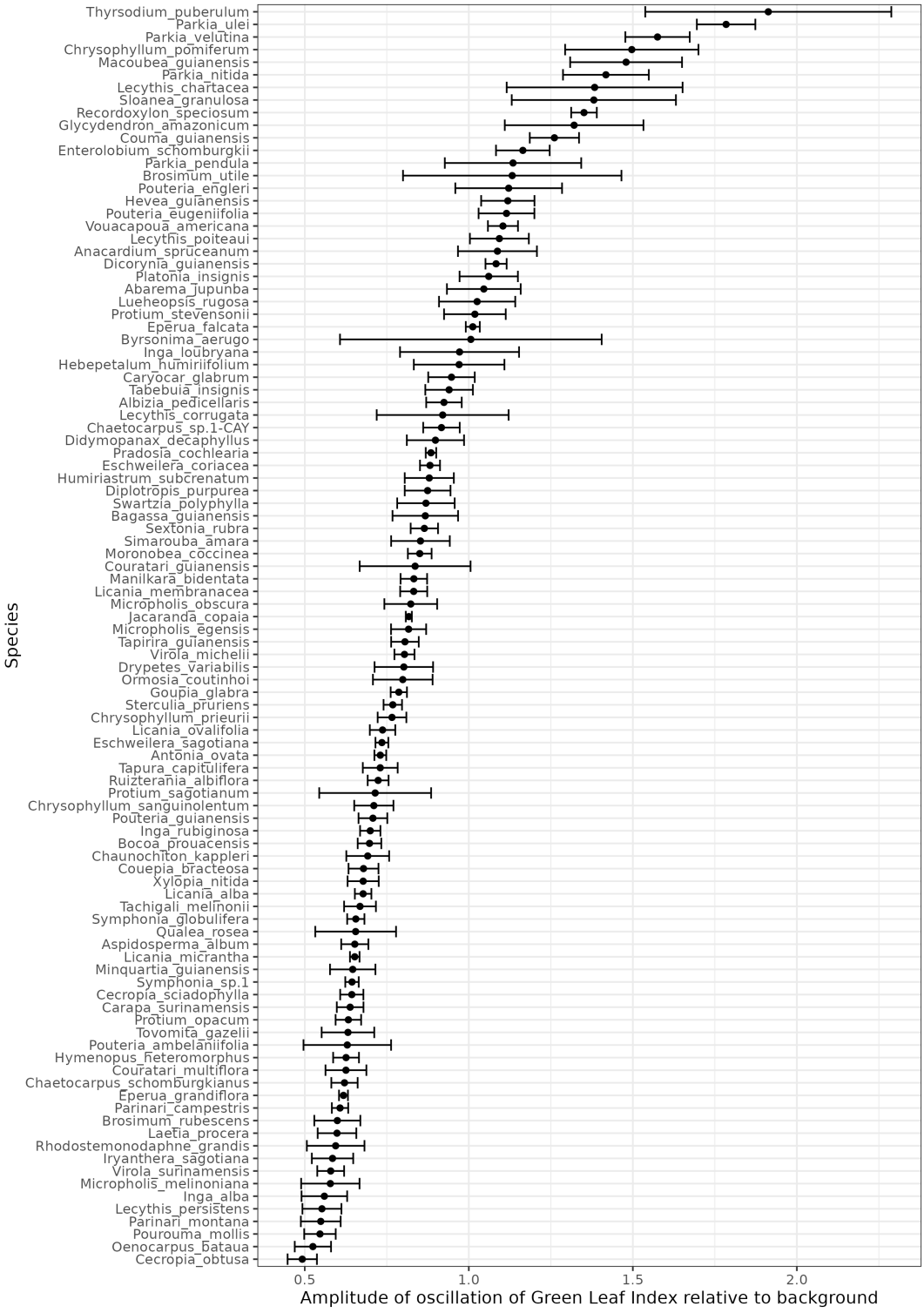


Fig. D.8 The amplitude of the phenological signals averaged at the species level. The amplitude of a GLI signal was taken to be the difference between its minimum and maximum recorded value.

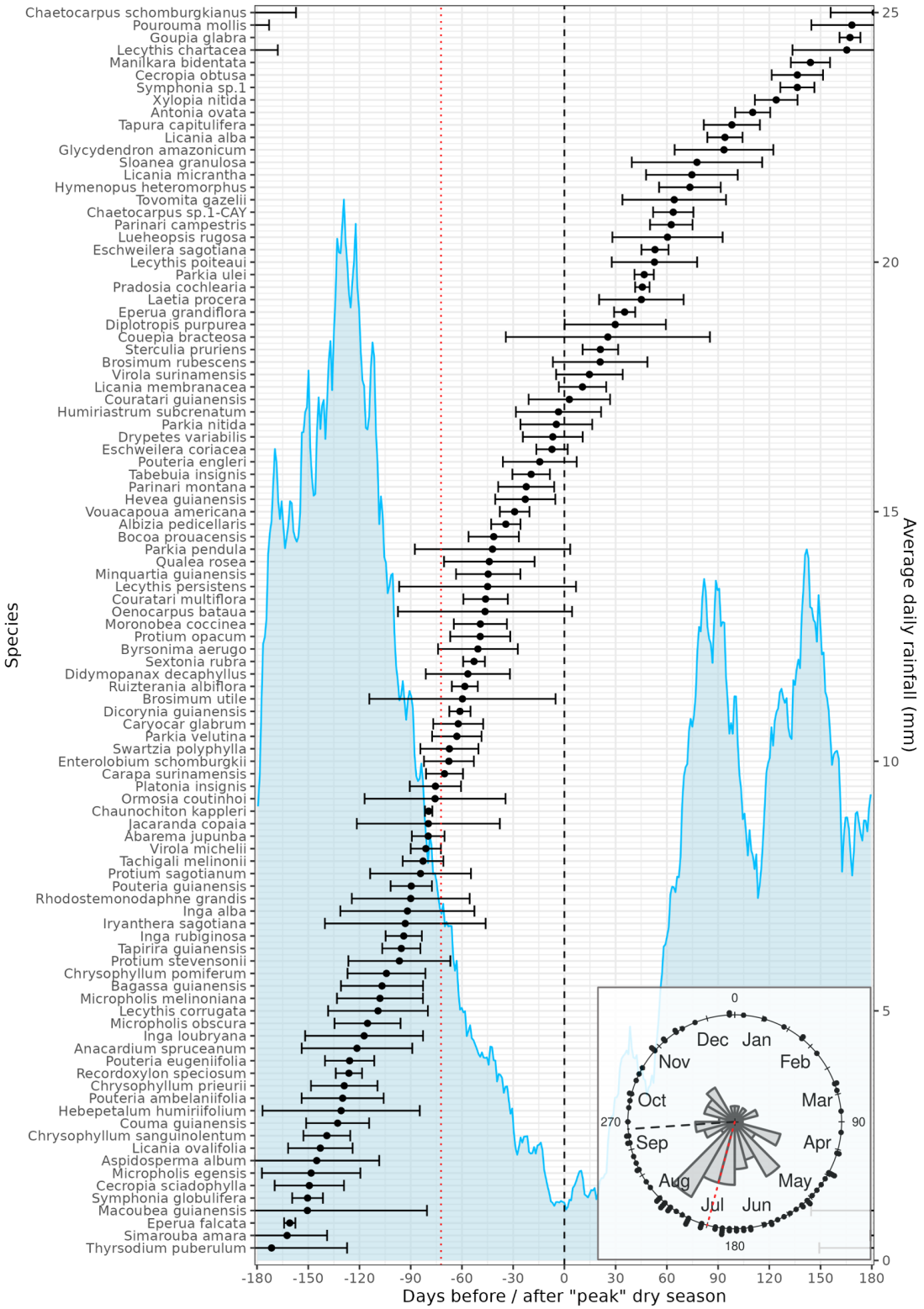


Fig. D.9 The average timing of minimum greenness (normalised GLI) for each species relative to the driest point of the year (20th of September) overlaid on the average daily rainfall at the site. Negative values are before the driest point and positive values are after. The dotted red line show the (circular) mean timing of peak greenness across all crowns. Inset is the radial histogram for crown shedding/ date showing the driest day and mean peak of greening with the same line style as the main plot.

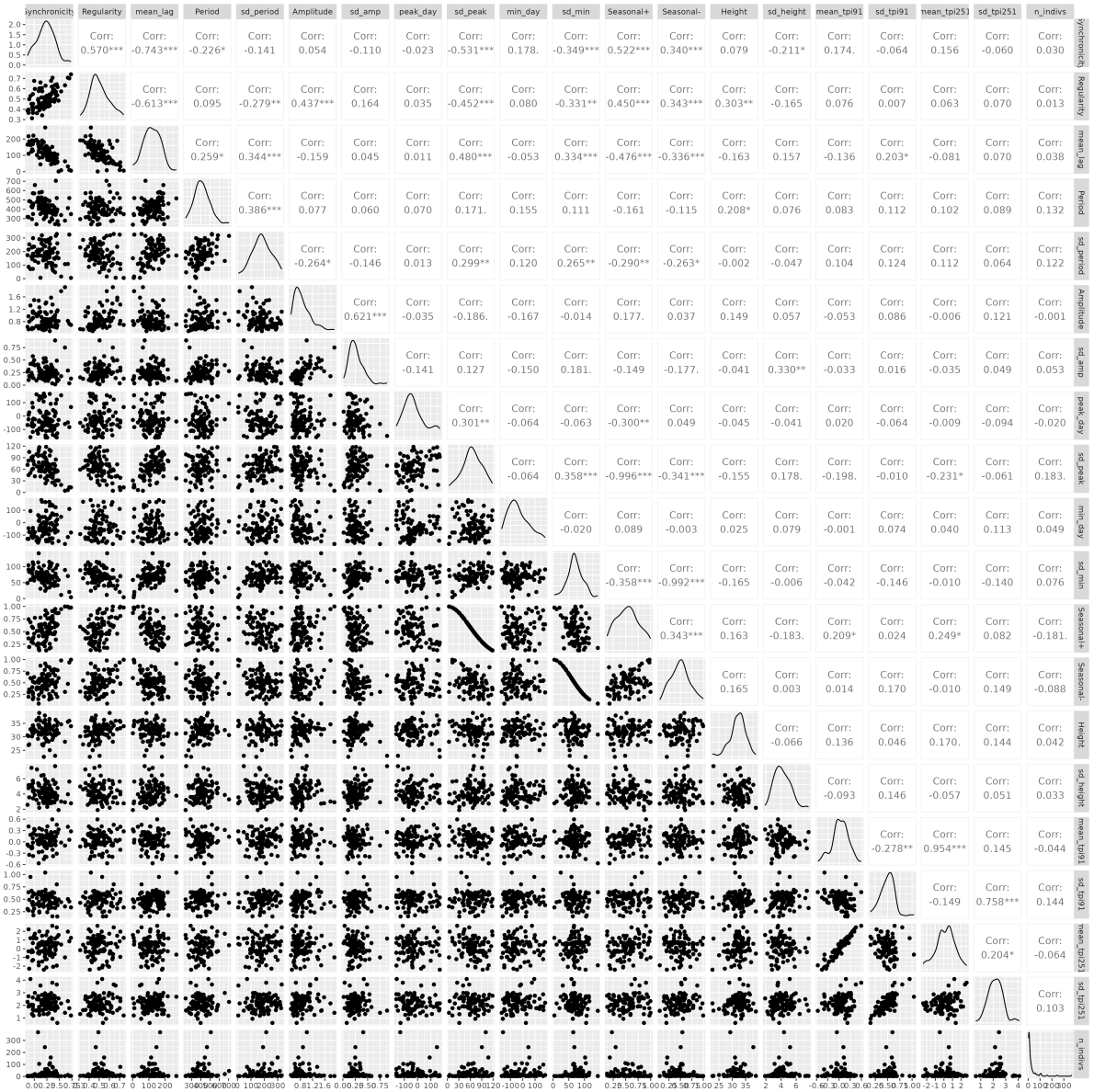


Fig. D.10 Pairwise correlations of the phenological metrics and height and TPI variables. *Synchronicity* is a measure of how well the signals of trees within a species match at any given point in time, *regularity* is a measure of how similar the signal is between trees within the a species (regardless of of well aligned in time they are in time) and *lag* is the mean shift in time (days) required to align the signal of trees within the same species. *Amplitude* is a measure of the strength of the observed signal.