



HAL
open science

Classification des types de prairies et estimation de la diversité taxonomique à partir de séries temporelles d'images satellites

Marc Lang

► **To cite this version:**

Marc Lang. Classification des types de prairies et estimation de la diversité taxonomique à partir de séries temporelles d'images satellites. [Stage] France. Ecole Nationale Supérieure Agronomique de Toulouse (INPT - ENSAT), FRA. 2015, 84 p. hal-02801734

HAL Id: hal-02801734

<https://hal.inrae.fr/hal-02801734>

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université de Toulouse
Institut National Polytechnique de Toulouse
École Nationale Supérieure Agronomique de Toulouse

RAPPORT DE PROJET DE FIN D'ÉTUDES

pour l'obtention du

Diplôme d'Ingénieur Agronome
Spécialité : Agrogéomatique

**Classification des types de prairies et estimation de la
diversité taxonomique à
partir de séries temporelles d'images satellites**



LANG Marc

UMR 1201 DYNAFOR
Dynamiques et écologie des paysages agriforestiers



REMERCIEMENTS

Je tiens tout d'abord à remercier chaleureusement Mathieu Fauvel et David Sheeren, mes encadrants, qui m'ont beaucoup apporté lors de ce stage. Ils ont su m'aiguiller, me conseiller et répondre à mes (nombreuses) questions tout en me laissant une grande autonomie. J'ai vraiment apprécié de travailler avec eux pour leur rigueur, pour leurs compétences et pour leur sympathie. Encore une fois merci à eux.

Je tiens ensuite à remercier Marc Deconchat, directeur du laboratoire DYNAFOR, pour m'avoir accueilli au sein de l'équipe. Merci également à l'équipe administrative aussi bien pour leur aide dans les démarches administratives que pour leurs explications sur l'art de relier un rapport de stage.

Je remercie Jean-François Dejoux du CESBIO pour son avis extérieur sur le stage et pour avoir fourni les images satellites sans lesquelles ce stage n'aurait pu avoir lieu. Merci à Jean-François Caux pour son expertise sur la zone d'étude et les précieux renseignements qu'il m'a fournis pour contacter les agriculteurs. Je remercie d'ailleurs grandement les agriculteurs qui ont accepté que je vienne faire des relevés botaniques sur leurs parcelles, et qui ont pris du temps pour discuter avec moi du mode de gestion de leurs prairies.

Je remercie Jérôme Willm pour m'avoir appris à reconnaître les plantes dans une prairie. Je suis maintenant un peu plus compétent dans la difficile discipline de la détermination des graminées.

Merci à Vincent, pour ses conseils avisés sur les choix sémiologiques, pour notre passion commune pour le thé et pour son éternelle bonne humeur. Merci également pour les pauses de non-fumeur, qui ont grandement amélioré ma productivité lors de ce stage.

Merci aux thésards du laboratoire, Romain, Hugo, Pierre-Alexis, Mailys, Anthony et Nicolas mais aussi à ceux d'Ecolab, Léo et Adrien, pour les nombreuses discussions sur des sujets divers et variés telles que les secrets du fonctionnement d'un lyophilisateur, l'intoxication alimentaire au cyanure en passant par l'intérêt esthétique d'une sucrière. Plus sérieusement, ils m'ont beaucoup éclairés sur le déroulement d'une thèse et m'ont conforté dans mon choix d'en réaliser une moi-même. Je les remercie également pour avoir participé à la bonne ambiance qui a régné tout au long du stage.

Merci à tous les stagiaires de DYNAFOR, qui ont partagé la même expérience que moi pendant ces six mois au sein de ce laboratoire. Ils m'ont permis de m'ouvrir sur d'autres thématiques que celui de mon stage.

Merci enfin à tous les autres membres de l'équipe qui n'ont pas été mentionnés ici mais qui ont participé de près ou de loin à la qualité de ce rapport.

SOMMAIRE

Remerciements.....	2
Résumé.....	1
I. Introduction.....	2
A. Présentation de la structure d'accueil.....	2
B. Etat de l'art et objectifs	4
1. Caractérisation des Prairies.....	4
2. La Télédétection : un outil adapté à l'étude des prairies à large échelle ?.....	6
3. Bilan de l'état de l'art.....	8
4. Objectifs et hypothèses du stage	9
5. Apport du précédent rapport de stage.....	10
II. Matériel et méthodes	12
A. Présentation de la zone d'étude	12
B. Présentation de la démarche générale.....	13
C. Données satellitaires	14
1. Caractéristiques générales	14
2. Images disponibles et sélection des données.....	14
D. Données auxiliaires.....	16
1. Carte des occupations du sol.....	16
2. Le registre Parcellaire Graphique (RPG).....	16
E. Relevés terrain	17
1. Sélection des prairies	17
2. Protocoles de relevés terrain	18
3. Validité des données botaniques	20
4. Construction des classes botaniques.....	21
5. Bilan de la campagne terrain	23
F. Traitements statistique des données.....	25
1. Traitements préliminaires	25
2. Classifications statistique supervisées	27
3. Modélisation linéaire	31
4. Logiciels utilisés et implémentation	33
III. Résultats et analyse	35
A. Masque des prairies.....	35
B. Classification sur le mode de conduite	36
C. Classifications issues la composition botanique.....	36
D. Modèles linéaires.....	38
IV. Discussion.....	39

A.	Une distinction entre fauche et pature possible	39
B.	Une distinction fondée sur les typologies botaniques difficile.....	40
1.	De nombreux paramètres non contrôlés	40
2.	Pertinence agronomique VS pertinence en télédétection des typologies.....	41
3.	Des données disponibles insuffisantes	42
C.	Des modèles linéaires construits peu explicatifs	44
V.	Bilan et perspectives	45
A.	Conclusion	45
B.	Bilan personnel.....	46
	Table des figures.....	48
	Table des tableaux	49
	Bibliographie	50
	Annexes	53

RESUME

La télédétection offre de nombreuses possibilités pour caractériser la végétation aussi bien par sa composition que par sa structure. Si la capacité de cet outil à caractériser les milieux mono-spécifiques comme les grandes cultures a été montrée à de nombreuses reprises, plus de difficultés sont rencontrées lors de l'étude de milieux pluri-spécifiques comme les prairies. En effet, le mélange des espèces dans un milieu renvoie des valeurs radiométriques difficiles à interpréter. Dans ce contexte, l'objectif de ce stage est de caractériser le mode de gestion et la composition botanique des prairies à partir d'une série temporelle d'images satellites Formosat-2.

Des relevés botaniques et des enquêtes sur le mode de conduite d'une cinquantaine de prairies ont été réalisés lors d'une campagne de terrain. Des typologies botaniques ont été construites avec une approche fonctionnelle de la composition en espèces, qui permet de rendre compte de la valeur d'usage agricole. Les prairies ont ainsi été distinguées selon leur précocité, leur potentiel de productivité, et la richesse en formes de vie (graminées, légumineuses, et diverses). Elles ont aussi été distinguées selon les différents modes de conduite qui ont été identifiés sur le terrain (prairies fauchées une fois, prairies fauchées deux fois, prairies pâturées et prairies mixtes). Des classifications supervisées ont été réalisées sur chacune de ces typologies et différents modèles linéaires ont été construits pour relier directement les taux de recouvrement des formes de vies avec les valeurs radiométriques enregistrées par les images.

Les résultats indiquent qu'il est possible de distinguer les prairies fauchées des prairies mises en pâture avec une précision globale de plus de 80%. En revanche, la distinction des classes botaniques est difficile, notamment en raison du manque d'informations sur des paramètres non contrôlés, et d'un échantillonnage parfois irrégulier. De même, les modèles linéaires construits n'expliquent que très peu la composition botanique à partir des variables spectrales utilisées.

The remote sensing offers many possibilities to determine the vegetation composition and also his structure. If the ability of this tool to characterize mono-specific land like arable crops is well known, problems are encountered when multi-specific lands like meadows are studied. Indeed, the reflected signal of the mix of species is hard to interpret. In this context, the aims of this internship are to characterize the cultural practices of the meadows and their floristic composition through the analysis of a time series of Formosat-2 satellite images.

Botanic survey and management practices surveys have been realized on about fifty meadows during a field tried. Botanic typologies were established, based on functional approach of species composition of the meadows. This approach gives some information on the agronomic use value of the meadow and they were thus distinguished according to their phenology, their potential productivity and their richness in life forms (grasses, legumes, and other dicots). They were also distinguished according to the different management practices that had been identified during the field tried (meadows mowed once, meadows mowed twice, grazed meadows or mix management meadow). Statistical supervised classifications were performed on each typology and linear model were built to predict live forms from spectral variables.

The results show that is possible to differentiate mowed meadows from pastures with an overall accuracy of more than 80%. By contrast, the results are not that good for classifications and modelisation on botanic data. This could be partly explained by the lack of information of non-controlled parameters such as the fertilization practices, and also by irregular sampling.

I. INTRODUCTION

A l'heure actuelle, les changements des paysages agricoles vont vers une diminution des milieux semi-naturels, ce qui entraîne des pertes de biodiversité conséquentes. La prairie a été identifiée comme un élément important pour la fourniture en biodiversité et en services écosystémiques (Le Roux et al., 2008; Robinson and Sutherland, 2002). Or, l'étude de cette fourniture est souvent difficile à réaliser à large échelle car elle nécessite des moyens humains et financiers importants.

La télédétection est prometteuse pour caractériser les ressources que peuvent fournir les infrastructures agro-écologiques (IAE) à l'échelle du territoire. Des projets comme le projet Tel-IAE, auquel fait part le laboratoire DYNAFOR, sont menés pour étudier le potentiel de la télédétection pour identifier et caractériser ces infrastructures.

Ce stage s'inscrit dans ce courant et vise à caractériser le mode de gestion et la composition botanique des prairies à partir de séries temporels d'images satellites.

A. PRESENTATION DE LA STRUCTURE D'ACCUEIL

Mon stage a été effectué au sein de l'Unité Mixte de Recherche DYNAFOR, rattachée à l'Institut National de la Recherche Agronomique (INRA). Cet institut public a été créé en 1946. Il est sous la tutelle du ministère de la recherche et de l'agriculture. L'INRA possède 13 départements pour remplir l'ensemble des missions qui lui sont confiées.

DYNAFOR est une UMR qui a été créée en 2003. Sa mission s'inscrit principalement dans l'axe de recherche n° 6 du centre INRA de Toulouse. Cette unité fédère des chercheurs de 2 départements de l'INRA : SAD (Science pour l'Action et le Développement) et EFPA (Ecologie des Forêts, Prairies et milieux Aquatiques) en plus d'enseignants-chercheurs de l'ENSAT (Ecole Nationale Supérieure Agronomique de Toulouse) intégrés à l'Institut National Polytechnique de Toulouse, et d'enseignants-chercheurs de l'EI Purpan (Ecole d'Ingénieurs de Purpan).

Les activités de DYNAFOR sont centrées sur la gestion durable des ressources forestières et de l'espace rural dans le cadre de l'écologie du paysage. L'objectif principal de DYNAFOR est de comprendre et de modéliser les relations entre des processus écologiques, des processus techniques et des processus socio-économiques dans la gestion de ressources naturelles renouvelables. Il s'agit de faire face aux enjeux actuels dans les espaces ruraux et forestiers, induits par les changements globaux qui concernent conjointement le climat, l'occupation des terres, la biodiversité et les activités humaines. Une finalité de l'Unité est de développer l'ingénierie écologique des territoires ruraux comme moyen d'en assurer leur pérennité et leur capacité à fournir les produits et services attendus par la société.

Les travaux du laboratoire sont répartis selon trois axes principaux :

- l'axe 1 : services écosystémiques du paysage,
- l'axe 2 : données, espace, géomatique et modélisation,
- l'axe 3 : biodiversité des forêts rurales et des milieux naturels.

L'organisation de DYNAFOR est présentée dans la figure suivante (Figure 1) :

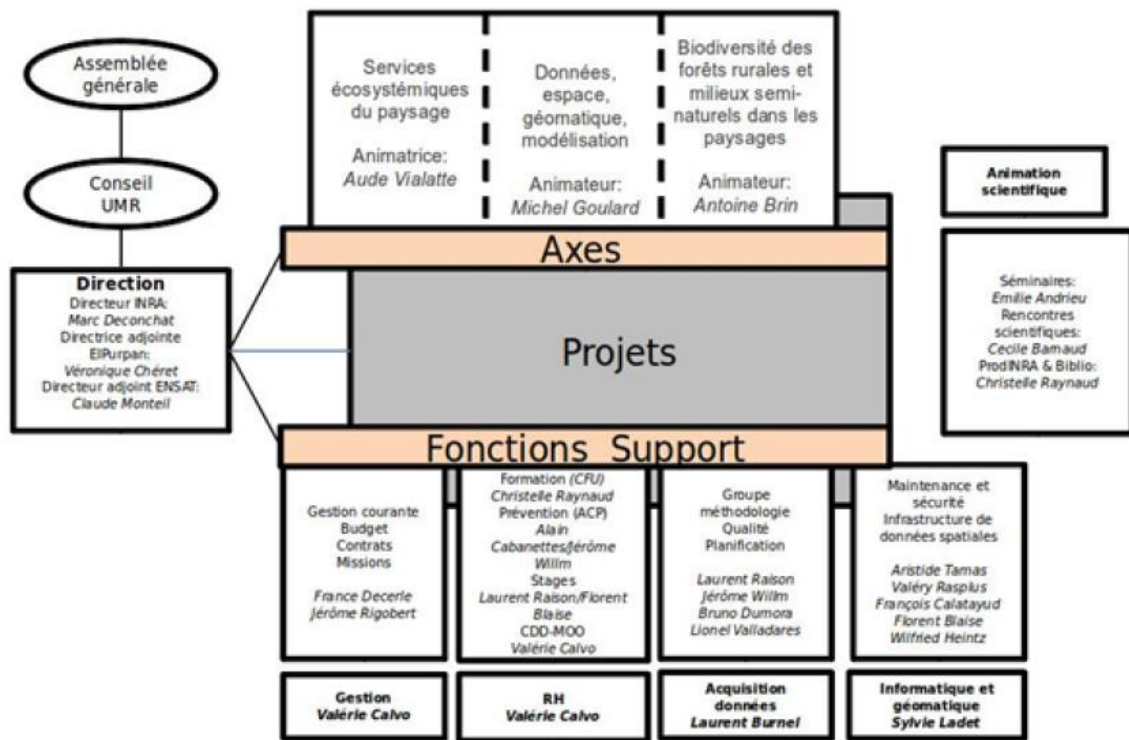


Figure 1: Organigramme du laboratoire DYNAFOR.

Les travaux réalisés pendant ce stage se situent au cœur des problématiques du deuxième axe. J'ai en effet traité des données spatiales que j'ai mobilisé avec des outils géomatiques et de télédétection pour modéliser le comportement spectrale des prairies.

J'ai également eu de nombreuses interactions avec les écologues travaillant dans le premier axe pour réfléchir sur les méthodes de caractérisation de la composition botanique des prairies. De plus, j'ai pu collaborer avec les techniciens du laboratoire, notamment Jérôme Willm, lors de la préparation de la campagne de terrain et pendant son déroulement.

Enfin, ces travaux présentent des perspectives intéressantes pour les chercheurs du premier axe qui pourront à terme utiliser les résultats de ce stage pour estimer les services écosystémiques des prairies.

Des interactions avec des organismes externes au laboratoire ont également eu lieu. Jean-François Dejoux a suivi avec intérêt le déroulement de ce stage. L'unité mixte de recherche dans laquelle il travaille, le CESBIO (Centre d'études spatiales de la biosphère), étudie le fonctionnement et la dynamique de la biosphère continentale à différentes échelles spatiales et temporelles, et notamment les cultures agricoles.

Des échanges ont également eu lieu avec la chambre d'agriculture de la Haute-Garonne sur des questions de caractérisation des exploitations agricoles de la zone d'étude.

B. ETAT DE L'ART ET OBJECTIFS

1. CARACTERISATION DES PRAIRIES

a. Contexte

Une prairie peut être définie de manière générale comme une surface agricole destinée à la culture de plantes fourragères, utilisées pour l'alimentation des animaux. Depuis les dernières décennies, le pourcentage de surfaces occupées par les prairies est en constante diminution en Europe et en France (Figure 2). Celle-ci peut s'expliquer en partie par l'évolution des modalités de l'activité d'élevage comme l'intensification des exploitations ou le manque de soutien par rapport à d'autres productions de la part des politiques publiques. De plus, ces surfaces sont généralement considérées comme peu productives et peu rentables économiquement (Baumont et al., 2011), et ce notamment à cause d'un manque de savoir-faire technique et de connaissance sur le fonctionnement des prairies.

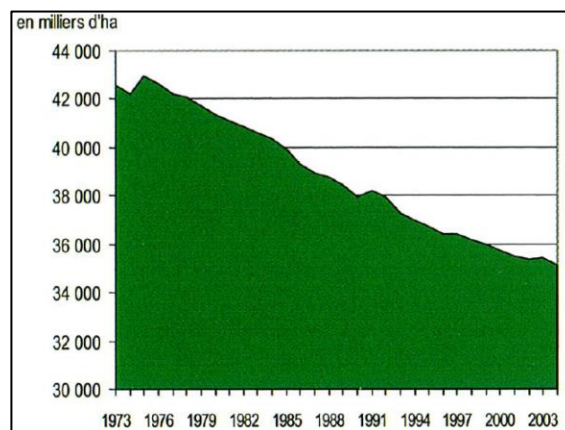


Figure 2 : Evolution des surfaces toujours en herbe de l'Union européenne des 9 depuis les années 1970 (Baumont et al., 2011).

Les enjeux liés à leur conservation et leur bonne gestion sont pourtant de taille, ne serait-ce que par la part toujours importante qu'elle représente aujourd'hui dans la SAU (surface agricole utile), qui est d'environ 45%. En effet, les prairies rendent de nombreux services. Elles jouent un rôle important sur le plan économique, notamment dans la gestion des stocks (Vignau-Loustau and Huyghe, 2008), mais aussi à travers des services écosystémiques comme les services d'approvisionnement (ressources génétiques, aliments et matières premières), de régulation (climat, eau, maladies), les services culturels (récréation, patrimoniale, valeurs esthétiques) ou encore de soutien (formation et rétention des sols, offre d'habitats, cycle de l'eau)(Mauchamp et al., 2012).

Les prairies apparaissent donc comme une unité de production agricole aux enjeux multiples qu'il convient de gérer de manière optimale, aussi bien pour le bénéfice des agriculteurs qui les possèdent et qui les exploitent que pour l'intérêt général à travers les différents services qu'elles rendent. Afin d'avoir une bonne gestion de ces prairies, il est nécessaire d'avoir une connaissance poussée de leur fonctionnement et de leur environnement. Or les prairies ne forment pas une réalité homogène et correspondent à une grande diversité de situations selon les pratiques exercées et les milieux dans lesquels elles se trouvent.

b. Typologies et diagnostics des prairies

Il est possible de caractériser les prairies selon plusieurs critères et selon plusieurs objectifs. Un exemple de typologie simple est celle utilisée pour la PAC (Politique Agricole Commune), établie en 1998 par le SCEES (Service Central des Études et Enquêtes Statistiques du ministère de l'Agriculture)(Dusseux, 2014). Basée sur la composition botanique et sur leur mode de gestion,

elle distingue trois types de prairies qui sont : les prairies permanentes, les prairies temporaires et les prairies artificielles. Les dernières sont caractérisées par une composition à plus de 80% en légumineuses. Les deux autres se distinguent uniquement par un ensemencement effectué il y a plus ou moins de 5 ans (Agreste). Cette typologie se révèle très générale et chaque type de prairie correspond encore à une grande diversité de pratiques et de composition.

D'autres typologies existent et se basent sur les critères suivants : les conditions pédoclimatiques (pluviométrie, nutrition minérale, acidité ...), la composition botanique, le mode de gestion (fauche, pâture, mixte). Il est également possible de les caractériser selon leur valeur d'usage, c'est-à-dire la capacité des prairies à remplir une fonction, celle-ci pouvant être différentes selon l'objectif en vue. On peut s'intéresser à la valeur d'usage environnementale comme par exemple le maintien des espèces pollinisatrices si l'approche des prairies envisagée est écologique mais on peut également s'intéresser à la valeur d'usage agricole comme la capacité des prairies à fournir un fourrage de qualité et/ou en quantité dans le cas d'une approche agronomique.

L'ensemble de ces entrées sont étroitement liées (Figure 3), et il est possible d'arriver à un même diagnostic, agricole et/ou environnemental avec différentes approches. Il existe par exemple de nombreuses relations entre le couple pratiques et conditions environnementales et la composition botanique des prairies (Duru et al., 2007; Hubert and Pierre, 2003; René Baumont et al., 2011). Ainsi, la valeur d'usage agricole peut être déterminée grâce à une typologie basée sur les pratiques et les conditions du milieu (Plantureux et al., 1992) mais également par une analyse de sa composition botanique.

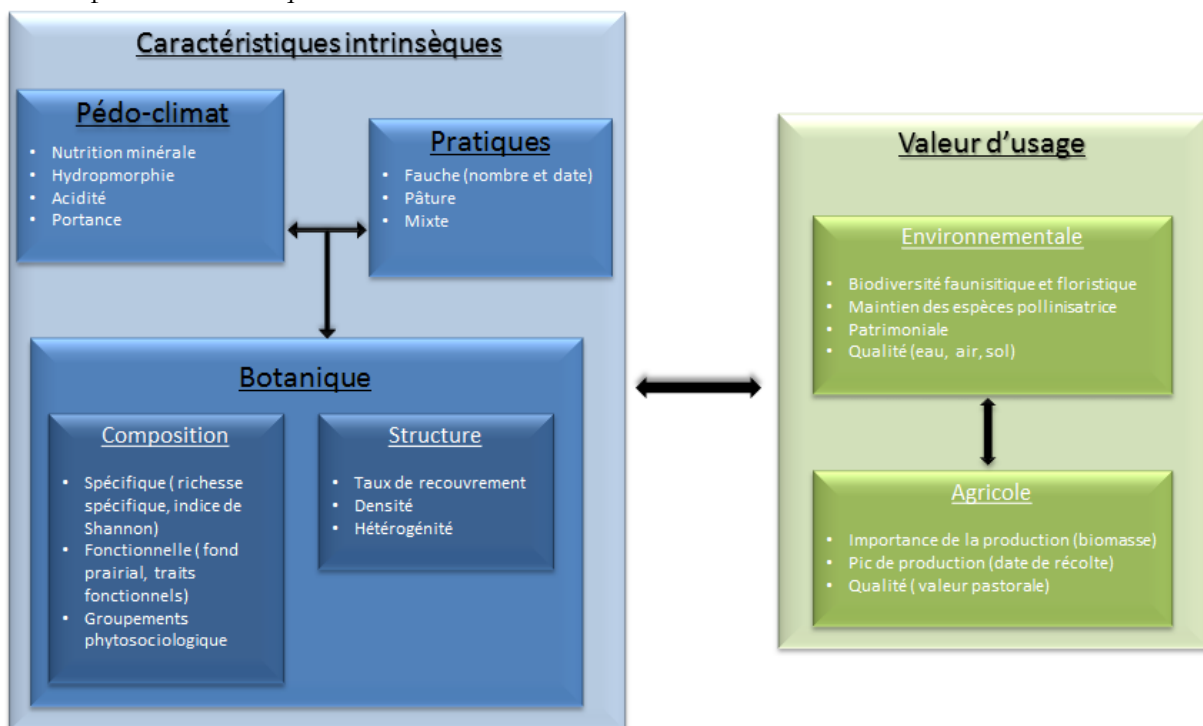


Figure 3 : Interdépendance des entrées possibles pour caractériser une prairie.

Il se pose ainsi la question de la méthodologie à mettre en place pour la réalisation du diagnostic, qui est un compromis entre précision de ce dernier et facilité de mise œuvre en termes de coût et de connaissances à mobiliser. Il est certes possible de réaliser des analyses en laboratoire sur la composition du sol et de déterminer ainsi la productivité potentielle d'une parcelle mais ce type de démarche est onéreux et fastidieux. D'autres typologies se basent sur des

critères taxonomiques au travers d'indices spécifiques de la qualité fourragère ou d'indices zootechniques des espèces. Mais ces approches n'ont généralement qu'une validité locale et sont difficilement généralisables (Duru et al., 2007).

Depuis quelques années, des méthodes reposant sur l'écologie fonctionnelle ont été développées. On ne cherche plus à regrouper les espèces végétales des prairies selon leur taxonomie ou leur appartenance à des groupes phytosociologiques, mais sur la base de *traits fonctionnels*, dont la valeur est standardisée. Un trait fonctionnel est une propriété physiologique, morphologique ou phénologique d'une espèce qui a un impact sur sa valeur sélective ou *fitness* (Violle et al., 2007). Ces traits traduisent la réponse des espèces à des facteurs environnementaux (disponibilité des nutriments ...) et aux pratiques agricoles (modalité de fauche et de pâture), et permettent de déterminer leur valeur d'usage potentielle agronomique et environnementale. Ainsi les propriétés d'une parcelle sont dépendantes des espèces rassemblées sous un ou plusieurs traits fonctionnels (Cruz et al., 2010; Theau et al., 2010).

Il n'est pas nécessaire d'identifier toutes les espèces présentes d'une parcelle pour parvenir à un diagnostic avec une approche fonctionnelle. Ce serait en effet trop difficile à mettre en œuvre par les acteurs du développement comme les conseillers en chambre d'agriculture, aussi bien en termes de temps de réalisation qu'en niveau de compétence exigé.

Les relevés peuvent se limiter aux espèces abondantes, qui constituent ce qu'on appelle le *fond prairial*, qui correspond aux espèces représentant 85% de l'abondance relatives (Vivier, 1971). Cruz et al. (2010), ont ainsi regroupé 38 espèces de graminées représentatives de plusieurs régions herbagères de France en 6 types fonctionnels de graminées (TFG), basé sur l'étude de 6 traits fonctionnels : le taux de matière sèche, la surface spécifique foliaire, la durée de vie des feuilles, la date de floraison, la hauteur maximale de la plante, et la résistance des feuilles à la cassure. L'affectation d'une prairie à un ou deux TFG permet d'avoir une bonne idée de la valeur d'usage agronomique potentielle.

2. LA TELEDETECTION : UN OUTIL ADAPTE A L'ETUDE DES PRAIRIES A LARGE ECHELLE ?

Cette partie de l'état de l'art a été réalisée en s'inspirant de la thèse de Pauline Dusseux (2014), qui a travaillé sur le suivi des prairies en milieu agricole à partir de séries temporelles d'images satellites à haute résolution.

a. Notions générales

La télédétection permet l'étude des surfaces terrestres de manière indirecte via l'acquisition par des capteurs adaptés du rayonnement électromagnétique qu'elles émettent ou réfléchissent. Elle peut être passive ou active, selon si le signal acquis a été émis depuis une source naturelle comme le soleil ou s'il est émis depuis une source artificielle comme des capteurs RADAR.

Les capteurs peuvent être caractérisés selon leur résolution spatiale, leur résolution spectrale et leur résolution temporelle.

La résolution spatiale définit la taille minimale des objets qui peuvent être détectés et se traduit par la taille, au sol, des pixels d'une image. Elle doit être adaptée à l'objet d'étude et diffère selon les applications. Elle peut varier du kilomètre pour les satellites de basse résolution (ex. AVHRR) pour des applications comme la météorologie ou l'étude du couvert forestier à l'échelle continentale (Seebach et al., 2011), à l'ordre du mètre pour des satellites à haute ou très haute résolution (SPOPT-4, Formosat-2, Pleiades ...).

La résolution temporelle est la capacité de revisite d'un capteur sur une même zone. Il convient de nuancer cette définition pour les capteurs optiques. Selon les conditions atmosphériques (couverture nuageuse), le nombre d'images exploitables peut parfois être beaucoup moins important que le nombre d'images acquises.

La résolution spectrale quant à elle définit la capacité d'un satellite à distinguer deux longueurs d'onde voisines. Elle se traduit par le nombre de bandes spectrales d'une image. Trois catégories d'images peuvent être définies : les images panchromatiques (une bande en N&B), les images multi-spectrales (de 3 à 5 bandes environ), et les images hyper-spectrales (plus de 100 bandes). Plus la résolution spectrale est importante, plus la signature spectrale d'un objet est précise et plus ce dernier est susceptible d'être bien discriminé.

Les bandes du rouge (de 600nm à 700nm) et du proche infra-rouge (700nm à 1500nm) sont les plus couramment utilisées pour caractériser la végétation. Plusieurs études ont par exemple montré que ces bandes étaient optimales pour étudier la composition floristique (Feilhauer et al., 2013) ou d'autres variables comme la biomasse (Fava et al., 2010). A partir de ces bandes, il est également possible de créer des indices de végétation, ou d'estimer des variables biophysiques. Parmi tous les indices spectraux, le NDVI (Normalized Difference Vegetation Index) est celui qui est le plus couramment utilisé. Il peut en effet être corrélé à l'activité photosynthétique des plantes (Homolová et al., 2013) et à la quantité de biomasse (Dusseux et al., 2014; Girard and Girard, 1999; Caloz and Collet, 2001). Son évolution au cours d'une année est ainsi facilement interprétable et traduit le cycle phénologique de la végétation. Il est néanmoins sensible aux effets de l'atmosphère, du sol, ou du taux de recouvrement de la végétation (Gitelson et al., 2002).

b. Télédétection et prairies

La télédétection permet l'étude et le suivi à large échelle de la végétation, comme les forêts, les cultures ou encore les prairies (Boyd and Danson, 2005). Elle constitue ainsi un outil d'aide à la décision pour les acteurs des politiques de conservation et d'aménagement du territoire.

S'il est possible d'établir des relations entre des compositions floristiques similaires et différentes phénologies des surfaces herbacées sur de très larges échelles (plusieurs centaines de kilomètres) (Viña et al., 2012), la cartographie des prairies dans un milieu agricole hétérogène et fragmenté ne peut s'envisager qu'avec une haute résolution spatiale (Lecerf, 2008) voire une très haute résolution spatiale (Franke et al., 2012).

La plupart des études sur les prairies en milieu agricole visent à les discriminer des autres types de couvert végétal ou d'estimer des paramètres descriptifs du couvert tels que la biomasse ou le taux de recouvrement (Fava et al., 2010). Avec une faible résolution temporelle intra-annuelle (de une à trois dates), il est en effet possible d'identifier et de localiser les prairies sans les différencier entre elles (Cingolani et al., 2004). Avec des données sur plusieurs années (résolution inter-annuelle), il est possible de déterminer la place des prairies dans les successions culturales et

leur durée (Vannier, 2012). Cependant, les informations issues de telles données satellitaires restent imprécises, avec notamment des confusions avec les cultures d'hiver qui ont un comportement spectral similaire. La répétitivité temporelle limitée des capteurs apparait ainsi comme une contrainte pour une détection fiable des prairies et *a fortiori* la caractérisation de leur mode de conduite et de leur composition botanique.

Des études à partir de séries temporelles plus importantes ou d'images à haute résolution spatiale ont permis de caractériser le mode de gestion des prairies. C'est le cas de Mansour et al. (2012) qui ont déterminé l'état de dégradation des pâtures en Afrique du Sud à partir de l'abondance d'espèces indicatrices et données hyperspectrales. Il semble également possible de détecter les dates précises de fauche à partir de méthodes statistiques de détection de changement (Hadj Said et al., 2011) ou même de distinguer les prairies fauchées, pâturées ou en gestion mixte (Dusseux et al., 2011) à partir de séries temporelles à haute résolution spatiale. En distinguant les prairies par leur intensité d'utilisation, c'est-à-dire par le nombre et la quantité des prélèvements de biomasse par fauche ou pâture, Franke et al. (2012) obtiennent également de bons résultats de classification avec une précision globale de 85,7%.

D'autres études ont montré la capacité de la télédétection à caractériser le couvert végétal des prairies et leur composition botanique. Il est par exemple possible de déterminer la surface spécifique foliaire, la biomasse ou la masse surfacique foliaire (Fava et al., 2010; Homolová et al., 2014). L'estimation de ces paramètres, fondée principalement sur l'indice de végétation NDVI peut cependant être perturbée par les conditions atmosphériques ou le recouvrement floral des prairies (Shen et al., 2010). A partir de différents indices de végétation dont le NDVI, Knox et al. (2013) ont montré qu'il était possible de déterminer le stade phénologique des plantes d'une prairie, stades phénologiques qu'il est possible de mettre en lien avec des traits d'espèces comme la date de floraison (Homolová et al., 2013)

D'autres ont pu estimer des indicateurs de la biodiversité tels que la richesse spécifique (Homolová et al., 2014) ou l'indice de Shannon (Fava et al., 2010), qui permet de quantifier l'hétérogénéité de la diversité taxonomique d'un milieu. Enfin, Feilhauer et al. (2013), ont pu établir des gradients floristiques sur des prairies en Allemagne. Même si ces études ont été faites à partir de données hyper-spectrales, elles constituent des résultats encourageants pour la caractérisation botanique des prairies avec des images satellites.

3. BILAN DE L'ETAT DE L'ART

Peu d'études sur la composition botanique des prairies à partir d'images satellites ont été réalisées. Mais des approches avec des données hyper-spectrales, obtenues à partir de spectromètres de terrain ou de vols avec capteurs embarqués, obtiennent des résultats intéressants. Cependant, l'obtention de ces données est très coûteuse et celles-ci ne permettent pas d'étude du paysage à une large échelle.

L'apport des données à haute résolution spatiale et temporelle n'a pas encore été étudié et semble une alternative prometteuse aux données hyper-spectrales, notamment avec une approche basée sur les traits fonctionnels, novatrice dans la caractérisation de la composition botanique par télédétection.

En effet, parmi toutes les méthodes de caractérisation, l'approche fonctionnelle semble la plus intégrative de toutes les entrées possible pour caractériser une prairie. Elle est en outre la plus

facile à mettre en œuvre puisqu'elle se base sur la notion de fond prairial (ou recouvrement majoritaire de la prairie), ce qui est donc le plus susceptible d'être détecté par télédétection.

S'il a été montré que des études ont été menées à partir de quelques dates sur la caractérisation du mode de conduite, aucune n'a été faite avec des séries d'images très fournies. La disponibilité d'un grand nombre d'images peut en effet être une piste intéressante pour caractériser finement le type des pratiques, c'est-à-dire déterminer le nombre et la date des interventions ainsi que leur nature, là où les études actuelles se limitent à distinguer les prairies fauchées des pâtures.

4. OBJECTIFS ET HYPOTHESES DU STAGE

L'objectif de ce stage, est d'étudier les possibilités de la télédétection pour caractériser les prairies, aussi bien par leur mode de conduite que par leur composition botanique, à partir d'une série temporelle de 17 images satellites à haute résolution spatiale FORMOSAT-2. Les études sur les prairies à partir de telles séries d'images en sont encore à leur début et des méthodes de traitement restent à développer.

Un premier objectif de ce stage est de savoir s'il est possible, d'une part de distinguer les prairies fauchées, les prairies pâturées et les prairies en gestion mixte ; et d'autre part, de distinguer le nombre d'interventions (fauche ou pâture) et leur date.

Nous faisons l'hypothèse que **les pratiques culturales, entraînant des prélèvements de biomasse plus ou moins importants sur les prairies, vont pouvoir être différenciées en étudiant l'évolution du NDVI au cours de l'année** (le NDVI étant corrélé à la biomasse bien qu'il traduit d'abord le niveau d'activité chlorophyllienne de la végétation). Ainsi, on suppose **qu'une fauche peut être distinguée par une diminution de cet indice plus rapide et plus importante que celle d'une prairie mise en pâture** (pour laquelle les prélèvements sont moins importants ou plus espacés dans le temps).

Bien que l'utilisation du NDVI puisse être sujette à certaines limites (saturation du signal en cas de biomasse importante par exemple), il reste un des indices les plus utilisés dans les études et est facilement interprétable. Pour ces raisons et parce que ce stage constitue une première étude exploratoire sur les prairies au sein de DYNAFOR, cet indice a ainsi été retenu.

Le deuxième objectif de ce stage est ainsi d'étudier la possibilité ou non de distinguer les prairies selon leur composition botanique fonctionnel. Les groupes fonctionnels de graminées établis par Cruz et al. (2010), vont permettre d'établir des types de précocité de la phénologie et de productivité potentielle. Les hypothèses qu'on cherche à vérifier sont que **les prairies précoces se distinguent par un maximum de NDVI apparaissant plus tôt que celui des prairies tardives** et que **les prairies au fort potentiel de productivité ont un maximum de NDVI plus élevé que les prairies au faible potentiel**.

Les différentes variables par lesquelles on souhaite caractériser le mode de gestion et la composition botanique sont résumées dans le Tableau 1. Les variables spectrales associées, sur lesquelles des hypothèses ont été formulées, y figurent également.

(a)	Mode de gestion		
Variable thématique à caractériser	Nombre d'interventions	Date des interventions	Nature de l'intervention
Variable spectrale à mobiliser	Nombre de diminution du NDVI	Début des diminutions	Amplitudes des diminutions
(b)	Composition botanique		
Variable thématique à caractériser	Précocité	Productivité	
Variable spectrale à mobiliser	Date du maximum du NDVI	Maximum du NDVI	

Tableau 1 : Synthèse des variables de gestion (a) et botanique (b) à déterminer et les variables spectrales associées.

5. APPORT DU PRECEDENT RAPPORT DE STAGE

En vue de l'obtention du Master Recherche en géomatique, un rapport de stage (Lang, 2015) a été rédigé en Juin 2015. Dans ce travail passé, une première étape a été de rassembler les prairies en se basant uniquement sur leurs caractéristiques spectrales communes, sans connaissance thématique *a priori*.

Le regroupement a été en fait en utilisant deux algorithmes de classification non supervisées communs : la méthode de K-means et la méthode de GMM. De nombreuses itérations ont été réalisées pour chaque méthode, et les deux algorithmes ont montré des résultats similaires quant au nombre de groupes spectraux isolés et à leurs caractéristiques.

Les profils temporels moyens du NDVI de ces groupes ont ensuite été analysés et les différents groupes de prairies trouvés peuvent systématiquement se rassembler en deux types distincts :

- Les groupes pour lesquelles une seule baisse de NDVI est observée entre la période de début mai à mi-septembre, période pendant laquelle des interventions de fauche ou de pâture sont susceptibles d'avoir lieu (voir Figure 4.b)
- Les groupes pour lesquels deux baisses du NDVI sont observées (Figure 4.a)

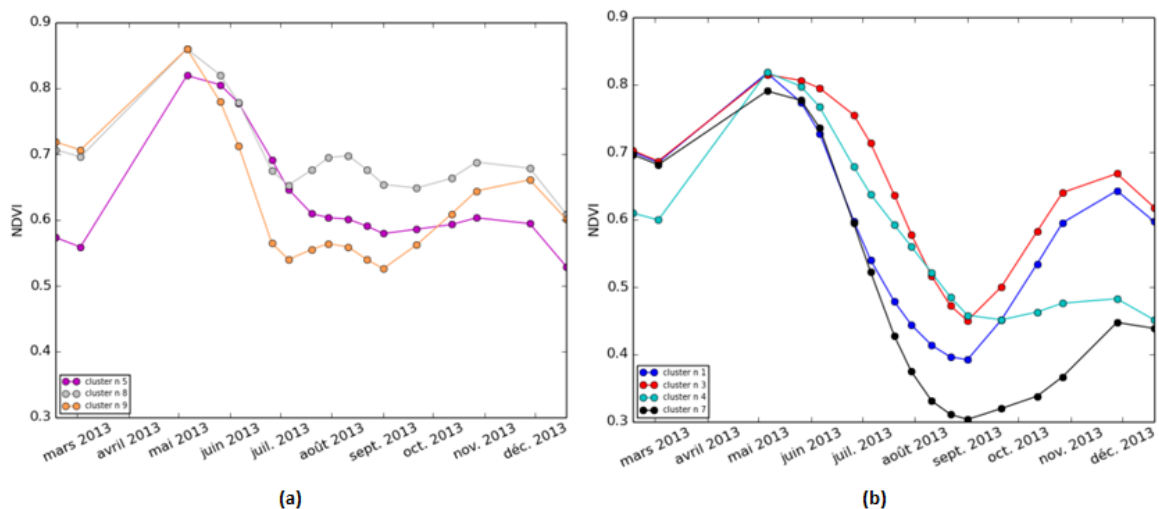


Figure 4 : Profils moyens de NDVI des groupes issus d'une classification non supervisée, caractérisés par deux diminutions (a) et par une diminution (b) (Lang, 2015).

La classification non supervisée apparaît être alors un outil prometteur pour déterminer le nombre d'intervention ayant eu lieu sur une prairie, en ne permettant toutefois pas d'en déterminer la nature. En effet, la comparaison de ces profils avec les données terrains collectées sur les modes de conduite n'a pas permis d'analogie entre les groupes issus des classifications non supervisées et un mode de gestion. C'est-à-dire que le profil moyen de NDVI des prairies fauchées une fois (et des autres modes de conduite) ne correspondait à aucun des profils des groupes de issu de classification.

Il n'a également pas été possible de déterminer des dates de fauches car les diminutions de NDVI observées (Figure 4) ont lieu sur de très longues périodes (de deux à quatre mois) alors que la diminution provoqué par une fauche est supposée être forte et sur une courte période de temps (chute brutale).

Dès lors, on peut se demander si les baisses de l'activité photosynthétique identifiées des profils ne sont pas la conséquence d'un effet combiné entre une intervention humaine et la sénescence des espèces végétales, liée à un manque d'eau, un ensoleillement trop important ou simplement à la fin de leur cycle phénologique.

Suite à ces premières conclusions plusieurs propositions ont été faites pour poursuivre les travaux sur les modes de conduites.

Une première approche basée sur l'utilisation de descripteurs des dynamiques de NDVI (date du maximum, pente entre deux dates etc.) va être étudiée par Maïlys Lopes, doctorante du laboratoire DYNAFOR, pour essayer de déterminer pour chaque prairie, la ou les dates d'intervention ainsi que leur nature.

Une autre possibilité pour déterminer le mode de gestion est la mise en œuvre de classifications supervisées, basées sur la connaissance du mode de conduite de certaines prairies. C'est cette dernière approche qui sera développée dans ce rapport.

Par ailleurs, les données botaniques n'étant pas encore analysées lors de la rédaction du précédent rapport, l'étude de la caractérisation des prairies par leur composition botanique n'avait pas été faite. Elle sera également développée ici, avec la mise en œuvre de classifications supervisées sur différentes typologies basées sur la composition fonctionnelles des prairies. Des modèles linéaires ont également été construits à titre exploratoire pour essayer d'expliquer directement les taux de recouvrement des différentes formes de vie et types fonctionnels de graminées.

II. MATERIEL ET METHODES

A. PRESENTATION DE LA ZONE D'ETUDE

La zone d'étude se situe pour sa grande majorité dans le département de Haute-Garonne, et pour une minorité dans le Gers. Elle se trouve dans la région de St-Lys et Rieume, au sud-ouest de Toulouse (Figure 5). D'une superficie d'environ 44000 ha, elle ne correspond à aucune entité géographique ou graphique particulière sinon aux limites des images disponibles. La superficie des surfaces déclarées en 2013 pour le registre parcellaire graphique (RPG) représente environ 55% de la surface totale. 11% de ces surfaces sont des prairies, dont 84% de prairies temporaires et 16% de prairies permanentes, alors qu'à l'échelle de la région Midi-Pyrénées ces surfaces sont respectivement de 45%, 34% et 66% en 2010 (Enquête "Utilisation du territoire - Teruti-Lucas" ; Agreste - Recensements agricoles 2010).

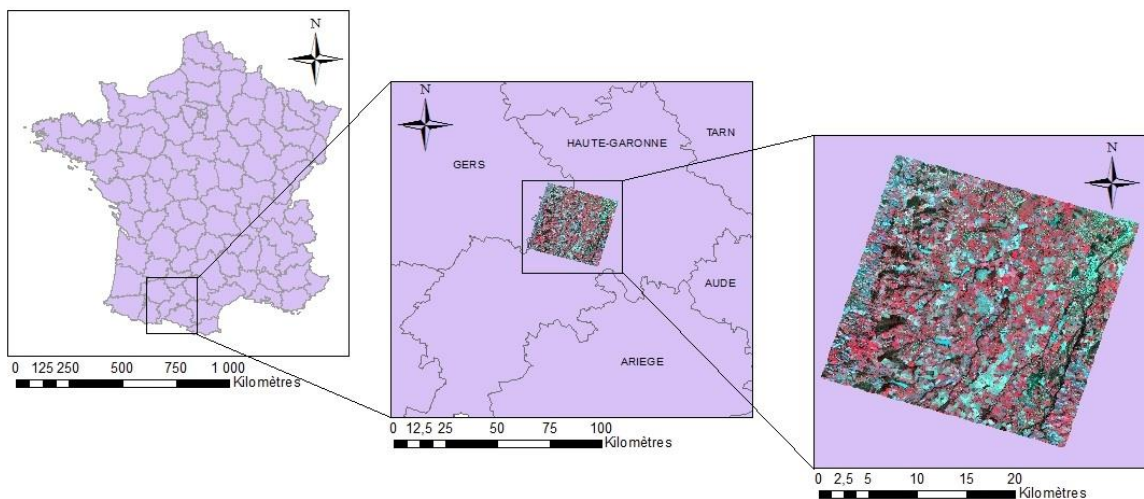


Figure 5 : Localisation du site d'étude et image Formosat-2 du 03/03/2013, fausses couleurs (RGB, PIR, R, V).

L'élevage n'est pas très présent dans cette zone, et est même en régression. Certains agriculteurs arrêtent leur activité d'élevage au profit de l'exploitation de grandes cultures (maïs, tournesol, ...). De plus, une partie des prairies déclarées ne sont en fait que des réserves foncières qui sont simplement entretenues mais pas réellement exploitées, en particulier dans la partie nord-est de la zone, située à la limite de la métropole toulousaine. On peut noter cependant un nombre plus important d'éleveurs dans la partie sud-ouest de la zone, à proximité de la région des coteaux de Gascogne, région où l'élevage est beaucoup plus présent.

Ce site d'étude n'est pas représentatif de la diversité des pratiques suivies dans la région mais est une zone test, sur laquelle de nombreuses données sont disponibles. Bien qu'elle ne soit pas familière au laboratoire DYNAFOR, elle permet de tester le potentiel de la télédétection et de développer des méthodes pour caractériser les prairies selon leur mode de conduite et leur composition botanique à partir de séries temporelles fournies. Si le CESBIO est plus familier de la zone, les prairies ne font pas partie de leur objet d'étude. Les recherches méthodologiques menées sur le site et se focalisant sur les prairies sont donc les premières.

B. PRESENTATION DE LA DEMARCHE GENERALE

La démarche suivie lors de ce stage est représenté sur la Figure 6 ci-dessous :

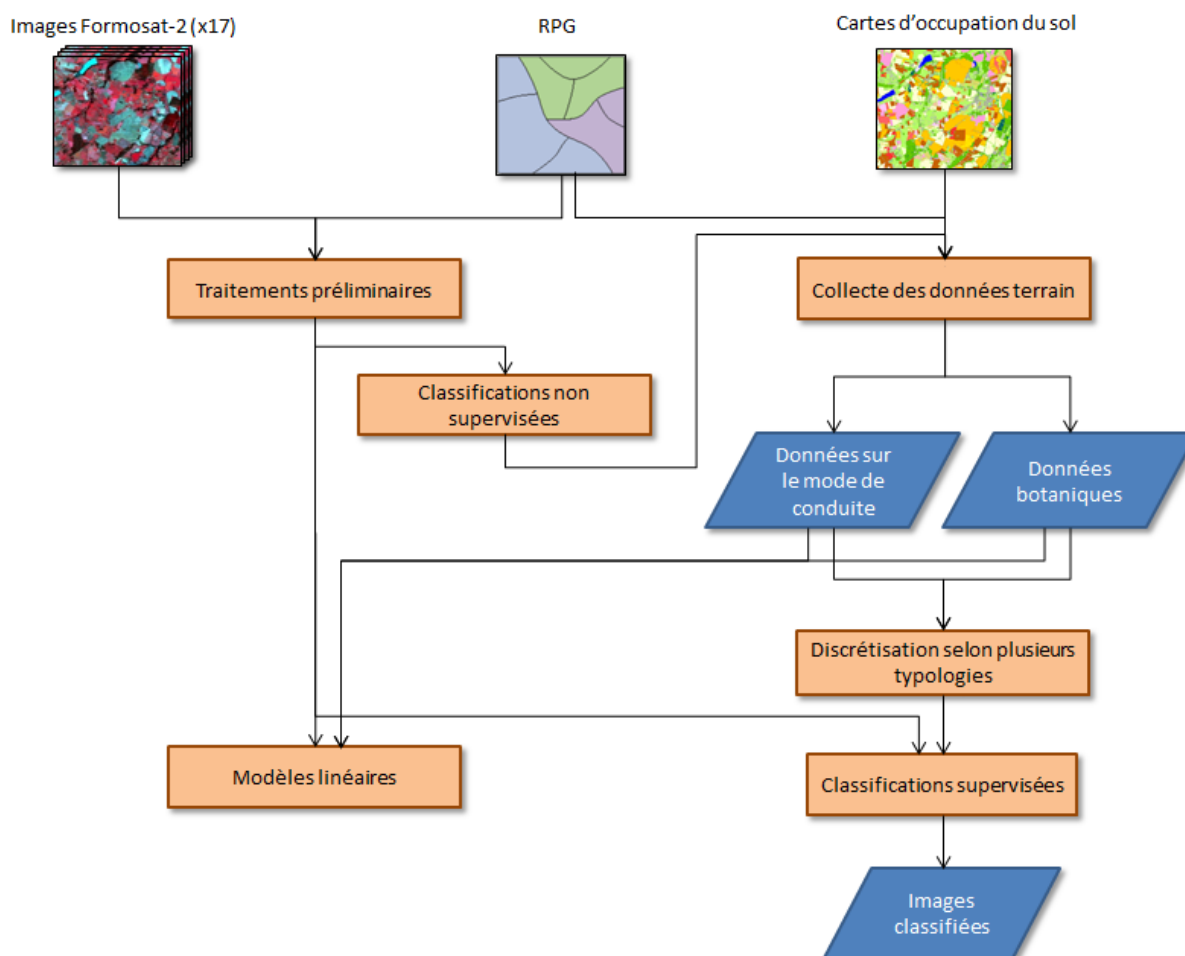


Figure 6 : Schéma fonctionnel de la démarche générale.

Une partie de traitements préliminaires des données a été effectuée afin de pouvoir réaliser en premier lieu des classifications non supervisées qui ont font l'objet de l'analyse décrite dans le rapport précédent (Lang, 2015). Ces classifications ont également fourni des informations aidant à sélectionner les prairies sur lesquelles des enquêtes du mode gestion et des relevés botaniques ont été réalisés lors d'une campagne de terrain.

Les données récoltées ont ensuite permis de répartir les prairies en différentes classes, selon différentes typologies et réaliser des classifications supervisées que nous détaillerons par la suite. Des modèles linéaires ont également été construits pour essayer d'expliquer directement les taux de recouvrement des formes de vie et des types fonctionnels de graminées.

C. DONNEES SATELLITAIRES

1. CARACTERISTIQUES GENERALES

Les images disponibles sont issues du satellite Formosat-2 lancé en mai 2004 et qui possède la particularité d'avoir une orbite héliosynchrone, ce qui permet une acquisition des images sous le même angle (Site d'Airbus Defense and Space, 10 juin 2015). Il est le seul satellite de haute résolution doté d'une capacité de revisite quotidienne, ce qui permet l'obtention de série temporelle à très haute fréquence (Site du CESBIO, 10 juin 2015). Le satellite fournit des images multi-spectrales à 4 bandes et une image panchromatique en noir et blanc. Les caractéristiques des capteurs du Formosat-2 sont présentées dans le tableau ci-dessous (Tableau 2).

N° canal	Couleur	Bande spectrale	Résolution
B1	Bleu (B)	0,45 – 0,52 μm	8x8m
B2	Vert (V)	0,52 – 0,60 μm	8x8m
B3	Rouge (R)	0,63 – 0,69 μm	8x8m
B4	Proche Infra Rouge (PIR)	0,76 – 0,90 μm	8x8m
P	noir et blanc	0,45 – 0,90 μm	2x2m

Tableau 2 : Caractéristiques des capteurs du satellite Formosat-2, (Site d'Airbus Defense and Space, 10 juin 2015).

2. IMAGES DISPONIBLES ET SELECTION DES DONNEES

Dans le cadre de la préparation de la mission Sentinel-2 (ESA) ainsi que la mission VEN μ S portée par l'UMR CESBIO, le laboratoire dispose de séries temporelles sur la zone d'études pour les années 2006 à 2014, avec 15 à 40 dates par an. Certaines images acquises présentent une couverture nuageuse plus ou moins importante (Figure 7). Lorsque celle-ci est très importante, la quantité de pixels exploitables n'est plus suffisante pour pouvoir utiliser l'image. On trouve aussi des zones pour lesquelles la valeur du signal enregistrée est saturée, notamment pour les surfaces en eau qui renvoie la lumière du soleil par effet de réflexion.

Le CESBIO, dans le cadre du dispositif Kalideos OSR MiPy coordonné par le CNES (Centre National d'Etudes Spatiales), nous a fourni les séries d'images (multi-spectrales) pré-traitées pour les années de 2007 à 2014. Celles-ci sont fournies avec un masque de nuages pour chaque date, réalisée grâce à une méthode de correction des effets atmosphériques développée par Hagolle et al. (2008). L'étude présentée dans ce rapport a été effectuée sur l'année 2013 et les dates disponibles sont présentées dans le tableau ci-dessous (Tableau 3).

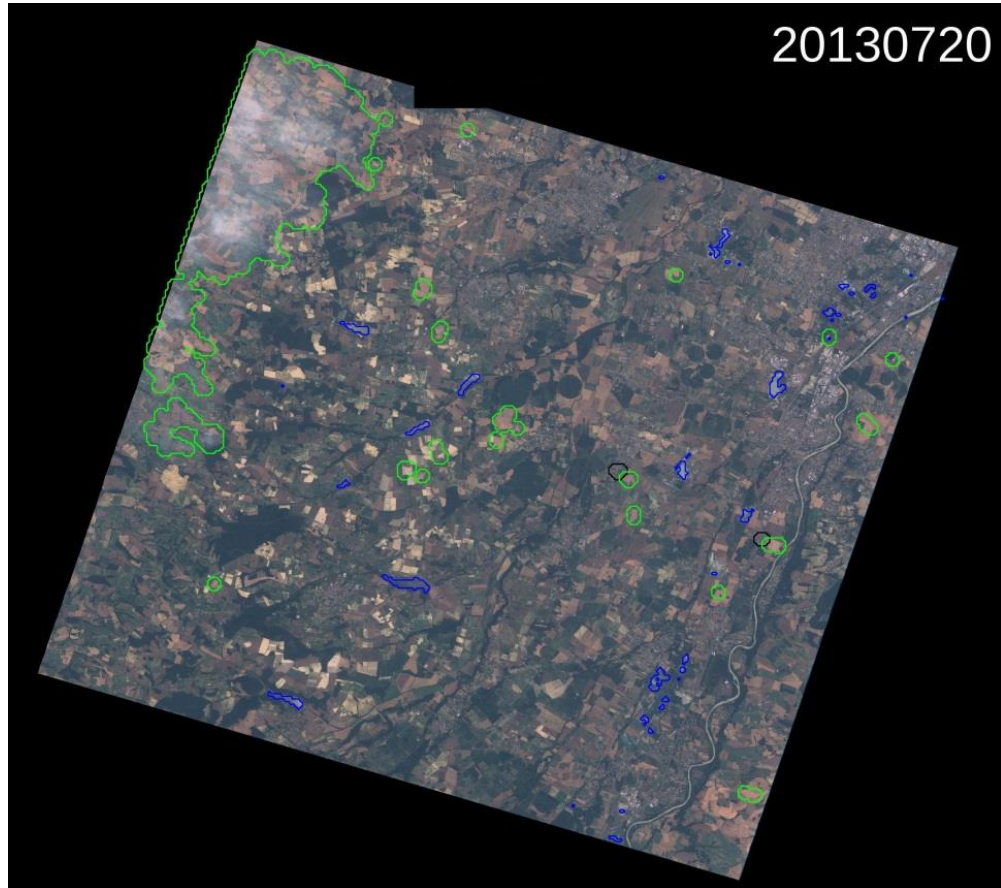


Figure 7 : Exemple de présence de nuages et de leur ombre (entourés en vert et en noir) et de saturation (entourées en bleu) sur une image de la zone d'étude (20/07/2014).

Date d'acquisition	16/02	03/03	03/05	26/05	06/06	26/06	06/07	20/07	30/07
Présence de nuages			X	X	X	X	X	X	
Date d'acquisition	11/08	22/08	01/09	21/09	12/10	27/10	28/11	20/12	
Présence de nuages	X	X	X	X	X	X		X	

Tableau 3 : Dates disponibles pour l'année 2013 et présence de nuages.

D. DONNEES AUXILLIAIRES

1. CARTE DES OCCUPATIONS DU SOL

Le CESBIO, dans le cadre de ses études, réalise chaque année une carte d'occupation du sol de la zone d'étude à partir de classification d'images des satellites SPOT-4 et Formosat-2. Les différentes cultures y sont représentées sur ces cartes, ainsi que les prairies, avec une précision annoncée d'environ 90% pour l'année 2013 (taux de pixels bien classés, d'après des documents internes du CESBIO). Le laboratoire nous a fourni les classifications pour les années de 2006 à 2013 qui serviront dans le cas de cette étude à comparer qualitativement les résultats obtenus. La carte d'occupation des sols de 2013 est présentée en ANNEXE 1 à titre d'exemple.

2. LE REGISTRE PARCELLAIRE GRAPHIQUE (RPG)

Le Registre Parcellaire Graphique a été mis en place dans le cadre de la PAC pour permettre la déclaration et l'identification des îlots agricoles de chaque exploitant auprès de l'ASP (Agence de Service et de Paiement). Les données sont créées par les exploitants agricoles sur la base de la BD-ORTHO de l'IGN (Institut National de l'Information Géographique et Forestière) à une échelle de précision de l'ordre du 1/50 000ème. Elles sont disponibles gratuitement sur la plateforme ouverte des données publiques française (data.gouv.fr).

Une version déjà traitée a été fournie par le CESBIO sous forme de couche vectorielle et qui comprend la zone d'étude. Chaque entité de la couche correspond à un îlot agricole, qui peut correspondre à plusieurs parcelles de cultures différentes. Si un îlot est constitué de plusieurs parcelles de cultures différentes, ces dernières sont spécifiées par ordre décroissant de surface. Les caractéristiques des champs attributaires de la couche sont présentées dans le Tableau 4 ci-dessous :

Nom du champ	Description	Exemple
ID_ILOT	Identifiant de l'îlot, commun à chaque année	2268782
ANNEE	Année pour à laquelle les données correspondent	2013
DEPARTEMEN	Département dans lequel se trouve l'îlot	31
ID_RPG_ILO	Identifiant de l'îlot, il peut être différent d'une année à l'autre	031-135388
ID_EXPLOIT	Identifiant de l'exploitation à laquelle l'îlot appartient	031-135388
SURFACE	Surface totale de l'îlot, en ha	9.87
CULT_X	Code de la X ^{ième} culture (X allant de 1 à 5)	19
SURF_X	Surface de la X ^{ième} culture dans l'îlot	9.77

Tableau 4 : Caractéristiques des données attributaires des couches du RPG disponible.

E. RELEVES TERRAIN

Afin d'étudier les relations entre la valeur radiométrique des pixels et le mode de conduite ainsi que la composition botanique des prairies, il est nécessaire d'effectuer des relevés sur la zone d'étude et mener des enquêtes auprès des agriculteurs. Le nombre de prairies à échantillonner est un compromis entre la nécessité d'avoir suffisamment de données et le coût, en termes de temps et de moyen, des relevés : trop peu de données ne permettraient pas une étude statistique satisfaisante et il n'est pas possible de déployer beaucoup de personnel pendant de longues périodes. Nous nous sommes fixé comme objectifs de sélectionner **une cinquantaine de prairies** sur lesquelles les données seront collectées.

1. SELECTION DES PRAIRIES

Afin d'obtenir des échantillons représentatifs de la diversité des prairies de la zone d'étude, il est nécessaire de bien sélectionner les prairies sur lesquelles effectuer des relevés. Plusieurs critères sont à prendre en compte pour capter l'ensemble de la diversité qu'on cherche à étudier. Il faut sélectionner des prairies :

- qui ont différents mode de conduite (fauche, pâture, mixte),
- qui ont des compositions botaniques différentes,
- qui sont bien réparties sur l'ensemble de la zone d'étude,
- pour lesquelles on dispose de l'accord des propriétaires pour réaliser des relevés, ce dernier critère étant le plus limitant

Cette sélection suppose d'avoir des connaissances *a priori* sur le terrain. Avec le RPG, il est possible de savoir quelles prairies sont déclarées temporaires et lesquelles sont déclarées permanentes. Bien que cette définition ne soit pas très pertinente pour notre étude, elle donne néanmoins une indication sur le mode de conduite des prairies (par exemple : la majorité des pâtures s'effectue sur des prairies permanentes).

Grâce aux cartes d'occupation du sol évoquées précédemment, il est également possible de réaliser une carte des âges des prairies (voir en ANNEXE 2). L'information de l'âge fournit une indication sur la végétation potentiellement présente, et sur le caractère temporaire ou non d'une prairie. Une prairie d'un an est souvent mono-spécifique et a des chances d'être retournée dans l'année pour l'exploitation d'une autre culture. Il est donc risqué (et moins intéressant pour l'étude) de les sélectionner pour effectuer un relevé.

Afin d'obtenir des indications supplémentaires sur le mode de conduite et la productivité des prairies, des classifications non-supervisées sont réalisées à partir de séries temporelles du NDVI. Ces approches servent également à identifier des comportements temporels des prairies variées et diversifiées.

Les coordonnées des agriculteurs ne sont en revanche pas disponibles. Une pré-sélection de prairies d'intérêt a donc été faite sur la base des informations décrites précédemment pour les présenter à un conseiller de la chambre d'agriculture de la Haute-Garonne, M. Jean François CAUX. Ce conseiller, de par ses fonctions, est en relation avec certains agriculteurs, et connaît le fonctionnement général des exploitations qu'il suit. De plus, il a une connaissance experte de la zone et de ses caractéristiques. Suite à un entretien avec lui, les coordonnées d'une dizaine

d'agriculteurs ont été récupérées avec des indications sur leur mode d'exploitations. D'autres coordonnées ont été obtenues via M Olivier Bégue du Conseil général de la Haute-Garonne.

Les agriculteurs ont ensuite été contactés par téléphone afin d'obtenir leur accord pour réaliser des relevés sur leurs parcelles et les interroger sur les pratiques suivies en 2013 (nombre et dates de fauches, pâture ou non). Enfin, la sélection finale a été faite en tenant en compte de toutes les informations disponibles. La méthode suivie pour la sélection des prairies à échantillonner est présentée sur le schéma de synthèse ci-dessous (Figure 8).

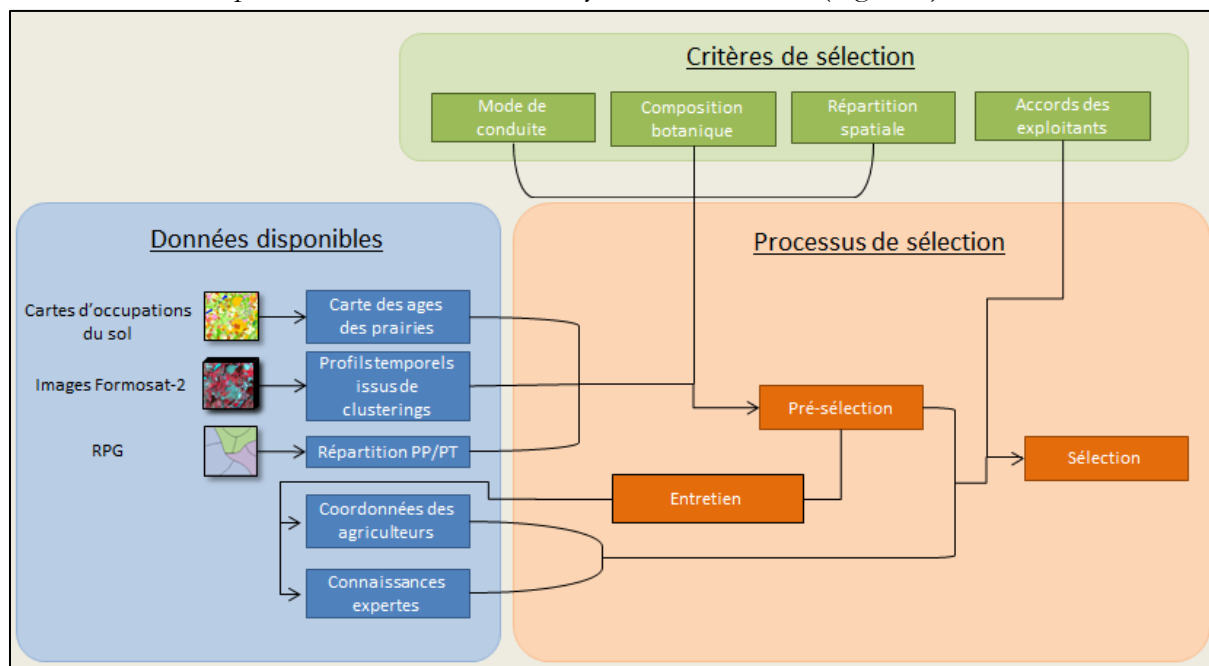


Figure 8 : Schéma de synthèse de la méthode de sélection des prairies à échantillonner.

2. PROTOCOLES DE RELEVÉS TERRAIN

a. Relevés botaniques

Pour les besoins de la thèse de Romain Carrié, doctorant au sein de DYNAFOR, des relevés botaniques sur les prairies de coteaux de Gascogne ont également été planifiés, en plus de ceux nécessaires pour cette étude. La période de relevé sur les coteaux et sur la zone d'étude du stage coïncidant, et dans un souci de mutualisation des moyens à disposition, en termes de personnel compétent en reconnaissance florale disponible, le protocole de relevé botanique a été élaboré conjointement.

Après concertation avec les écologues du laboratoire, le choix du type de relevé botanique s'est porté sur un relevé d'abondance-dominance basé sur l'échelle de Braun-Blanquet (1965). La végétation est étudiée selon un transect allant de part et d'autre de la prairie, ce qui permet de prendre en compte les différents faciès de la prairie s'il y en a plusieurs. Un premier parcours de la prairie est effectué pour avoir une appréciation générale des espèces présentes. Lors d'un deuxième passage, l'abondance et la dominance des espèces rencontrées sont estimées visuellement selon une note allant de * (espèces rare, quelques individus relevés) à 5 (plus de 75% de recouvrement) (Tableau 5).

Indices	*	+	1	2	3	4	5
correspondance	1 individu ou très peu	Quelques dizaines d'individus	Centaine d'individus ou recouvrement de 1-5%	Très abondant ou recouvrement 5-25%	25-50% de recouvrement	50-75% de recouvrement	75-100% de recouvrement

Tableau 5 : Echelle de notation d'abondance dominance inspiré de Braun-Blanquet (1965).

Cette méthode présente l'avantage d'être rapide à mettre en œuvre, environ 30 minutes par parcelle. Elle ne permet pas de faire un relevé exhaustif des espèces présentes mais des espèces constituant le fond prairial, qui est notre objet d'étude. Il est important de noter que la somme des notes peut représenter un recouvrement supérieur à 100%. Les espèces n'ayant pas toute la même hauteur, deux espèces peuvent avoir une note de 5 car le taux de recouvrement est considéré à des strates différentes.

A partir de ces indices, il est possible de calculer le recouvrement de chaque espèce ramené à un total de 100%, et calculer les recouvrements pour les légumineuses, les diverses (autres espèces dicotylédones), les graminées et pour chaque type fonctionnel de graminée, A, B, b, C, D, tels que décrits par Cruz et al., (2010).

Pour les besoins de la thèse de R. Carrié, des relevés de hauteur et de recouvrement floral ont également été réalisés. Ces données ne sont pas utilisées dans le cadre de ce stage et les protocoles ne seront pas explicités dans ce rapport mais les données sont conservées en vue d'une valorisation ultérieure (par exemple : calcul de la productivité fourragère).

L'ensemble des relevés a fait l'objet d'une campagne de terrain de 4 jours, les 28 et 29 avril ainsi que les 4 et 5 mai. Les parcelles ont été regroupées par zone et par journée. Des réunions avant chaque sortie ont eu lieu pour standardiser la méthode de relevé et faire l'état d'avancement de la campagne. Des exemples de fiche terrain sont disponibles en ANNEXE 3 et ANNEXE 4. Les recouvrements des différentes formes de vie et TFG sont quant à eux présentées en ANNEXE 5.

b. Enquête sur le mode de conduite

Quelques questions ont été posées sur le mode de gestion des prairies aux agriculteurs qui pouvaient être présents au moment des relevés. Les questions qui leurs ont été posées sont les suivantes :

- La prairie était-elle fauchée et/ou pâturée en 2013 ?
- Si il y eu des fauches, à quelles dates ont elles eu lieu ?
- De quand date le dernier semis de la prairie ?

La dernière question est plus en lien avec la composition botanique que le mode de conduite. La date du dernier semis donne une indication sur la stabilité de la végétation. En effet, celle-ci se stabilise entre 3 et 4 ans après un semis. Pendant ce laps de temps, elle est susceptible de changer de manière importante d'une année à l'autre.

3. VALIDITE DES DONNEES BOTANIKUES

a. Somme des températures

La croissance et le développement des plantes se déroulent selon différentes étapes dont la durée varie selon l'espèce et la température. La croissance est ainsi plus rapide les années où la température est plus élevée à la fin de l'hiver et au début du printemps. Par conséquent à une même date, le stade de développement n'est pas le même d'une année à l'autre. Il n'est donc pas possible d'indiquer la durée de croissance d'une plante en mois (ou en jours).

La somme des températures depuis le 1^{er} février exprimée en degrés jours (°C.jours) sur une base de 0°C est un bon indicateur qui rend compte du développement des plantes, et qui permet de s'affranchir des données climatiques spécifique d'une année. Il est ainsi possible de mesurer en condition standard des durées de développement. (Baumont et al., 2011)

Pour calculer une somme de température, la température minimale et maximale de la journée sont moyennées. Ensuite si la température est inférieure à 0°C, on enregistre 0°C, et si la température est supérieure à 18°C, on enregistre 18°C. Il faut ensuite sommer toutes les températures à partir du 1 février (Theau et Zerourou, 2008).

La somme des températures a été calculée pour l'année 2013 à partir des températures journalières de la zone d'étude. Ces données ont été obtenues par l'intermédiaire du CESBIO, ayant installé une station météo sur le site. La correspondance entre la somme des températures et les dates d'acquisition est présentée dans le Tableau 6 ci-dessous.

Date d'acquisition	16/02	03/03	03/05	26/05	06/06	26/06	06/07	20/07	30/07
Somme des Températures	98	151	838	1083	1236	1564	1741	1994	2174
Date d'acquisition	11/08	22/08	01/09	21/09	12/10	27/10	28/11	20/12	
Somme des Températures	2390	2588	2768	3108	3452	3706	3990	4112	

Tableau 6 : Correspondances entre les dates d'acquisition des images satellites pour l'année 2013 et la somme des températures en °C.jours.

Les profils temporels des centres de classes issues des classifications supervisées sur la composition botanique seront représentés avec en fonction des somme de température plutôt qu'en fonction des mois de l'année. Il est ainsi directement possible de voir s'il y a une cohérence entre le signal phénologique des classes et la somme des températures.

b. Fenêtre temporelle d'étude

L'étude de la composition botanique ne peut être faite que pour les dates d'acquisition qui correspondent à la période avant la première fauche ou mise pâture. Au-delà, les valeurs de NDVI enregistrées reflètent majoritairement l'effet des pratiques sur la végétation. L'ensemble du signal peut toutefois être informatif. Par exemple, pour deux prairies fauchées à la même date, la

repousse peut être différente selon les espèces présentes. Seule l'évolution de la végétation avant la première intervention a été étudiée pendant le stage.

D'après les enquêtes sur le mode de conduite, la période de fauche la plus récente pour l'année 2013 est mi-mai. Il n'est donc possible de prendre que les trois premières dates (Tableau 6) pour les études sur la composition botanique.

c. Année des relevés VS année d'acquisition des images

Les relevés botaniques ont été effectués cette année (2015) alors que l'année d'acquisition des images satellites est de 2013. Il est donc possible que la composition botanique des prairies diffère entre la date d'acquisition et le moment où les relevés ont été faits, surtout pour les prairies récemment semées.

D'après Michel Duru, chercheur au laboratoire AGIR de l'INRA, la végétation se stabilise et n'évolue plus que très peu dans une prairie, au plus tôt 3 ans après un semis. Les relevés réalisés sur des prairies dont la date de semis est postérieure à 2010 ont donc une probabilité non négligeable de correspondre à une composition différente de celle de 2013.

4. CONSTRUCTION DES CLASSES BOTANIQUES

Les données botaniques sous forme de pourcentage, c'est-à-dire sous forme de variables continues, ne peuvent être utilisées telles quelles pour la réalisation des classifications supervisées. Elles doivent en effet être discrétisées en plusieurs classes. Les données botaniques ont ainsi été séparées selon différentes typologies.

a. Les formes de vie

Une discrétisation des différentes formes de vie, inspirée de Baumont et al. (2011), distingue les prairies selon qu'elles soient :

- Très riches en graminées (taux de recouvrement >75%)
- Riches en légumineuses (taux de recouvrement > 10%)
- Riches en « diverses » (taux de recouvrement > 15%)

Par « diverses » sont désignées toutes les espèces qui ne sont pas des graminées ou des légumineuses. Elles sont souvent considérées comme étant de « mauvaises herbes » car elles ont une valeur fourragère faible, voire nulle. *A contrario*, les légumineuses, qui constituent la famille botanique des Fabacées, ont souvent une valeur fourragère élevée (Hubert and Pierre, 2003.).

Ainsi l'information des proportions des différentes formes de vie donne une première indication de la valeur fourragère d'une prairie. La partition réalisée selon les seuils définis précédemment nous donne six classes possibles de prairies comme on peut le voir sur le Tableau 7. On peut noter que de par la définition des seuils, il n'est pas possible qu'une prairie soit riche en légumineuses, en diverses et très riche en graminées. A l'inverse, il n'est pas possible d'observer une prairie qui ne soit riche dans aucune forme de vie.

Richesse	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6
Riche en légumineuses ?	-	-	-	oui	oui	oui
Riche en diverses ?	-	oui	oui	-	-	oui
Très riche en graminées ?	oui	-	oui	-	oui	-

Tableau 7 : Classes de richesse en formes de vie.

Aucune hypothèse n'est formulée ici quant à des différences attendues entre les dynamiques de NDVI des différentes classes. Cette typologie est testée ici à titre exploratoire puisque d'une part elle présente un intérêt pour le diagnostic agronomique des prairies et que d'autre part parce que les données sur les formes de vie sont disponibles.

b. Classes de productivité et de précocité

Les types fonctionnels de graminées (TFG) de Cruz et al. (2010) sont définis d'après une étude des valeurs de certains traits fonctionnels comme par exemple la Surface Spécifique Foliare (SSF) et la date de floraison (FLO). La SSF représente le rapport entre la surface des feuilles sur la masse sèche (exprimé en m^2/g) et est corrélé avec la productivité primaire et donc *a priori* avec le NDVI. La date de floraison est quant à elle représentative de la phénologie d'une plante. Ainsi les TFG peuvent être utilisés pour **caractériser** la **productivité potentielle** et la **précocité** d'une prairie (Tableau 8).

	Type A	Type B	Type b	Type C	Type D
Productivité (SSF)	Elevée	Elevée	Elevée	Faible	Faible
Précocité (FLO)	Très précoce	Précoce	Tardif	Précoce	Tardif

Tableau 8 : Caractérisation de la productivité potentielle et de la précocité à partir du type fonctionnel de graminées pérennes (Cruz et al., 2010)

A large échelle, Duru et al. (2007) proposent de caractériser la **productivité potentielle** des prairies selon trois classes de TFG : $C+D < 10\%$, $10-40\%$ ou $> 40\%$ (du pourcentage de recouvrement total des graminées), ce qui correspond respectivement à des prairies à la productivité potentielle élevée, normale ou faible. On suppose ici que **les prairies avec un potentiel de productivité élevé se distinguent des prairies avec un faible potentiel par une valeur de NDVI plus élevé que les prairies au potentiel faible.**

Theau et al. (2010) proposent quant à eux de caractériser la **précocité** des prairies selon trois classes de TFG :

- Les prairies très précoces, dont le TFG majoritaire est A.
- Les prairies précoces, dont le cumul des TFG B et C est majoritaire.
- Les prairies tardives, dont le cumul des TFG b et D est majoritaire.

L'hypothèse est ici que **les prairies précoces se distinguent des prairies tardives par une date du maximum de NDVI antérieure à celle des prairies tardives.**

Enfin une dernière typologie a été réalisée en croisant les résultats des deux précédentes, c'est à dire que chaque classe correspond à un potentiel de productivité et un type de précocité distinct, ce qui correspond à un total de neuf classes théoriques.

5. BILAN DE LA CAMPAGNE TERRAIN

a. Nombre de Prairies échantillonnées

A l'issue de la campagne de terrain, des relevés ont été réalisés sur 51 parcelles et des informations sur le mode de conduite ont été collectées sur 41 parcelles (Tableau 9). Quatre types de mode de conduite ont été identifiés :

- les prairies fauchées une fois,
- les prairies fauchées deux fois,
- les prairies pâturées (non fauchées),
- les prairies fauchées puis pâturées.

Mode de conduite	Nombre de parcelles total	Nombre de parcelles avec relevé botanique
Une fauche	20	20
Deux fauches	3	3
Pâturage	11	8
Fauche puis pâturage	7	4
Total	41	35

Tableau 9 : Nombre de parcelles en fonction du mode de conduite et de la réalisation de relevés botaniques.

La répartition des 51 parcelles avec des données botaniques selon les différentes typologies est présentée dans le Tableau 10 qui suit :

(a)	Formes de vie (richesse)						
	Gram	Div	Gram + Div	Leg	Gram + Leg	Div + Leg	
Nombre prairies	8	29	1	1	0	12	
Pixels (%)	14,76	62,42	2,44	2,06	0	18,31	
(b)	Potentiel de productivité						
	Elevée (prod3)	Normal (prod2)	Faible (prod1)				
Nombre prairies	41	9	1				
Pixels (%)	79,88	17,68	2,44				
(c)	Précocité						
	Très précoce (prec3)	Précoce (prec2)	Tardive (prec1)				
Nombre prairies	20	21	10				
Pixels (%)	35,96	49,84	14,19				
(d)	Potentiel de productivité et précocité						
	Prod3 – Prec3	Prod3 – Prec2	Prod3 – Prec1	Prod2 – Prec3	Prod2 – Prec2	Prod2 – Prec1	Prod1- Prec3
Nombre prairies	11	18	4	5	2	1	1
Pixels (%)	23,27	45,06	11,54	10,25	4,78	2,65	2,44

Tableau 10 : Nombre de parcelles et le pourcentage de pixels correspondant en fonction des classes de richesse en formes de vie (a), des classes de potentiel de productivité (b), des classes de précocité (c) et des classes de productivité et précocité combinées (d). « Gram » = Graminées, « Leg » = Légumineuses, « div » = Diverses.

On peut noter qu'il n'y a aucune prairie très riche en graminées et riche en légumineuses qui ait été recensée lors de la campagne terrain. De même, les classes de productivité et précocité « Prod1-Prec2 » et « Prod1-Prec1 » n'ont pas été rencontrées.

b. Description des parcelles visitées

(i) Parcelles homogènes VS parcelles hétérogènes

Lors de la sortie terrain, différentes répartitions de la végétation ont été observées au sein des prairies. Ces différences ne sont pas reflétées par les relevés botaniques qui attribuent une note d'abondance-dominance unique par espèce, pour chaque prairie. Pourtant, si certaines parcelles présentent effectivement une répartition spatiale de la végétation homogène, certaines ont une structure en tâches qui peuvent être de l'ordre de la dizaine de mètres carrés (Figure 9).

D'autres prairies ne présentent pas une structure aussi fragmentée mais on observe un gradient progressif entre deux faciès. Cette hétérogénéité est susceptible d'influencer le signal capté par le satellite.

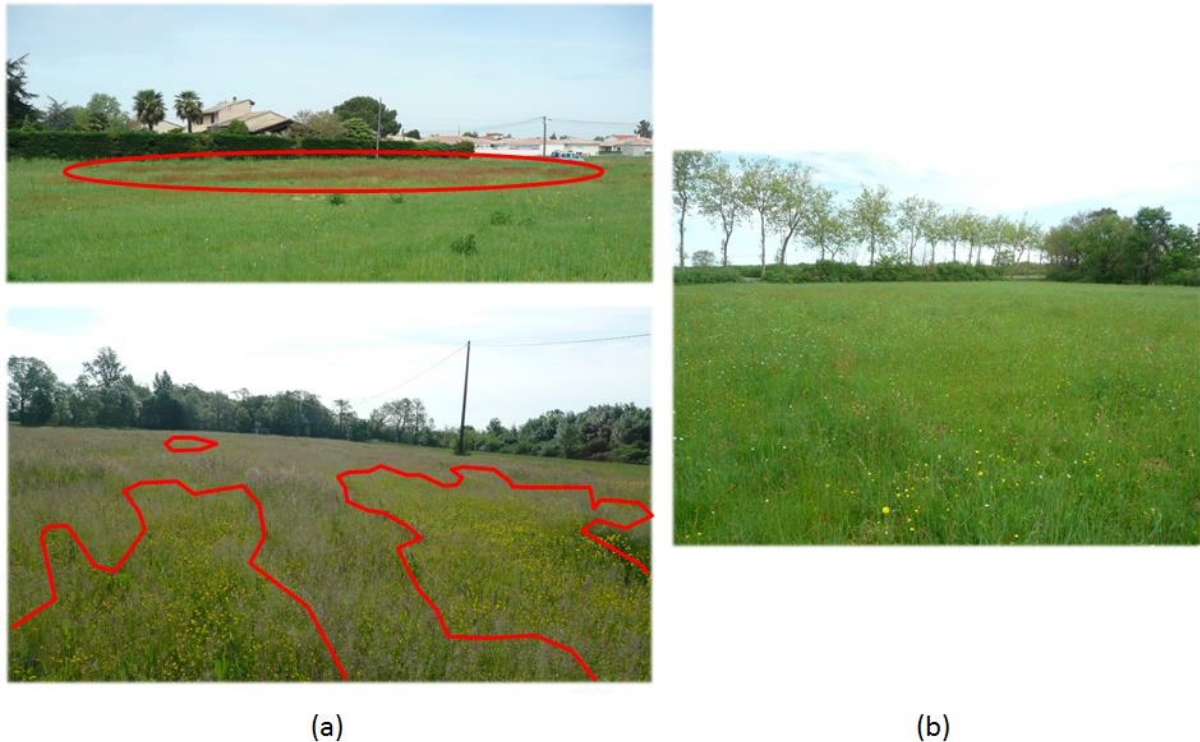


Figure 9 : Photos de prairies : (a) qui ont une composition taches (entourées rouge) et (b) qui ont une composition homogène.

(ii) *Effet de bordure*

Si l'effet de bordure sur la végétation n'a pas été observé de manière systématique sur les prairies échantillonnées, il est toutefois présent sur certaines parcelles. La flore n'y est pas la même près des haies ou des lisières. L'influence des bordures n'excède cependant jamais plus de 10m.

F. TRAITEMENTS STATISTIQUE DES DONNEES

1. TRAITEMENTS PRELIMINAIRES

Les traitements préliminaires consistent en deux étapes principales :

- La première consiste à créer, à partir des images satellites Formosat-2, une image qui contient les valeurs de NDVI de chaque pixel à chaque date, en tenant compte de l'éventuelle présence de nuages sur les images d'origine.
- La deuxième consiste à créer un masque isolant les pixels des prairies de l'image sur lesquels seront appliquées les méthodes de classification et de modélisation.

a. Pré-traitement des images satellitaires

Les pré-traitements des images satellitaires sont présentés dans le diagramme ci-dessous (Figure 10) :

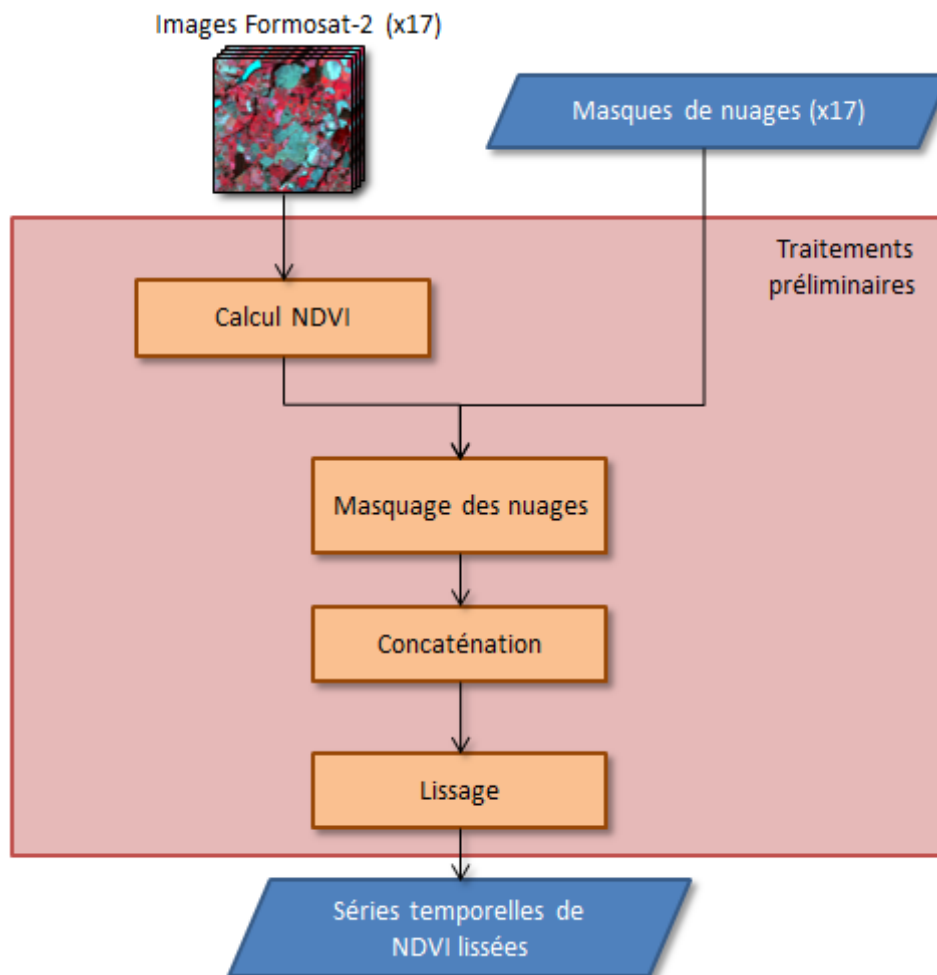


Figure 10 : Chaîne des pré-traitements appliqués aux images satellitaires.

(i) Création de séries ndvi temporelle

L'indice de végétation NDVI est calculé selon la formule présentée ci-dessous. On dispose ensuite de 17 images de NDVI.

$$NDVI = \frac{PIR - R}{PIR + R}$$

Avec « PIR » la valeur de réflectance dans le Proche Infra Rouge et « R » la valeur dans le rouge

(ii) Application d'un masque de nuages

Un des inconvénients des capteurs optiques est qu'ils ne captent pas les signaux des objets se trouvant sous la couverture nuageuse. Or sur certaines images de la série disponible, on observe la présence de telles couvertures pour plusieurs dates (Tableau 3). Grâce aux masques de nuages fournis par le CESBIO, de nouvelles images masquées sont créées.

(iii) Lissage des données

Le lissage des données consiste à remplacer dans une série temporelle les valeurs qui sont manquantes soit à cause de la présence de nuages, soit à cause de défauts du capteur au moment de l'acquisition. Plusieurs méthodes de lissage existent, chacune ayant des avantages et des inconvénients. L'étude de l'apport de ces différentes méthodes a été réalisée Mailys Lopes, doctorante au sein de DYNAFOR. Celle qui a été retenue est le lissage de Whittaker (Eilers, 2003).

b. Création d'un masque de prairie

Lors de ce stage, l'objet d'étude est uniquement les prairies et il convient de créer un masque pour les distinguer du reste des autres occupations du sol (forêts, culture, bâti, eau etc.). Grâce à la déclaration des agriculteurs sur le RPG, l'emplacement des prairies peut être localisé. La création du masque s'effectue selon la démarche suivante (Figure 11):

- Une sélection des îlots de prairies pures issus du RPG: en effet, un îlot peut contenir plusieurs parcelles avec des cultures différentes qu'il n'est pas possible de différencier.
- Une érosion de 10m (un peu plus d'un pixel) est appliquée aux prairies afin de s'affranchir de différents biais en bordure. En effet, le contour des entités présentes dans la couche du RPG correspond aux limites géométriques des îlots. Selon l'angle de prise de vue des images, une partie de l'îlot peut donc être cachée par le feuillage d'une lisière de forêt ou d'un arbre, qui possèdent des caractéristiques spectrales différentes des prairies. Un deuxième biais possible est que certains îlots sont mal dessinés et des routes peuvent y être incluses. Enfin, la végétation des prairies est différente en bordure et ne doit pas être prise en compte lors des traitements statistiques.
- Les prairies ayant une aire inférieure à 1280 m² (20 pixels) sont ensuite enlevées de la sélection. Cette opération permet de limiter quelque peu le volume des données à traiter.
- Enfin, la couche vecteur contenant la sélection est rasterisée afin d'être utilisée pour la suite des traitements.

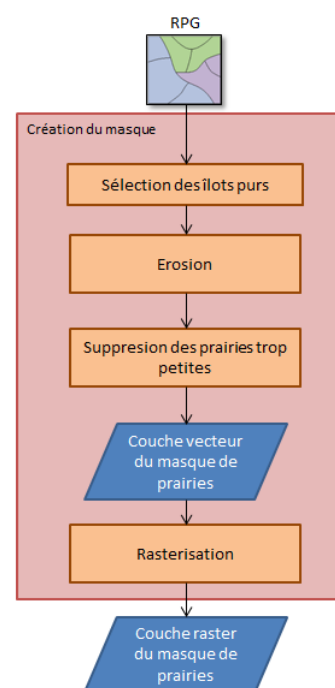


Figure 11 : Schéma explicatif de la création d'un masque de prairie.

2. CLASSIFICATIONS STATISTIQUE SUPERVISEES

a. Objectifs et fonctionnement

L'objectif d'une classification supervisée est de regrouper automatiquement les pixels d'une image selon un nombre de classes d'intérêt prédéfini, et dont on a une connaissance thématique *a priori* (par exemple, des classes de précocité). La méthode de classification utilisée dans ce stage

est celle du *maximum de vraisemblance*. Parmi les plus courantes, facile à utiliser et à interpréter, elle a été sélectionnée pour ces premières classifications sur les typologies retenues. Le fonctionnement général de cette méthode est expliqué dans les paragraphes qui suivent.

Grâce à la récolte données terrain, la classe thématique (label) d'un certain nombre de pixels est connue. A partir de ces derniers, qui constituent ce qu'on appelle des échantillons d'apprentissage, chaque classe est modélisée par une loi de probabilité gaussienne $\mathcal{N}(\mu, \sigma^2)$ où μ est la moyenne des valeurs de NDVI de chaque pixel de la classe, et σ^2 la variance.

Après cette étape d'apprentissage, il est donc possible de calculer pour chaque pixel de l'image, dont le label n'est cette fois-ci pas connu, la probabilité d'appartenance à chacune des classes. A l'issue de la classification, chaque pixel se voit ainsi attribué du label de la classe pour laquelle la probabilité d'appartenance est la plus élevée (Caloz and Collet, 2001).

Dans le cas de notre étude, seuls les pixels appartenant à des prairies sont classés, selon les différentes typologies expliquées précédemment. Le masque des prairies est donc appliqué sur les séries temporelles de NDVI avant l'attribution d'un label aux pixels. Les différentes étapes de la classification supervisée sont présentées en Figure 12 ci-dessous.

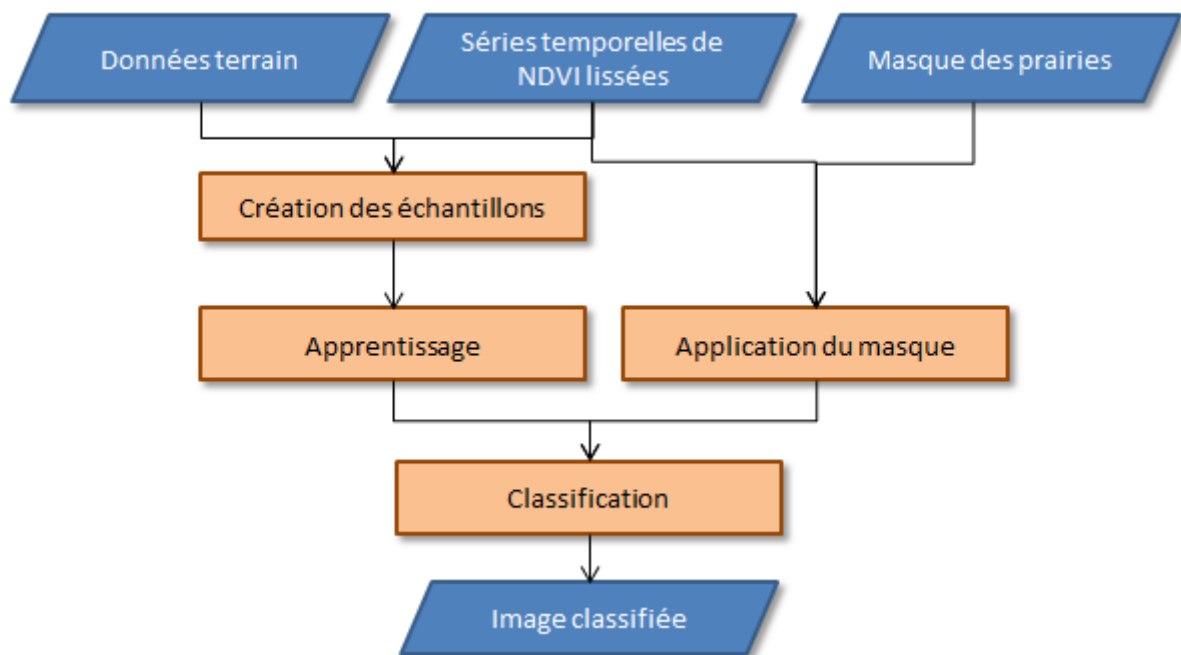


Figure 12 : Schéma explicatifs des étapes d'une classification supervisée

b. Evaluation de la qualité d'une classification

(i) Principe

Afin d'évaluer la qualité d'une classification, il est nécessaire de comparer le label réel de certains pixels (connu d'après les données terrain), au label qui leur est attribué par le modèle appris. Ces pixels, appelés échantillons de vérification (ou validation), ne doivent pas être utilisés lors de la phase d'apprentissage afin que l'évaluation ne soit pas biaisée. Généralement, les trois quarts des échantillons sont utilisés pour l'apprentissage du modèle tandis que le quart restant, indépendant, est réservé à l'évaluation de sa qualité. Dans notre cas, il n'y a pas suffisamment d'échantillons pour certaines classes pour procéder à une validation de cette manière. Par

exemple, il n'y a qu'une prairie de productivité faible dans le cas de la typologie des potentiels de productivité (Tableau 10). L'évaluation de la classification peut alors différer selon la partition réalisée entre échantillons d'apprentissage et échantillons de vérification. Afin d'éviter ce genre de biais, l'évaluation est réalisée trente fois, sur trente jeux d'échantillons d'apprentissage et de validation différents. Chaque partition entre les deux jeux d'échantillons est réalisée de manière aléatoire. La précision générale de la classification est alors estimée par la moyenne des précisions évaluées. La Figure 13 illustre le principe d'évaluation d'une classification.

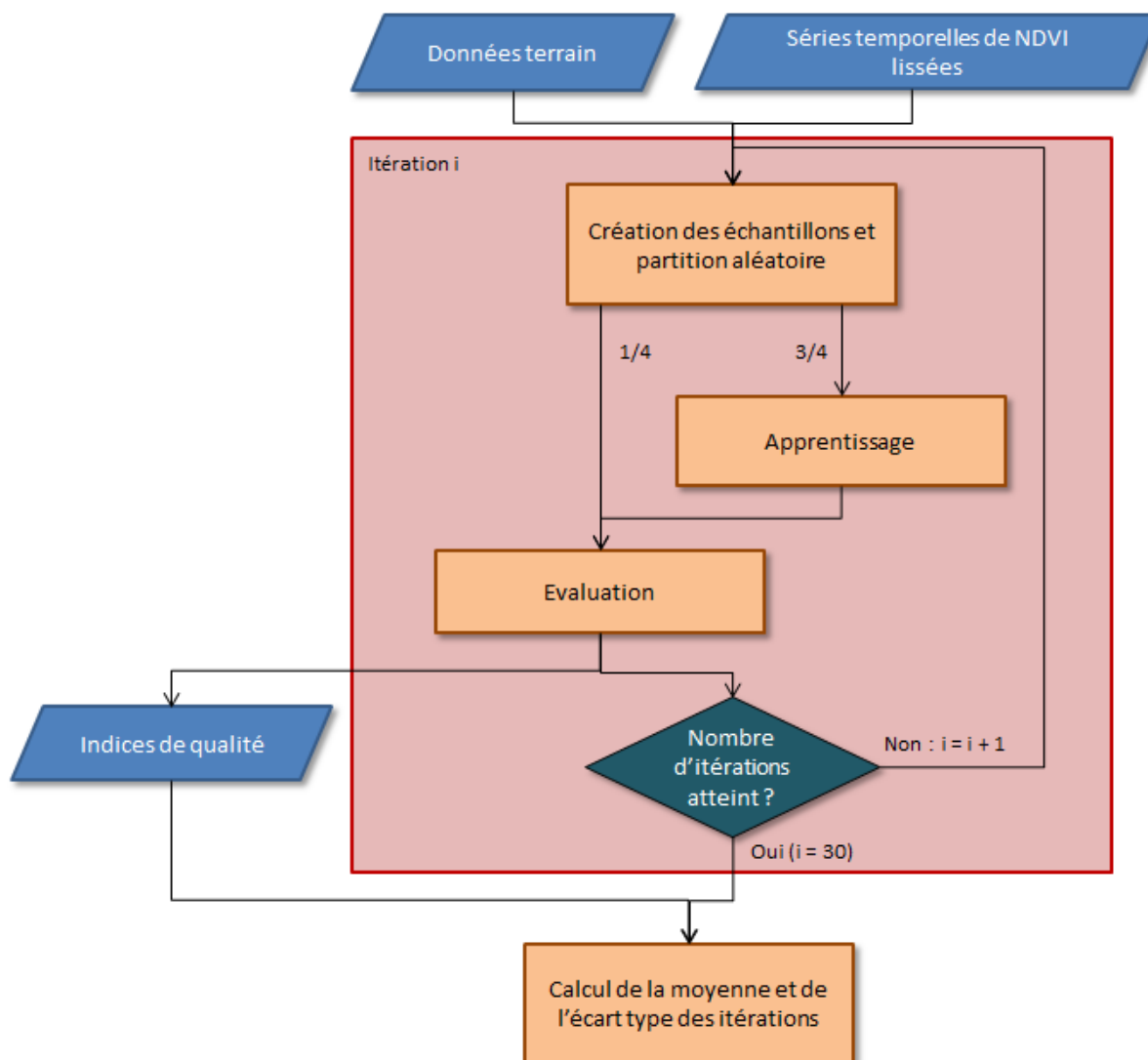


Figure 13 : Schéma explicatif de la méthode de validation des classifications

(ii) *Indices de précision et matrice de confusion*

Il est possible d'évaluer la qualité d'une classification à partir d'une matrice de confusion qui comprend les éléments suivants :

- en colonnes les données correspondantes aux classes attribuées par la classification,
- en lignes les données des classes issues des données terrain.

Les pixels bien classés figurent sur la diagonale de la matrice et ceux ne s’y trouvant pas correspondent à des confusions entre les classes. La précision totale désigne ainsi le nombre de pixels bien classés (en diagonale) sur le nombre de pixels total. Il est également possible de calculer pour chaque classe deux types de précision.

La précision réalisateur qui correspond, pour une classe i , au rapport entre le nombre de pixels bien classés dans cette classe i , et le nombre total de pixels de classe i existant sur le terrain. A partir de celle-ci, il est possible d’en déduire l’erreur de déficit (omission) qui est à égale 100% - la précision réalisateur.

Pour une classe i , la précision utilisateur correspond quant à elle au rapport du nombre de pixels bien classés sur le nombre de pixels total classés dans la classe i . Il est cette fois-ci possible d’en déduire l’erreur d’excédent (commission) qui est égale à 100% - la précision utilisateur (Girard and Girard, 1999).

		Classification				
		Classe 1	Classe 2	Classe 3	Total	Précision réalisateur
Terrain	Classe 1	85	3	0	88	97%
	Classe 2	0	115	0	115	100%
	Classe 3	18	12	25	55	45%
	Total	103	130	25	258	
	Précision utilisateur	82%	88%	100%		

Tableau 11 : Exemple de matrice de confusion

Dans la matrice présentée dans le Tableau 11, on peut voir par exemple qu’aucun pixel appartenant à la classe 2 n’a été classé dans les autres classes, ce qui correspond à une précision réalisateur de 100%. Tous les pixels appartenant à la classe 2 d’après le terrain sont effectivement classés en 2 dans la classification réalisée. En revanche, 3 pixels de la classe 1 et 12 pixels de la classe 3, d’après le terrain, ont été attribués à la classe 2, ce qui correspond à une précision utilisateur de 88%.

Certains pixels peuvent être classés correctement par chance et la précision totale ne tient pas compte de ce phénomène. L’indice du Kappa, qui reflète la qualité globale de la classification, permet de prendre en compte la proportion d’accords aléatoires lors de son calcul. Par exemple, une valeur de 0,75 exprime que la classification utilisée évite 75% des erreurs obtenues par une méthode de classification travaillant complètement aléatoirement (Girard and Girard, 1999).

Pour chaque classification, la matrice de confusion et les différents indices de précision qui en découlent sont calculés sur 30 itérations par la méthode de validation croisée décrite précédemment. Les valeurs obtenues sont donc des moyennes auxquelles il est possible d’associer un écart-type.

3. MODELISATION LINEAIRE

a. Principe

Une autre approche pour essayer de caractériser les prairies par leur composition botanique est d'essayer d'expliquer directement les pourcentages en formes de vie ou en TFG des prairies par les valeurs de NDVI enregistrés à différentes dates. Pour ce faire, des modèles de régression linéaire ont été mis en œuvre. L'objectif de tels modèles est d'expliquer les valeurs prises par une variable Y , dite variable à expliquer, à partir d'une série de variables explicatives X_1, X_2, \dots, X_p , ce pour un nombre d'individus n . Le modèle théorique peut se formuler de la manière suivante :

$$Y_i = a_{0i} + a_{1i}X_{1i} + a_{2i}X_{2i} + \dots + a_{1p}X_{1p} + \varepsilon_i \text{ où } i = 1, 2, \dots, n$$

Où a_1, a_2, \dots, a_p sont les paramètres à estimer du modèle telle que la somme au carré des résidus $\sum_{i=1}^n (\hat{Y}_i - Y_i)^2$ soit la plus petite possible (Hastie et al., 2013), avec \hat{Y}_i les valeurs prédites et Y_i la valeur réelle. Pour notre étude les individus i sont des prairies, les pourcentages de formes de vie et de TFG représentent chacun une variable Y à expliquer (voir le Tableau 12), et les variables X sont des variables spectrales, qui décrivent le comportement spectral d'une prairie.

Or, une prairie, pour laquelle on dispose d'un seul relevé botanique, est composée de plusieurs pixels qui ont chacun des valeurs de NDVI pour trois dates. La question est donc de savoir quelles variables X choisir pour décrire spectralement la prairie : faut-il choisir la moyenne des pixels la composant pour chaque date ? Le maximum pour chaque date ? La variance pour chaque date ? etc.

Variables à expliquer
% de graminées
% de légumineuses
% de « diverses »
% TFG A
% TFG B
%TFG b
%TFG C

Tableau 12 : Variables à expliquer dans les modèles linéaires (en pourcentage du recouvrement total)

b. Extraction des variables descriptives

Classification non-supervisée

Pour vérifier l'homogénéité spectrale de chacune des prairies, des classifications non-supervisées de type Gaussian Mixture Model (GMM) sont réalisées. Elles permettent de regrouper les pixels ayant des caractéristiques spectrales communes en plusieurs groupes, de manière à ce que la variabilité intra-groupe soit la plus faible possible et la variabilité inter-groupe

la plus forte possible (Lang, 2015). La prairie est ainsi considérée comme spectralement homogène si le nombre de groupes optimal trouvé est de un.

Si le nombre optimal obtenu est égal à un, la moyenne et la variance de chaque date du groupe homogène issu de la classification sont alors retenues comme descripteurs de la prairie.

Le nombre de variables descriptives retenues est donc pour l'instant de six : une moyenne et une variance pour chaque date du NDVI des pixels d'une prairie, issues d'une classification non-supervisée (3 + 3 variables).

Analyse en Composantes Principales

Pour chaque prairie, une Analyse en Composante Principales (ACP) a également été réalisée sur les valeurs de NDVI des trois premières dates des pixels la constituant. Cette méthode statistique permet de transformer les trois variables « dates » en trois nouvelles variables décorréliées, qui sont appelées axes principaux. Ces axes sont des combinaisons linéaires des précédentes variables dont les coordonnées permettent de rendre compte de la variabilité spectrale d'une prairie, notamment de la covariance entre les dates (information non représentée avec les simples variances à chaque date). En analysant les ACP réalisées sur chaque prairie, il s'avère que pour 82% des prairies sur lesquelles des relevés ont été effectués, le premier axe principal permet d'expliquer plus de 75% de la variabilité.

Les coordonnées du premier axe propre constituent ainsi trois nouvelles variables descriptives supplémentaires qui reflètent la variabilité des prairies en tenant compte des covariances entre les dates.

Pour chaque variable à expliquer, plusieurs combinaisons de variables explicatives ont été testées : un modèle avec seulement les moyennes à chaque date, un seulement avec la variance à chaque date, un autre avec les coordonnées du premier axe principale, et enfin d'autres avec des combinaisons de ces trois types de variables. Les combinaisons sont explicitées sur le Tableau 13 :

Combinaisons de variables explicatives	Nombre de variables explicatives
Moyennes (m)	3
Variances (v)	3
Coord du 1 ^{ier} axe principal (c)	3
m + v	6
m + c	6
v + c	6
m+v+c	9

Tableau 13 : Combinaisons de variables explicatives testées dans les modèles linéaires

c. Algorithmes utilisés

Trois algorithmes différents ont été utilisés pour déterminer les paramètres a du modèle. Le premier, le modèle de Ridge, permet de trouver des valeurs de coefficient a dont la valeur n'est pas trop élevée. Les deux autres, Ridge et OMP(Orthogonal Matching Pursuit), permettent de sélectionner un nombre réduit de variables explicatives, c'est-à-dire de fixer certains coefficients a à zéro (Hastie et al., 2013).

Il y a donc 3 algorithmes qui ont été utilisées, pour modéliser 7 variables à expliquer avec pour chacune d'entre elles 7 combinaisons de variables explicatives possibles, soit un total de 147 modèles qui ont été testés.

d. Qualité des modèles

L'évaluation de la qualité des modèles est estimée par le coefficient de détermination R^2 qui correspond, dans le cas de modèle linéaires, à la part de la variance du jeu de donnée expliquée par le modèle. La valeur du R^2 , comprise entre 0 et 1, indique à quelle point le jeu de donnée correspond au modèle construit. Une valeur de 1 correspond à une correspondance parfaite entre les données et le modèle créé. D'autres indices mathématiques existent pour analyser les résultats des modélisations linéaires mais dans le cadre de l'approche exploratoire suivie pour ce stage, seul l'analyse du R^2 a été faite.

4. LOGICIELS UTILISES ET IMPLEMENTATION

Les traitements des données vectorielles présentées, la visualisation des images, ainsi que les rendus cartographiques ont été réalisés avec le logiciel Q-GIS.

Tous les traitements des séries temporelles d'images, ainsi que l'exécution des modèles linéaires ont quant à eux été réalisés dans l'environnement Python 2.7. Bien que le développement de fonctions et de chaînes de traitements ne soit pas un des objectifs en tant que tel de mon stage, cette activité y a occupé une place importante. Au total, plus de 2000 lignes de codes ont été produites pour automatiser les chaînes de traitement.

Les algorithmes complexes de classifications statistiques, de modélisation linéaire, ou encore de lissage de séries temporelles utilisés ont été importé depuis les bibliothèques Python OTB (Orfeo ToolBox) et Scikit-Learn. Les fonctionnalités d'ouverture et d'écriture des images satellites ont quant à elles été importées depuis la bibliothèque de traitements d'images GDAL (Geospatial Data Abstraction Library).

Ainsi, le développement informatique réalisé a plus consisté à la mise en place et l'automatisation de chaînes de traitements qu'au développement de nouvelles fonctionnalités. Trois chaînes principales ont été développées :

- Une pour la mise en œuvre de classifications non supervisées (Lang, 2015)

- Une deuxième pour la mise en œuvre de classifications supervisées
- Une dernière pour la réalisation de modélisations linéaires.

Elles ont permis pour chaque classification ou modélisation, de tester de nombreux paramètres et de synthétiser les résultats.

Plusieurs principes ont été respectés lors du développement algorithmique. Au début de chaque fonction, une documentation, appelée *docstring* a été réalisée en expliquant de manière structurée le rôle de la fonction ainsi que la nature et l'usage des paramètres d'entrées et de sortie, ce qui constitue une aide à la compréhension du code.

De plus, la documentation a été écrite en respectant certaines normes de formatage qui permettent la production d'une documentation utilisateur, utile pour utiliser la fonction sans avoir besoin d'en connaître les détails. Un exemple *docstring* est présent en ANNEXE 6.

Le code est également agrémenté de commentaires pour expliciter certaines fonctionnalités élémentaires ou appels de fonction.

Enfin, les commentaires ont été écrits en anglais pour faciliter un éventuel partage des ressources avec des équipes à l'internationale.

Le développement a ainsi été réalisé avec un souci de clarté pour permettre à des utilisateurs ou développeurs de pouvoir reprendre le code le plus aisément possible. La chaîne de traitement de classification non-supervisée illustre bien le respect de ses principes et est représentatif du développement réalisé lors de ce stage (voir ANNEXE 7).

III. RESULTATS ET ANALYSE

A. MASQUE DES PRAIRIES

A l'issue de la création du masque réalisé à partir du RPG, 564 îlots sont obtenus pour un total de 190166 pixels. Les prairies du masque sont présentées dans la figure ci-dessous (Figure 14) :

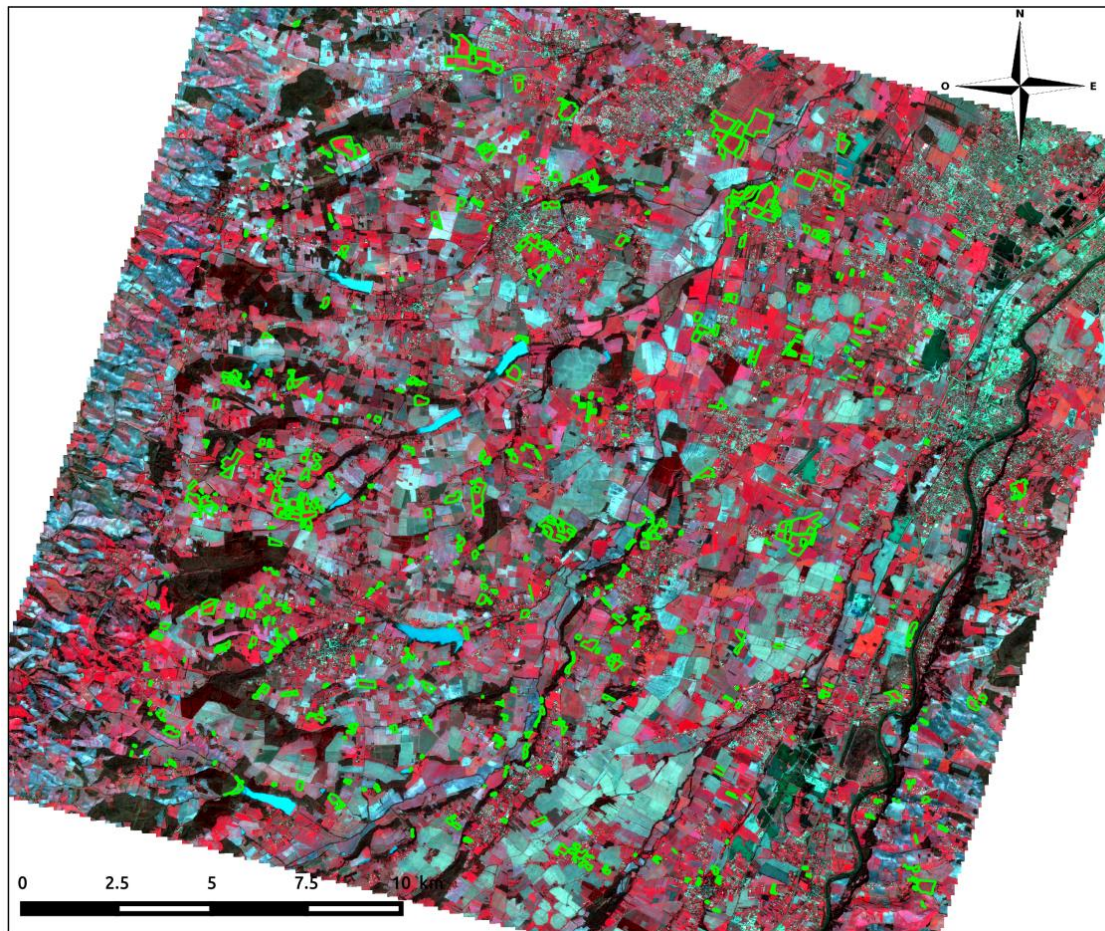


Figure 14 : Sélection des prairies issues du RPG (entourées en vert), extrait d'image Formosat-2 du 03/03/2015 en fausses couleurs (RGB: PIR, R, V).

B. CLASSIFICATION SUR LE MODE DE CONDUITE

Les résultats de la classification supervisée sur les différents modes de conduite sont présentés dans la matrice de confusion suivante (Tableau 14) :

Mode de conduite		Classification				
		1 fauche	2 fauches	pâturation	Fauche puis pâturation	Précision réalisateur
Terrain	1 fauche	0.67	0.07	0.22	0.03	67%(±2)
	2 fauches	0.10	0.75	0.11	0.04	75%(±4)
	pâturation	0.07	0	0.88	0.05	88%(±3)
	Fauche puis pâturation	0.14	0	0.33	0.53	53%(±5)
	Précision utilisateur	85%(±2)	63%(±5)	59%(±3)	76%(±4)	
Précision globale : 71% (±1) Kappa : 0.58 (±0.02)						

Tableau 14 : Matrice de confusion pour la classification supervisée sur les modes de conduite. Les résultats sont normalisés par rapport aux données terrain (moyennes de validations croisées sur 30 itérations, avec l'écart-type associé entre parenthèse).

La précision globale de 71% de pixels bien classés est plutôt bonne même si la valeur du kappa indique un degré d'accord modéré. La classe pâturation est celle qui présente le meilleur taux d'accord avec 88% des pixels bien classés. Cependant, elle est aussi celle qui a la plus faible précision utilisateur (ou le taux d'excédents le plus élevé). En effet, il y a des confusions (33%) importantes avec la classe « fauche puis pâturation ». Enfin, les classes « 1 fauche » et « 2 fauches », qui ont des taux d'accord acceptables (entre 67 et 75%), se confondent mutuellement.

C. CLASSIFICATIONS ISSUES LA COMPOSITION BOTANIQUE

Les résultats des classifications supervisées sur les typologies issues de la composition botanique sont présentés dans le Tableau 15 ci-dessous :

	Forme de vie	Potentiel de productivité	Précocité	Productivité et précocité
Précision Globale	42% (±1)	73% (±1)	0.58 (±1)	51% (±1)
Kappa	0.37 (±0.08)	0.47 (±0.01)	0.24 (±0.01)	0.41 (±0.05)

Tableau 15 : Indices de qualité globale des différentes classifications supervisées sur les classes issues de la composition botanique (moyennes de 30 itérations, avec l'écart-type associé entre parenthèse).

La précision totale des classifications est faible sauf pour celle sur les classes de potentiel de productivité. En revanche l'indice de Kappa de cette dernière tant à indiquer que si cette précision assez élevée, certaines classes sont mal classées. La qualité globale des classifications peut donc être qualifiée de mauvaise.

Les précisions *réalisateur* et *utilisateur* de chaque classe pour toutes les typologies sont présentées dans le Tableau 16. Les matrices de confusion détaillées figurent quant à elles en ANNEXE 8.

(a)	Formes de vie (richesse)						
	Gram	Div	Gram + Div	Leg	Div + Leg		
Précision réalisateur	18% (± 2)	59% (± 1)	72% (± 6)	67% (± 6)	0% (± 0)		
Précision utilisateur	13% (± 1)	66% (± 1)	9% (± 1)	22% (± 3)	0% (± 0)		
(b)	Potentiel de productivité						
	Elevée (prod3)	Normale (prod2)	Faible (prod1)				
Précision réalisateur	72% (± 1)	87% (± 2)	0% (± 0)				
Précision utilisateur	96% (± 1)	38% (± 1)	0% (± 0)				
(c)	Précocité						
	Très précoce (prec3)	Précoce (prec2)	Tardive (prec1)				
Précision réalisateur	36% (± 1)	84% (± 1)	25% (± 2)				
Précision utilisateur	64% (± 2)	57% (± 1)	57% (± 4)				
(d)	Potentiel de productivité et Précocité						
	Prod3 – Prec3	Prod3 – Prec2	Prod3 – Prec1	Prod2 – Prec3	Prod2 – Prec2	Prod2 – Prec1	Prod1- Prec3
Précision réalisateur	52% (± 2)	56% (± 1)	29% (± 2)	72% (± 3)	19% (± 3)	78% (± 3)	1% (± 1)
Précision utilisateur	62% (± 2)	60% (± 1)	62% (± 4)	28% (± 2)	53% (± 9)	42% (± 5)	2% (± 4)

Tableau 16 : Précision réalisateur et utilisateur pour chaque classes de la classification supervisée sur la richesse de formes de vie (a), la productivité potentielle (b), la phénologie (c), sur la productivité potentielle et la phénologie (d)(moyennes de 30 itérations, avec l'écart-type associé entre parenthèse).

La plupart des classes ont des précisions réalisateur et utilisateur faibles (<60%) voire très faible (<40%). La faible qualité globale des classifications n'est donc pas due à quelques classes mal classées mais à une faible précision sur l'ensemble des classes.

Deux classes, « div+leg » pour la typologie des richesses en formes de vie, et « faible » pour la typologie des potentiels de productivités, sont totalement confondus et ont de ce fait une précision nulle. Les classe des productivités faible n'est constituée que d'une seule prairie, avec peu de pixels, ce qui peut expliquer pourquoi cette classe sous représentée est mal classée.

Les quelques classes, (« prod2-prec2 » ou la classe « Précoce » par exemple) qui ont des taux d'accord assez élevée ont en revanche une faible précision utilisateur et sont surclassés.

L'ensemble des images classifiées sont présentées en ANNEXE 9 à 11.

D. MODELES LINEAIRES

Les valeurs du coefficient de détermination R^2 pour les 147 modèles testés sont présentées en ANNEXE 14. Seules les valeurs de R^2 supérieures à 0.3 sont présentées sur la Tableau 17 ci-dessous :

Algorithme	Variable à expliquer	Combinaison de variables explicatives	Valeur du R^2
Ridge	% b	m+c	0.33
Ridge	% b	m+v+c	0.41
Lasso	%B	m+v+c	0.44
Lasso	% b	m+v+c	0.36
OMP	% b	m+v+c	0.39

Tableau 17 : Coefficient de détermination les plus significatifs parmi les 174 modèles réalisés

Seuls cinq modèles présentent une valeur de R^2 supérieure à 0.3. Pour ces modèles, la variable à expliquer qui est la plus retenue est le taux de recouvrement en graminées de type b. La combinaison de variables explicatives la plus retenue est quant à elle celle avec les moyennes et les variances de NDVI des prairies à chaque date, associées aux coordonnées de l'axe principale. De manière générale, l'algorithme Lasso est celui pour lequel la part de la variance expliquée est le plus souvent nulle (ANNEXE 14).

IV. DISCUSSION

A. UNE DISTINCTION ENTRE FAUCHE ET PATURE POSSIBLE

Au vu des résultats présentés dans la matrice de confusion (Tableau 14), la distinction entre les prairies uniquement fauchées et les prairies sur lesquelles il y a de la pâture est possible. En effet, si les classes « 1 fauche » et « 2 fauche » sont regroupées ensemble ainsi que les classes « pâture » et « fauche puis pâture », la précision totale de la classification atteint plus de 80% de pixels bien classés.

En revanche, à l'instar du précédent rapport (Lang, 2015), il n'est pas possible de déterminer des dates de fauches d'après l'évolution du centres des classes de fauche (Figure 15), car les diminutions de NDVI observées ont lieu sur de très longues périodes (de deux à quatre mois) alors que la diminution attendue provoquée par une fauche est supposée être forte et sur période de temps courte.

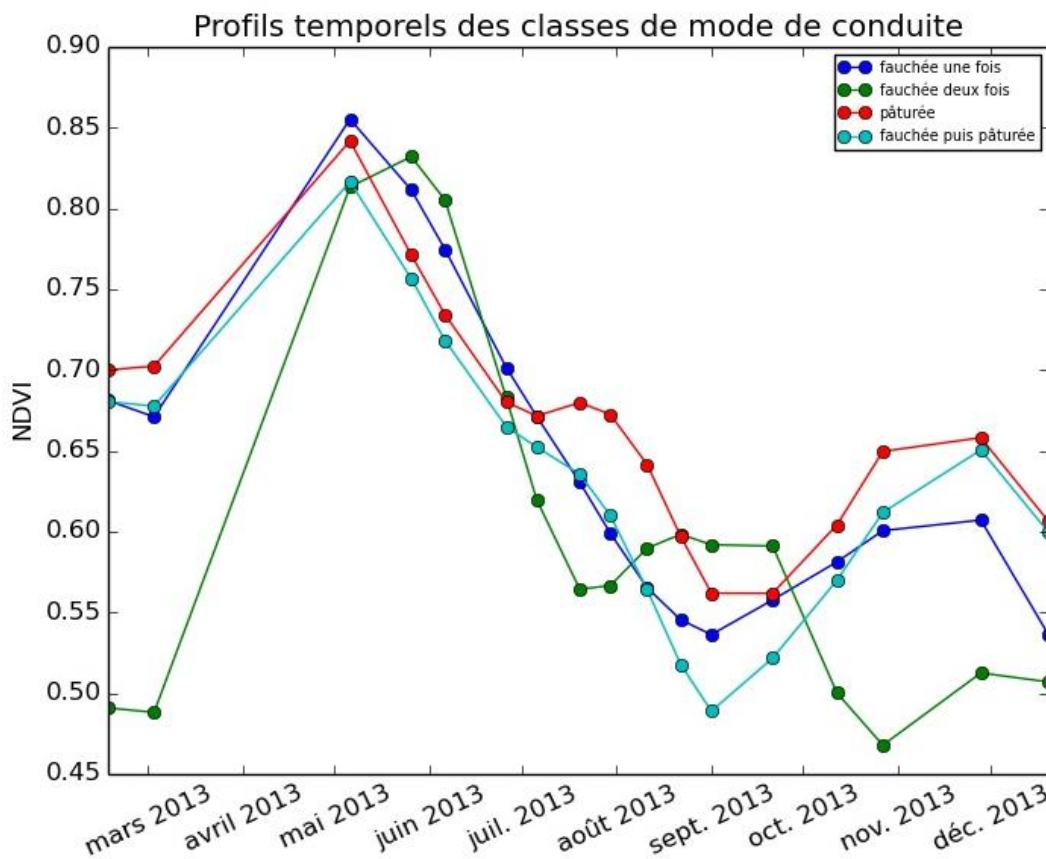


Figure 15 : Profils de NDVI des centres des classes de la classification supervisée sur les modes de conduite

B. UNE DISTINCTION FONDEE SUR LES TYPOLOGIES BOTANIQUES DIFFICILE

Nous avons vu que les classifications visant à discriminer les prairies sur la base de leur productivité, leur précocité ou une combinaison de ces deux critères (valables pour les graminées) ne fournissaient pas des résultats satisfaisants. Plusieurs éléments d'explication sont proposés dans les parties suivantes pour expliquer la faible qualité des classifications.

1. DE NOMBREUX PARAMETRES NON CONTROLES

Il a été montré dans la section I.B. qu'il est possible de caractériser une prairie selon de nombreuses entrées interdépendantes (Figure 3), ce qui rend la question complexe. L'approche par la composition botanique fonctionnelle utilisée dans ce rapport intègre dans une certaine mesure ces différentes entrées, comme la fertilité des prairies.

Cependant, sur la Figure 16, la classe de productivité potentielle élevée a, pour les trois dates, des valeurs NDVI plus faibles que les classes avec un potentiel normal et faible, ce qui est l'inverse de l'hypothèse attendue. Ces classes représentent en fait des **potentiels** qui ne se réalisent que pour des conditions de climat favorables et des pratiques adaptées. Si il est raisonnable de considérer que les conditions climatiques sont à peu près les mêmes pour toutes les prairies de notre zone d'étude, ce n'est en revanche pas le cas pour les pratiques de fertilisation qui peut différer d'une prairie à l'autre. Ainsi, l'effet de la disponibilité des nutriments des prairies sur les valeurs de NDVI observée de chaque classe est peut-être non négligeable et à prendre en compte lors de l'étude des valeurs de NDVI sur une prairie.

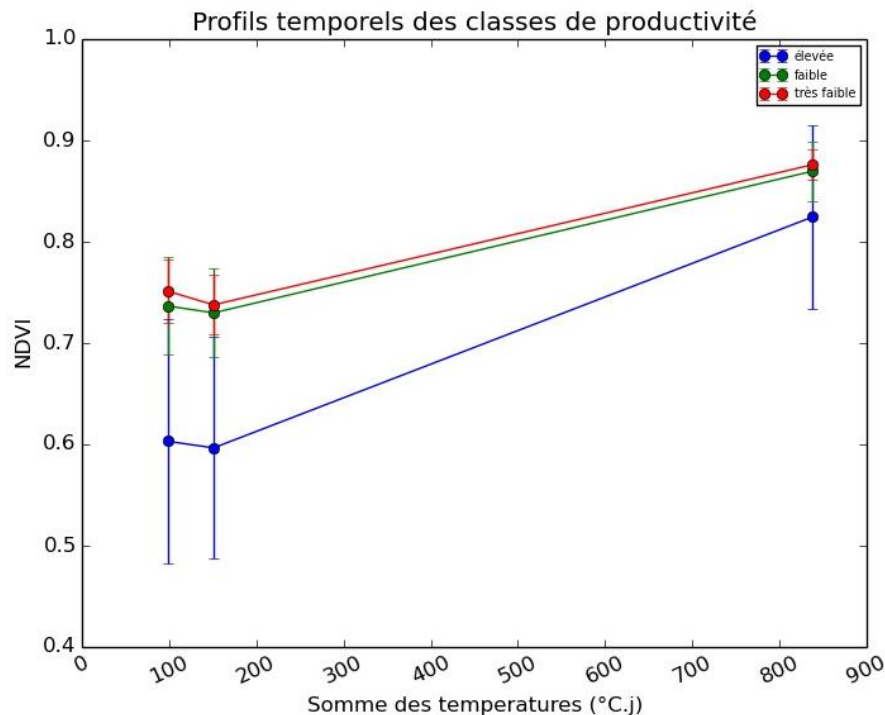


Figure 16 : Profils de NDVI des centres des classes de la classification supervisée sur les productivités potentiels. Les barres d'erreurs correspondent à l'écart-type des classes à chaque date.

De la même manière les différences observées sur la première date sont peut-être principalement le reflet de la gestion des prairies lors du précédent hiver. Selon qu'il y ait eu une pâture hivernale ou non, le stade de développement de la végétation peut en effet être différent.

Enfin, un dernier paramètre non contrôlé peut expliquer la variabilité des classes et la faible qualité des classifications. Pour 22 des prairies sur lesquelles des relevés botaniques ont été réalisés, la dernière date de semis n'est pas connue. Il n'est donc pas certain pour ces prairies que la végétation observée soit la même que celle qui était présente en 2013, date d'acquisition des images satellites de l'étude. La végétation ne se stabilise en effet qu'après trois ans au minimum après le dernier semis.

2. PERTINENCE AGRNOMIQUE VS PERTIENCE EN TELEDETECTION DES TYPLOGIES

Les méthodes de classifications supervisées nécessitent l'utilisation de classes d'apprentissage. Or, les données de départ sont des abondances sous forme de variables continues, qui ont été discrétisé en différents groupes (richesse en formes de vie, productivité et précocité). Le choix des seuils de discrétisation peut avoir une influence non négligeable sur la qualité des classifications. Les valeurs des seuils qui ont été utilisés pour la construction des classes permettent d'établir des typologies qui donnent des indications pertinentes pour évaluer le potentiel agronomique d'une parcelle. La qualité des classifications réalisées soulève un doute quant à la pertinence de la distinction des classes construite grâce à l'étude des dynamiques de NDVI.

En effet, pour la classification des richesses en formes de vie, si une abondance supérieure à 10% en légumineuses est considérée comme forte (en donnant indication sur la valeur fourragère de la prairie), l'influence de ces légumineuses sur le signal capté par les satellites risque d'être faible voire négligeable. Le choix de la discrétisation des données initiales pourrait expliquer la variabilité à chaque date des différentes classes de richesse en formes de vie. On peut en effet voir sur la Figure 17 que les profils de NDVI moyens des classes de formes de vies sont très similaires et ne se distinguent pas clairement les uns les autres.

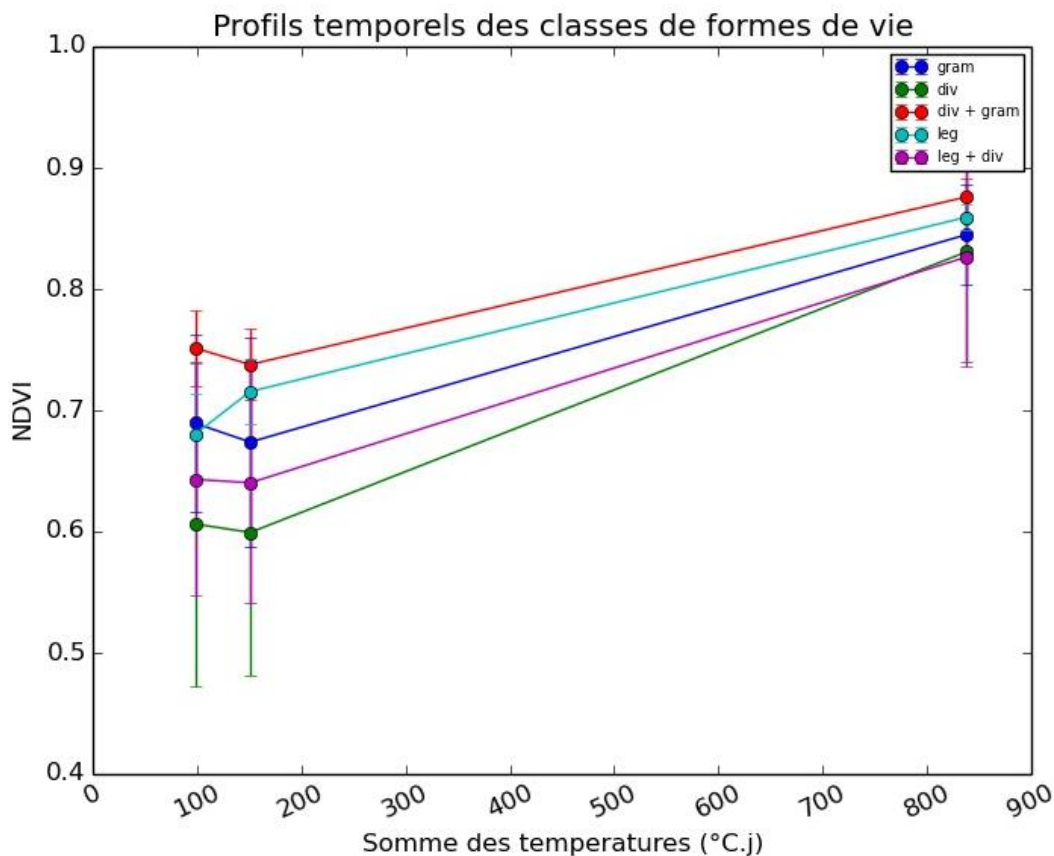


Figure 17 : Profils de NDVI des centres des classes de la classification supervisée sur la richesse en formes de vie. Les barres d'erreurs correspondent à l'écart-type des classes à chaque date.

La construction de la typologie sur la précocité et de la productivité potentielle des prairies est basée sur l'abondance relative des différents TFG. Or, l'abondance cumulée dans les prairies des graminées constitutives de ces TFG représente parfois moins de la moitié de toutes les espèces présentes. Si ce n'est pas une limite pour estimer la valeur d'usage agronomique, l'influence des espèces non constitutives des TFG peut s'avérer non négligeable sur les valeurs de NDVI enregistrées.

3. DES DONNEES DISPONIBLES INSUFFISANTES

a. Echantillonnage temporelle

L'étude de la composition botanique des prairies ne peut être réalisée que sur la période qui précède la première intervention, ce qui diminue considérablement le nombre de dates disponibles.

Les profils temporels de NDVI des centres des classes de précocité sont présentés sur la Figure 18. Outre le fait que la variabilité des classes soit importante (lié à ce qui a été dit précédemment), il n'est pas possible de conclure quant à l'hypothèse, à savoir que les prairies précoces se distinguent par un maximum de NDVI apparaissant plus tôt que celui des prairies tardives. D'une part, le manque d'images entre les sommes de température de 200 et 800 °C.jours

de dates ne permet pas de bien suivre l'évolution de l'activité photosynthétique des prairies pendant le début de la phase de croissance des plantes. D'autre part, l'incertitude quant aux dates d'interventions ne permet pas de prendre en compte les valeurs après 850°C.jours, et donc de suivre les couverts végétaux en fin de croissance. Un nombre plus important de dates (i.e. d'images) entre 700 et 1500°C.jours pourrait permettre une meilleure séparabilité des classes.

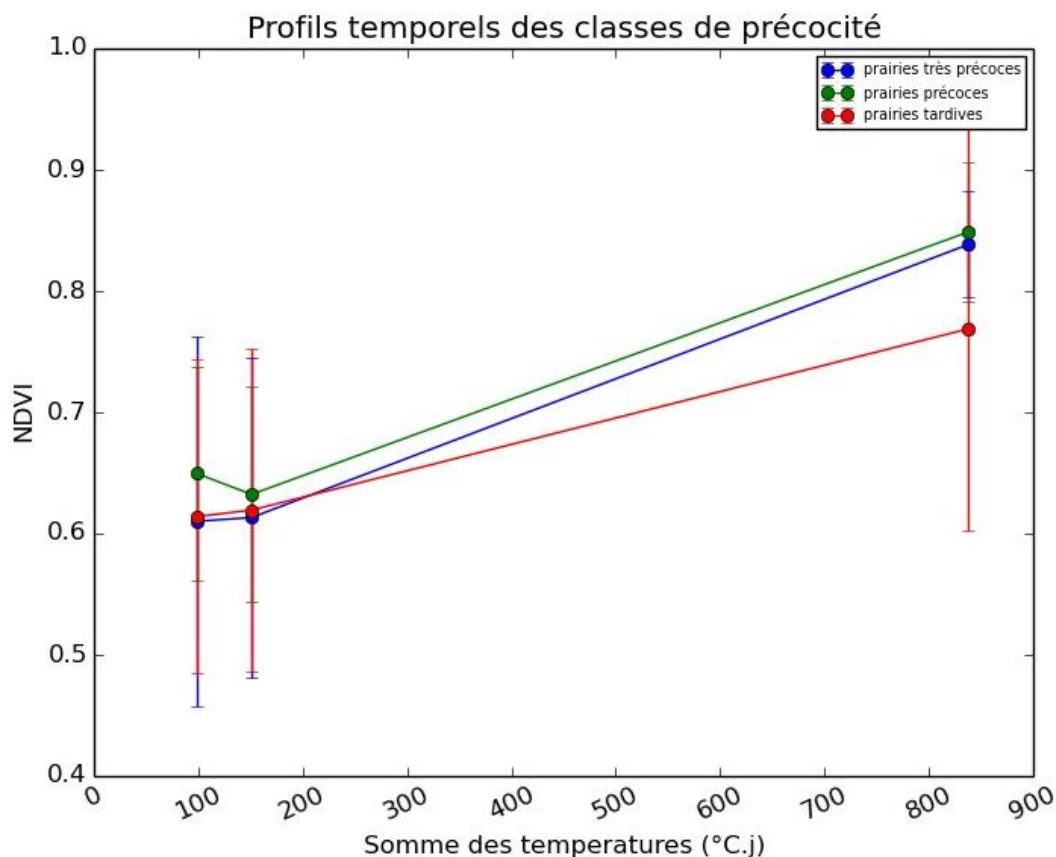


Figure 18 : Profils de NDVI des centres des classes de la classification supervisée sur la phénologie. Les barres d'erreurs correspondent à l'écart-type des classes à chaque date.

b. Echantillonnage thématique

L'échantillonnage irrégulier peut être une de raison expliquant les confusions entre certaines classes. En effet, certaines catégorie de prairies sont sur échantillonnées (classe des prairies riches en diverses) et d'autres sont sous échantillonnées, comme par exemple la classe «gram + div » de la classification des richesses en formes de vie et la classe de productivité potentielle faible. Il n'y a en effet pas assez de prairies pour chaque modalité de certaines typologies, et l'échantillonnage n'est ainsi pas toujours représentatif de la variabilité des typologies étudiées.

De fait, l'accord des agriculteurs pour accéder aux prairies n'est pas toujours évident à obtenir et il n'est pas possible de savoir *a priori* le type de végétation présente sur les parcelles.

L'ensemble des facteurs pouvant expliquer la mauvaise séparabilité des classes et la faible qualité des classifications sont résumés en Figure 19 ci-dessous. Ces différents facteurs sont interdépendants. Par exemple :

- la pertinence de l'utilisation d'une typologie dépend des données disponibles pour son étude et des paramètres qu'elle prend en compte.
- contrôler tous les paramètres pour isoler l'effet d'un seul paramètre implique l'utilisation d'une typologie complexe avec de nombreuses modalités, qui nécessite un échantillonnage important.

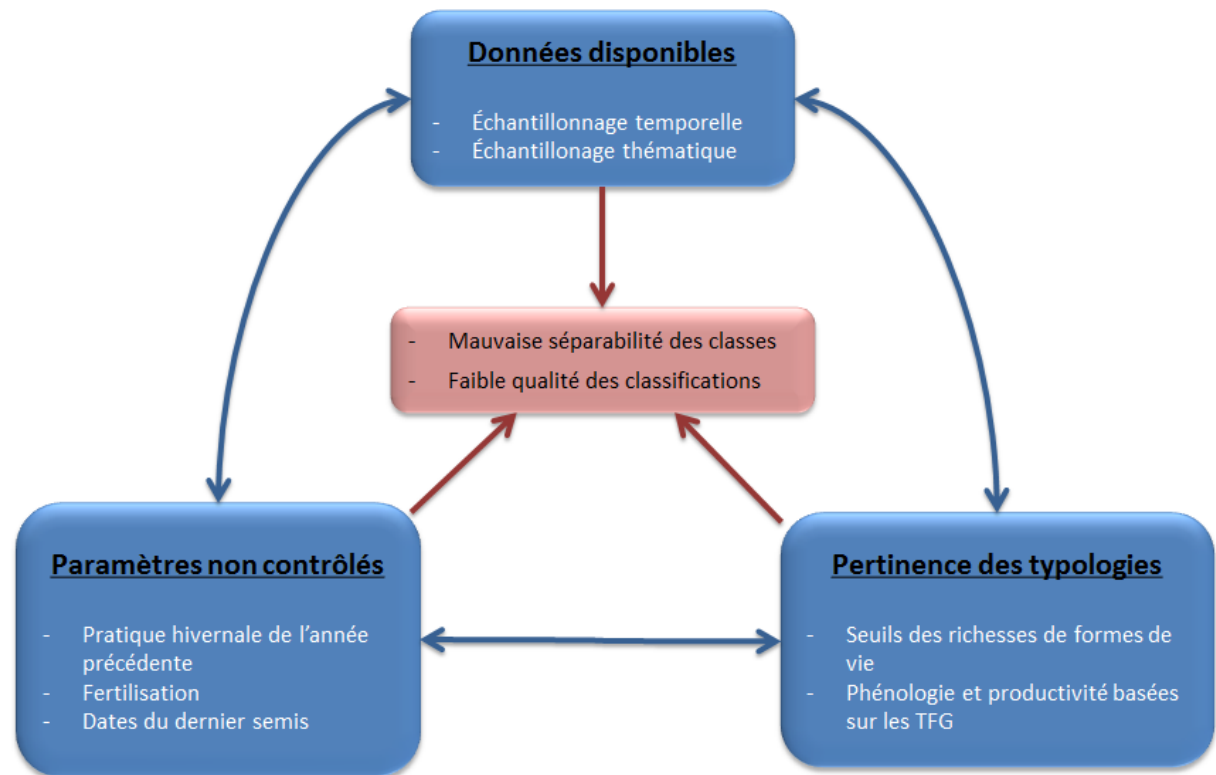


Figure 19 : Schéma de synthèse des différents facteurs pouvant expliquer la faible qualité des classifications réalisées

C. DES MODELES LINEAIRES CONSTRUITS PEU EXPLICATIFS

Les premiers résultats montrent que la plupart des modèles réalisés expliquent très mal le jeu de données. Même les meilleures valeurs du coefficient de détermination, qui sont de l'ordre de 0.4, restent faibles.

A l'inverse des classifications supervisées, les modèles linéaires ne nécessitent pas une discrétisation des données en différentes classes. De fait, ce ne sont donc pas des problèmes liés à la pertinence des typologies construites ou au manque de données pour certaines classes qui sont à l'origine de ces mauvais résultats. En revanche, les biais liés aux paramètres non contrôlés pourraient expliquer la faible capacité explicative des modèles. De plus, comme souligné dans le paragraphe précédent, le faible nombre d'images disponible pendant la phase de croissance des plantes peut également influencer le pouvoir explicatifs des données.

Les premiers résultats des modèles linéaires ne sont ainsi pas encourageants quant à la possibilité d'expliquer la composition botanique des prairies directement avec les valeurs de NDVI.

V. BILAN ET PERSPECTIVES

A. CONCLUSION

L'objectif de ce stage était d'étudier les possibilités des séries temporelles d'images satellites FORMOSAT-2 pour la caractérisation des prairies par leur mode de gestion et par leur composition botanique. Plusieurs hypothèses ont été formulées. La première était que les évolutions du NDVI au cours d'une année permettaient d'identifier les changements de l'activité photosynthétique des prairies liés aux interventions agricoles, comme les fauches ou les mises en pâture.

Grâce à la connaissance des espèces présentes dans une prairie, il est possible de distinguer les prairies selon leur productivité potentielle et leur précocité. Une seconde hypothèse était ainsi que la valeur et la date du maximum de NDVI permette de caractériser la productivité et la précocité des prairies.

Afin de vérifier ces hypothèses, une campagne de terrain a été réalisée pour connaître la composition botanique et le mode de gestion d'une cinquantaine de prairie. A partir de ces données, les taux de recouvrement des graminées, des légumineuses, des espèces « diverses » et des différents type fonctionnels de graminées A, B, b, C et D ont pu être calculés. Des classes de productivité potentielle, de précocité et de richesse en formes de vie ont ensuite été construites pour la réalisation de classification supervisée. Une approche plus directe a également été testée avec la construction de modèles linéaires reliant les différents taux de recouvrement calculés avec les valeurs de NDVI des prairies.

A l'inverse des classifications non supervisées qui, comme détaillé dans le précédent rapport (Lang, 2015), ne permettent pas d'identifier le mode de gestion des prairies, les résultats de la classification supervisée sur les modes de conduite montrent qu'une distinction entre les prairies fauchées et les prairies pâturées est possible. Afin de caractériser plus finement le mode de conduite, des méthodes basées sur l'utilisation de descripteurs de la dynamique des profils de NDVI des prairies vont être étudiées une doctorante du laboratoire. L'objectif suivi est ainsi d'essayer de déterminer pour chaque prairie, les dates précises d'intervention et leur nature.

En revanche, la qualité des classifications sur les classes botaniques est médiocre et les prairies n'ont pu être distinguées selon leur potentiel de productivité, leur précocité, ou encore selon leur richesse en formes de vie. De nombreux facteurs sont susceptibles d'expliquer ces mauvais résultats. En effet de nombreux paramètres comme les pratiques de fertilisation ne sont pas contrôlés et la définition des typologies ainsi que la discrétisation des données botaniques n'apparaissent pas très adaptés par rapport au type de données utilisées. De plus, même si la série temporelle contient un nombre d'images particulièrement élevé, le nombre de dates disponibles pendant la croissance végétale n'est pas très important.

Plusieurs choses sont à prendre en compte pour améliorer les résultats. Il est nécessaire de contrôler plus de paramètres afin de pouvoir isoler un à un les effets affectant les valeurs de NDVI. Ceci entraînerait cependant la création d'un nombre de classes plus important avec la construction de typologies beaucoup plus complexes et difficiles à interpréter. De plus, cela nécessiterait un échantillonnage des prairies beaucoup plus conséquent. Etant donnée la difficulté que représente le contrôle des paramètres, l'utilisation d'un dispositif expérimentale sur des parcelles de l'INRA est une solution à envisager.

De plus, l'approche fonctionnelle, au lieu d'être réalisée uniquement sur les graminées, pourrait être élargie à l'ensemble des espèces prairiales grâce à l'utilisation d'abaques décrivant la valeur de traits fonctionnels pour de nombreuses espèces végétales. En utilisant des traits corrélés au NDVI, les résultats de classification pourraient être améliorés.

Enfin, la disponibilité prochaine des images du satellite Sentinel-2 pourrait permettre l'obtention d'un plus grand nombre de dates pendant la période de croissance végétale. Ainsi, l'évolution phénologique pourrait être étudiée avec plus de précision.

Ce stage rend ainsi compte de la complexité de la caractérisation des prairies par télédétection. De nombreux facteurs sont à prendre en compte et de nombreuses méthodes sont encore à tester pour cet objet d'étude encore peu étudié à l'heure actuelle.

B. BILAN PERSONNEL

Ce stage m'est apparu très complémentaire car il m'a permis de combiner mon intérêt pour les questions agronomiques et environnementales ainsi que celui que j'ai pour la géomatique et la télédétection.

En effet, j'ai pu étudier avec curiosité les prairies pendant ces quelques mois, sujet qui est nouveau pour moi. Il m'a permis de mobiliser les connaissances acquises lors de ma formation d'ingénieur agronome.

De plus, ce stage m'a permis de consolider et d'approfondir les connaissances acquises durant cette année universitaire. J'ai beaucoup appris en télédétection avec la mise en œuvre de différentes approches de classification et l'analyse des profils temporels de l'indice de végétation NDVI.

J'ai également approfondi mes compétences en programmation puisque l'ensemble des traitements des images s'est fait par l'intermédiaire de scripts python. Le projet de fin d'année sur le développement d'une méthode de classification objet de la physiologie de la végétation basée sur ce même langage m'a été d'une grande aide.

J'ai pu mobiliser mes connaissances dans l'utilisation du logiciel SIG QGIS à travers le traitement de données vectorielles comme le RPG, ainsi que mes connaissances dans la gestion de base de données lors de l'acquisition des données de terrain.

Outre les compétences techniques que j'ai pu développer et mobiliser, l'expérience de l'intégration au sein d'une équipe pluri-disciplinaires a été très enrichissante car j'ai pu collaborer avec de nombreuses personnes et participer à des échanges productifs, aussi bien pour mon stage que pour ma culture scientifique personnelle. De la mise en œuvre de méthodes statistiques à l'approche fonctionnelle botanique des prairies, de la gestion d'une campagne de terrain au

développement informatique de chaîne de traitements, en passant par des entretiens avec des acteurs du monde agricole, ce stage m'aura permis de développer un plus la polyvalence de mon profil.

Enfin j'ai aussi découvert le monde de la recherche, avec la rédaction d'un état de l'art et la découverte d'un nouvel objet d'étude, mais également avec les hypothèses qui se doivent d'être formulées et les multiples remises en question qui en découle. Tout ceci m'a motivé pour continuer dans la recherche en télédétection et je poursuivrai l'étude de l'apport des séries temporelles en doctorat l'année prochaine, cette fois pour la caractérisation des milieux naturels méditerranéens.

TABLE DES FIGURES

Figure 1: Organigramme du laboratoire DYNAFOR.....	3
Figure 2 : Evolution des surface toujours en herbe de l'union européenne des 9 depuis les années 1970 (Baumont et al., 2011).....	4
Figure 3 : Interdépendance des entrées possibles pour caractériser une prairie.	5
Figure 4 : Profils moyens de NDVI des groupes issus d'une classification non supervisée, caractérisés par deux diminutions (a) et par une diminution (b) (Lang, 2015).....	10
Figure 5 : Localisation du site d'étude et image Formosat-2 du 03/03/2013, fausses couleurs (RGB, PIR, R, V).....	12
Figure 6 : Schéma fonctionnel de la démarche générale.....	13
Figure 7 : Exemple de présence de nuages et de leur ombre (entourés en vert et en noir) et de saturation (entourées en bleu) sur une image de la zone d'étude (20/07/2014).....	15
Figure 8 : Schéma de synthèse de la méthode de sélection des prairies à échantillonner.	18
Figure 9 : Photos de prairies : (a) qui ont une composition tâches (entourées rouge) et (b) qui ont une composition homogène.	25
Figure 10 : Chaîne des pré-traitements appliqués aux images satellitales.....	26
Figure 11 : Schéma explicatif de la création d'un masque de prairie.....	27
Figure 12 : Schéma explicatifs des étapes d'une classification supervisée.....	28
Figure 13 : Schéma explicatif de la méthode de validation des classifications	29
Figure 14 : Sélection des prairies issues du RPG (entourées en vert), extrait d'image Formosat-2 du 03/03/2015 en fausses couleurs (RGB: PIR, R, V).	35
Figure 15 : Profils de NDVI des centres des classes de la classification supervisée sur les modes de conduite	39
Figure 16 : Profils de NDVI des centres des classes de la classification supervisée sur les productivités potentiels. Les barres d'erreurs correspondent à l'écart-type des classes à chaque date.	40
Figure 17 : Profils de NDVI des centres des classes de la classification supervisée sur la richesse en formes de vie. Les barres d'erreurs correspondent à l'écart-type des classes à chaque date.	42
Figure 18 : Profils de NDVI des centres des classes de la classification supervisée sur la phénologie. Les barres d'erreurs correspondent à l'écart-type des classes à chaque date.	43
Figure 19 : Schéma de synthèse des différents facteurs pouvant expliquer la faible qualité des classifications réalisées.....	44

TABLE DES TABLEAUX

Tableau 1 : Synthèse des variables de gestion (a) et botanique (b) à déterminer et les variables spectrales associées.....	10
Tableau 2 : Caractéristiques des capteurs du satellite Formosat-2, (Site d'Airbus Defense and Space, 10 juin 2015).	14
Tableau 3 : Dates disponibles pour l'année 2013 et présence de nuages.	15
Tableau 4 : Caractéristiques des données attributaires des couches du RPG disponible.	16
Tableau 5 : Echelle de notation d'abondance dominance inspiré de Braun-Blanquet (1965).	19
Tableau 6 : Correspondances entre les dates d'acquisition des images satellites pour l'année 2013 et la somme des températures en °C.jours.	20
Tableau 7 : Classes de richesse en formes de vie.	22
Tableau 8 : Caractérisation de la productivité potentielle et de la précocité à partir du type fonctionnel de graminées pérennes (Cruz et al., 2010).....	22
Tableau 9 : Nombre de parcelles en fonction du mode de conduite et de la réalisation de relevés botaniques.....	23
Tableau 10 :Nombre de parcelles et le pourcentage de pixels correspondant en fonction des classes de richesse en formes de vie(a), des classes de potentiel de productivité (b), des classes de précocité (c) et des classes de productivité et précocité combinées(d). « Gram » = Graminées, « Leg » = Légumineuses, « div » = Diverses.....	24
Tableau 11 : Exemple de matrice de confusion.....	30
Tableau 12 : Variables à expliquer dans les modèles linéaires (en pourcentage du recouvrement total).....	31
Tableau 13 : Combinaisons de variables explicatives testées dans les modèles linéaires.....	32
Tableau 14 : Matrice de confusion pour la classification supervisée sur les modes de conduite. Les résultats sont normalisés par rapport aux données terrain (moyennes de validations croisées sur 30 itérations, avec l'écart-type associé entre parenthèse).....	36
Tableau 15 : Indices de qualité globale des différentes classifications supervisées sur les classes issues de la composition botanique (moyennes de 30 itérations, avec l'écart-type associé entre parenthèse).	36
Tableau 16 : Précision réalisateur et utilisateur pour chaque classes de la classification supervisée sur la richesse de formes de vie (a), la productivité potentielle (b), la phénologie (c), sur la productivité potentielle et la phénologie (d)(moyennes de 30 itérations, avec l'écart-type associé entre parenthèse).	37
Tableau 17 : Coefficient de détermination les plus significatifs parmi les 174 modèles réalisés	38

BIBLIOGRAPHIE

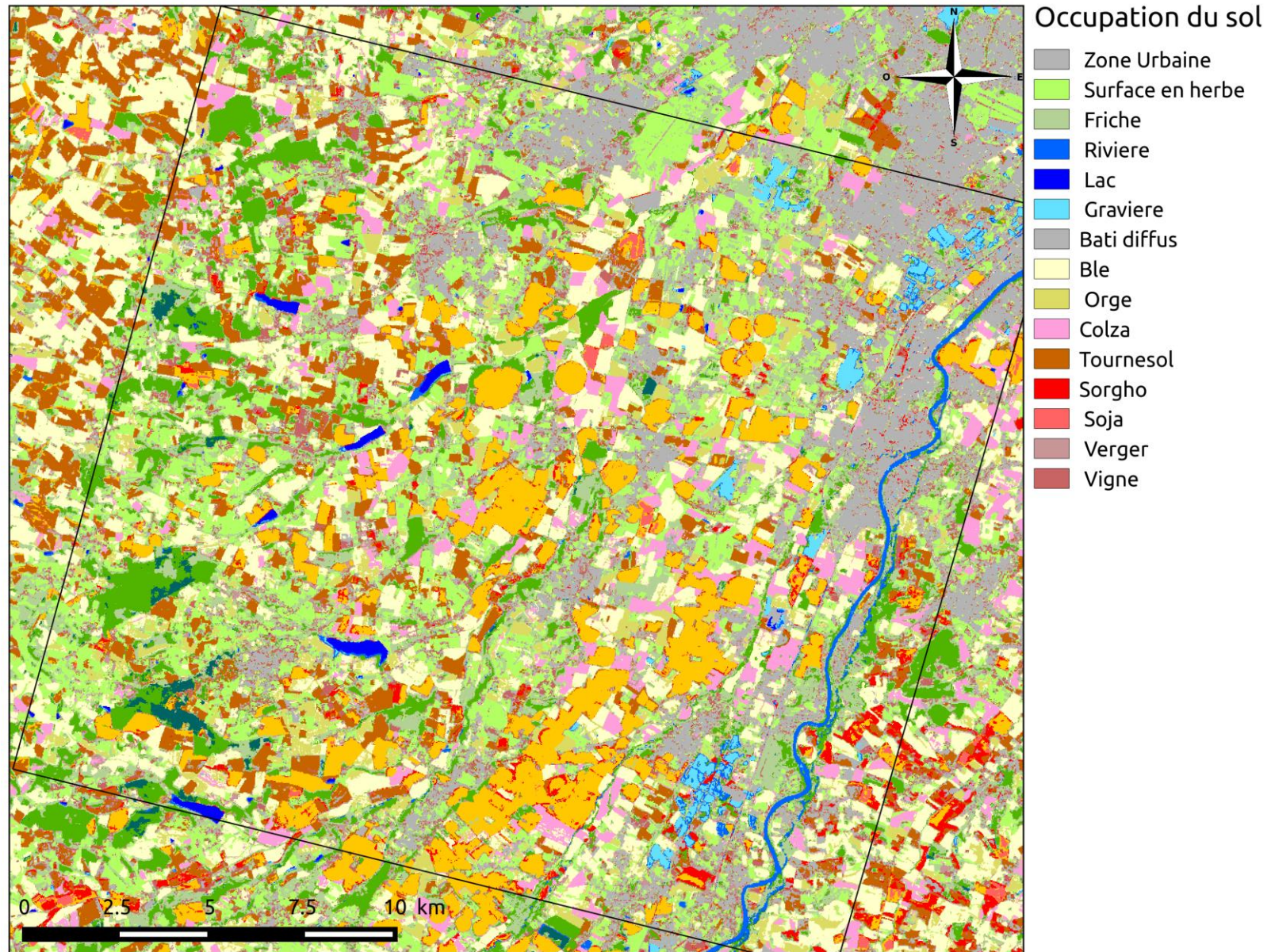
- Boyd, D.S., Danson, F.M., 2005. Satellite remote sensing of forest resources: three decades of research development. *Progress in Physical Geography* 29, 1–26.
- Braun-Blanquet, J., 1965. *Plant Sociology. The study of plant communities*. Hafner, London.
- Cingolani, A.M., Renison, D., Zak, M.R., Cabido, M.R., 2004. Mapping vegetation in a heterogeneous mountain rangeland using landsat data: an alternative method to define and classify land-cover units. *Remote Sensing of Environment* 92, 84–97. doi:10.1016/j.rse.2004.05.008
- Cruz, P., Theau, J.P., Lecloux, E., Jouany, C., Duru, M., 2010. Typologie fonctionnelle de graminées fourragères pérennes: une classification multitraits. *Fourrages* 201, 11–17.
- Duru, M., Cruz, P., Theau, J.P., Jouany, C., Ansquer, P., Al Haj Khaled, R., Therond, O., 2007. Typologies de prairies riches en espèces en vue d'évaluer leur valeur d'usage: bases agro-écologiques et exemples d'application. *Fourrages* 192, 453–476.
- Dusseux, P., 2014. Exploitation de séries temporelles d'images satellites à haute résolution spatiale pour le suivi des prairies en milieu agricole. Rennes 2.
- Dusseux, P., Hubert-Moy, L., Lecerf, R., Gong, X., Corpetti, T., 2011. Identification of grazed and mown grasslands using a time series of high-spatial-resolution remote sensing images, in: *Analysis of Multi-Temporal Remote Sensing Images (Multi-Temp)*, 2011 6th International Workshop on the. Presented at the Analysis of Multi-temporal Remote Sensing Images (Multi-Temp), 2011 6th International Workshop on the, pp. 145–148. doi:10.1109/Multi-Temp.2011.6005069
- Dusseux, P., Vertès, F., Corpetti, T., Corgne, S., Hubert-Moy, L., 2014. Agricultural practices in grasslands detected by spatial remote sensing. *Environ Monit Assess* 186, 8249–8265. doi:10.1007/s10661-014-4001-5
- Eilers, P.H., 2003. A perfect smoother. *Analytical chemistry* 75, 3631–3636.
- Fava, F., Parolo, G., Colombo, R., Gusmeroli, F., Della Marianna, G., Monteiro, A.T., Bocchi, S., 2010. Fine-scale assessment of hay meadow productivity and plant diversity in the European Alps using field spectrometric data. *Agriculture, Ecosystems & Environment*, Special section Harvested perennial grasslands: Ecological models for farming's perennial future 137, 151–157. doi:10.1016/j.agee.2010.01.016
- Feilhauer, H., Thonfeld, F., Faude, U., He, K.S., Rocchini, D., Schmidlein, S., 2013. Assessing floristic composition with multispectral sensors—A comparison based on monotemporal and multiseasonal field spectra. *International Journal of Applied Earth Observation and Geoinformation* 21, 218–229. doi:10.1016/j.jag.2012.09.002
- Franke, J., Keuck, V., Siegert, F., 2012. Assessment of grassland use intensity by remote sensing to support conservation schemes. *Journal for Nature Conservation* 20, 125–134. doi:10.1016/j.jnc.2012.02.001
- Girard, M.C., Girard, C.M., 1999. *Traitement des données de télédétection*. Dunod.
- Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment* 80, 76–87. doi:10.1016/S0034-4257(01)00289-9
- Hadj Said, M., Laghrouche, M., Malek, M., Hadria, R., Oliosio, A., Courault, D., Ameer, S., 2011. Détection supervisée des dates des opérations de fauches des prairies irriguées dans la plaine de la crau, France. *Revue Télédétection* 10, 129–136.
- Hagolle, O., Dedieu, G., Mougenot, B., Debaecker, V., Duchemin, B., Meygret, A., 2008. Correction of aerosol effects on multi-temporal images acquired with constant viewing angles: Application to Formosat-2 images. *Remote Sensing of Environment, Remote Sensing Data Assimilation Special Issue* 112, 1689–1701. doi:10.1016/j.rse.2007.08.016

- Hastie, T.J., Tibshirani, R.J., Friedman, J.H., 2013. The elements of statistical learning: data mining, inference, and prediction, 2. ed., corr. at 7. printing. ed, Springer series in statistics. Springer, New York, NY.
- Homolová, L., Malenovský, Z., Clevers, J.G.P.W., García-Santos, G., Schaepman, M.E., 2013. Review of optical-based remote sensing for plant trait mapping. *Ecological Complexity* 15, 1–16. doi:10.1016/j.ecocom.2013.06.003
- Homolová, L., Schaepman, M.E., Lamarque, P., Clevers, J.G.P.W., de Bello, F., Thuiller, W., Lavorel, S., 2014. Comparison of remote sensing and plant trait-based modelling to predict ecosystem services in subalpine grasslands. *Ecosphere* 5, art100. doi:10.1890/ES13-00393.1
- Hubert, F., Pierre, P., 2003. Guide pour un diagnostic prairial, Chambre d'agriculture. ed.
- Knox, N.M., Skidmore, A.K., van der Werff, H.M.A., Groen, T.A., de Boer, W.F., Prins, H.H.T., Kohi, E., Peel, M., 2013. Differentiation of plant age in grasses using remote sensing. *International Journal of Applied Earth Observation and Geoinformation* 24, 54–62. doi:10.1016/j.jag.2013.02.004
- Lang, M., 2015. Classification des types de prairies et estimation de la diversité taxonomique à partir de séries temporelles d'images satellites. Master en géomatique. Toulouse. Université Toulouse II.
- Lecerf, R., 2008. Suivi des changements d'occupation et d'utilisation des sols d'origine anthropique et climatique à l'échelle régionale par télédétection moyenne résolution (application à la Bretagne) (phdthesis). Université Rennes 2.
- Le Roux, X., Barbault, R., Baudry, J., Burel, F., Doussan, I., Garnier, E., Herzog, F., Lavorel, S., Lifran, R., Roger-Estrade, J., 2008. Agriculture et biodiversité. Valoriser les synergies. Expertise scientifique collective, synthèse du rapport, INRA (France).
- Mansour, K., Mutanga, O., Everson, T., Adam, E., 2012. Discriminating indicator grass species for rangeland degradation assessment using hyperspectral data resampled to AISA Eagle resolution. *ISPRS Journal of Photogrammetry and Remote Sensing* 70, 56–65. doi:10.1016/j.isprsjprs.2012.03.006
- Mauchamp, L., Gillet, F., Mouly, A., Badot, P.-M., 2012. Les prairies : biodiversité et services écosystémiques. Presses universitaires de Franche-Comté, Université de Franche-Comté.
- Plantureux, S., Bonischot, R., Guckert, A., 1992. Utilisation d'une typologie des prairies permanentes du Plateau lorrain pour le diagnostic agronomique. *Fourrages* 132, 381–394.
- Régis Caloz, Collet, C., 2001. Traitements numériques d'images de télédétection, PUQ/AUF. ed, Précis de télédétection.
- René Baumont, Plantureux, S., Farrié, J.-P., Launay, F., Michaud, A., Pottier, E., 2011. Prairies permanentes, des références pour valoriser leur diversité, Institut de l'Élevage. ed.
- Robinson, R.A., Sutherland, W.J., 2002. Post-war changes in arable farming and biodiversity in Great Britain. *Journal of Applied Ecology* 39, 157–176. doi:10.1046/j.1365-2664.2002.00695.x
- Seebach, L.M., Strobl, P., Miguel-Ayanz, J.S., Bastrup-Birk, A., 2011. Identifying strengths and limitations of pan-European forest cover maps through spatial comparison. *International Journal of Geographical Information Science* 25, 1865–1884. doi:10.1080/13658816.2011.562211
- Shen, M., Chen, J., Zhu, X., Tang, Y., Chen, X., 2010. Do flowers affect biomass estimate accuracy from NDVI and EVI? *International Journal of Remote Sensing* 31, 2139–2149. doi:10.1080/01431160903578812
- Theau, J.P., Cruz, P., Fallour, D., Jouany, C., Lecloux, E., Duru, M., 2010. Une méthode simplifiée de relevé botanique pour une caractérisation agronomique des prairies permanentes. *Fourrages* 401, 19–25.
- Theau, J.P., Zerourou, A., 2008. Herb'âge, une méthode de calcul des sommes de températures pour la gestion des prairies, in: Symposium International INRA-UMR AGIR

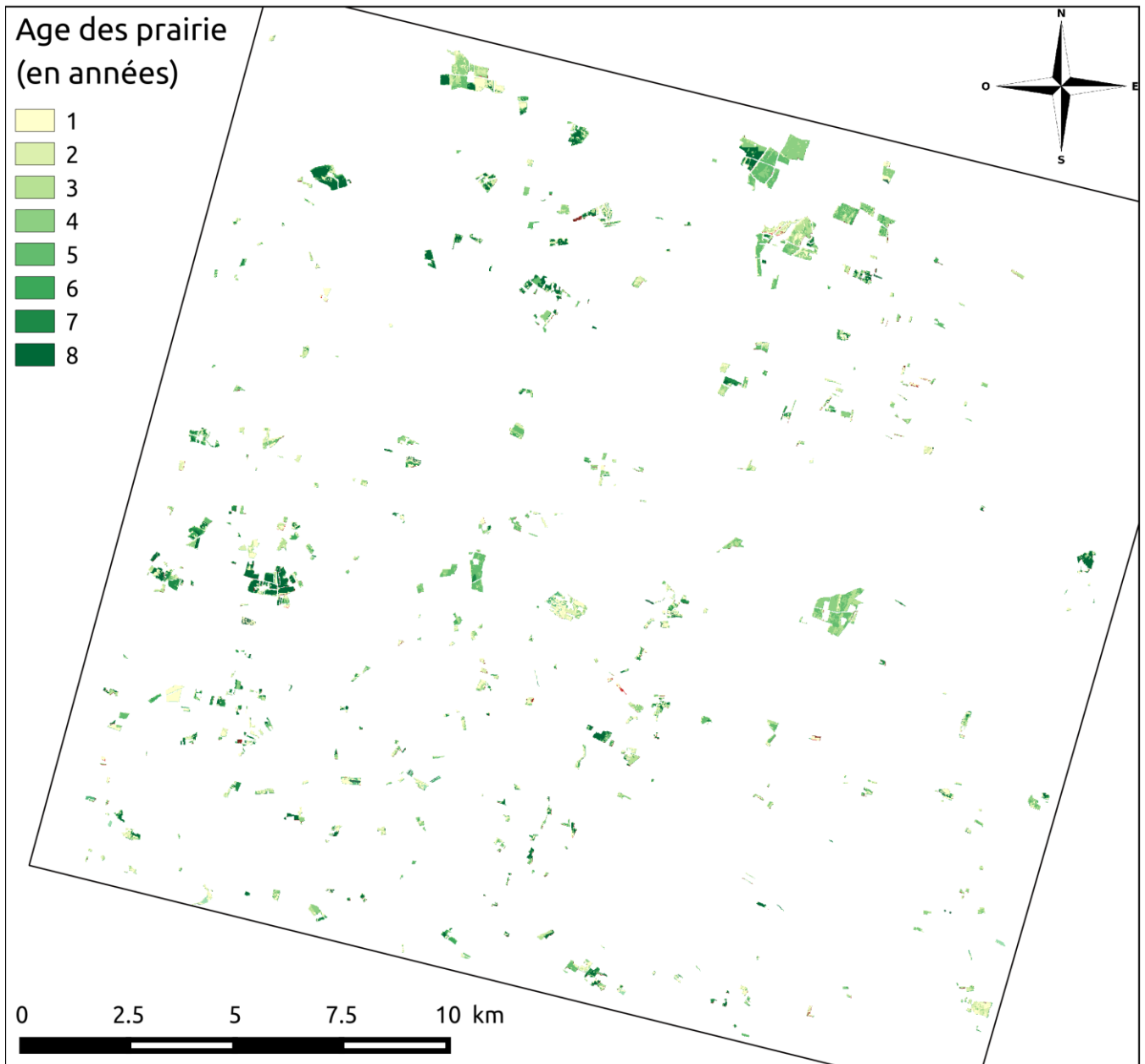
- (Agrosystèmes et Développement Territorial): Outils Pour La Gestion Des Prairies Permanentes. 2005-07-062005-07-08, Castanet-Tolosan, FRA.
- Vannier, C., 2012. Analyse spatiale de structures paysagères en contexte agricole bocager. *Cybergeo : European Journal of Geography*. doi:10.4000/cybergeo.25354
- Vignau-Loustau, L., Huyghe, C., 2008. Stratégies fourragères. France Agricole Editions.
- Viña, A., Tuanmu, M.-N., Xu, W., Li, Y., Qi, J., Ouyang, Z., Liu, J., 2012. Relationship between floristic similarity and vegetated land surface phenology: Implications for the synoptic monitoring of species diversity at broad geographic regions. *Remote Sensing of Environment* 121, 488–496. doi:10.1016/j.rse.2012.02.013
- Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., Garnier, E., 2007. Let the concept of trait be functional! *Oikos* 116, 882–892. doi:10.1111/j.0030-1299.2007.15559.x
- Vivier, M., 1971. Les prairies permanentes du Bessin et du Pays d'Auge. France.

ANNEXES

ANNEXE 1 : Carte d'occupation du sol pour l'année 2013, fournie par le CESBIO (zone d'étude entourée en noir).....	54
ANNEXE 2 : Carte des âges de prairies, établie à partir de carte d'occupation du sol.....	55
ANNEXE 3 : Fiche de terrain pour un travail d'une demi-journée (zone de travail entourée en noire) localisant les prairies (entourées en rouge) sur lesquelles effectuer des relevés botaniques. Fond de carte issu de la BD Topo IGN (2013).....	56
ANNEXE 4 : Fiche de terrain détaillant la localisation des prairies (entourées en rouge) sur lesquelles effectuer des relevés botaniques. Fond de carte issu de la BD Ortho IGN (2010).....	57
ANNEXE 5 : Résultats de la campagne de terrain : recouvrement (en %) des différentes formes de vie et des types fonctionnels de graminées.....	59
ANNEXE 6 : Exemple de <i>docstring</i> d'une fonction développée pendant le stage. Dans le code (a) et formatée automatiquement pour les utilisateurs (b).....	60
ANNEXE 7 : Code développé pour une chaîne de traitement de classification non supervisée de type GMM. La première fonction fait appel à des fonctions qui sont présentées à la suite.....	61
ANNEXE 8 : Matrices de confusion des classifications supervisées sur les classe de précocité (a), les classes de potentiel de productivité (b), les classes de richesse en forme de vie (c), et les classes de précocité et potentiel de productivité (d). Les résultats sont normalisés par rapport aux données terrain (moyennes de 30 itérations, avec l'écart-type associé entre parenthèse).....	70
ANNEXE 9 : Classification supervisée des prairies selon leur mode de conduite.....	72
ANNEXE 10 : Classification supervisée des prairies selon leur richesse en formes de vie.....	73
ANNEXE 11 : Classification supervisée des prairies selon leur précocité.....	74
ANNEXE 12 : Classification supervisée des prairies selon leur potentiel de productivité.....	75
ANNEXE 13 : Classification supervisée des prairies selon leur précocité et leur potentiel de productivité .	76
ANNEXE 14 : Coefficient de détermination R^2 obtenus pour avec des modèles linéaires réalisés avec l'algorithme de Lasso (a), de Ridge (b) et OMP (c).....	77



ANNEXE 1 : Carte d'occupation du sol pour l'année 2013, fournie par le CESBIO (zone d'étude entourée en noir).

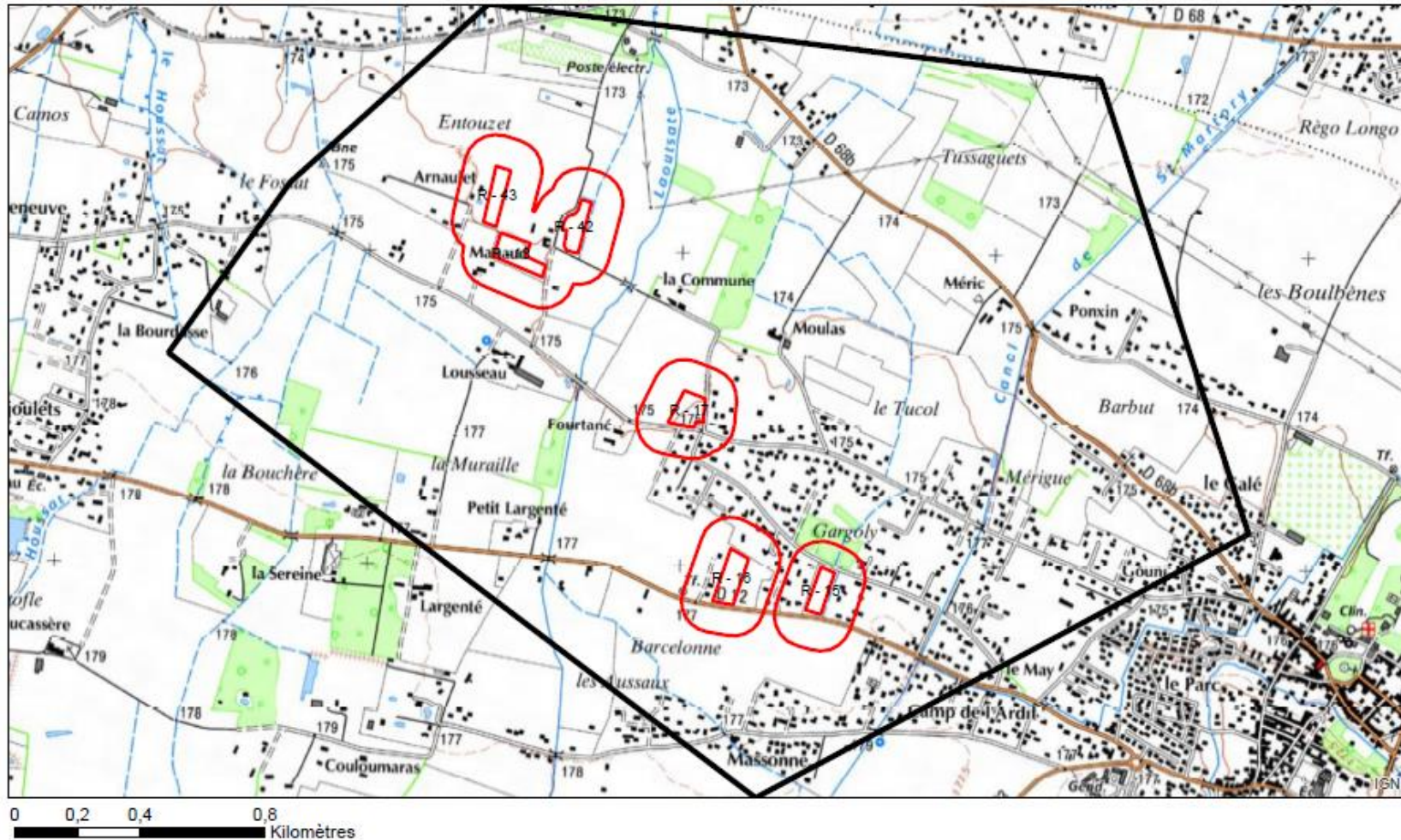


ANNEXE 2 : Carte des âges de prairies, établie à partir de carte d'occupation du sol.

Relevés bota prairie, mai-avril 2015
Région de St-Lys - Rieumes

Qui ? : _____ Quand ? : _____

Jour n° 41



ANNEXE 3 : Fiche de terrain pour un travail d'une demi-journée (zone de travail entourée en noire) localisant les prairies (entourées en rouge) sur lesquelles effectuer des relevés botaniques. Fond de carte issu de la BD Topo IGN (2013)

**Relevés bota prairie, mai-avril 2015
Région de St-Lys - Rieumes**

Qui ? : _____

Quand ? : _____

risque : non

Vue n° 12



0 25 50 100 150 200
Mètres

ANNEXE 4 : Fiche de terrain détaillant la localisation des prairies (entourées en rouge) sur lesquelles effectuer des relevés botaniques. Fond de carte issu de la BD Ortho IGN (2010)

Id	Recouvrement total	Graminées	légumineuses	diverses	type A	type B	type b	type C	type D
1	180	50.3	0.0	49.7	0.0	1.7	0.0	0.0	0.0
5	171.5	73.8	0.0	26.2	59.8	5.2	0.0	8.7	0.0
6	151	88.1	2.0	9.9	51.3	11.9	0.0	24.8	0.0
7	179	83.2	0.0	16.8	43.3	5.0	0.0	34.9	0.0
8	112.5	46.7	16.0	37.3	2.7	41.3	0.0	2.7	0.0
9	148.5	71.7	0.0	28.3	12.1	22.2	25.3	10.1	0.0
10	182.5	62.2	0.0	37.8	16.4	9.9	34.2	0.0	0.0
11	208	33.9	4.3	61.8	25.2	0.0	1.4	7.2	0.0
12	199	36.9	4.5	58.5	9.0	7.5	18.8	0.0	0.0
13	346.5	41.3	29.7	29.0	18.0	4.3	18.0	0.0	0.0
14	303.5	58.0	2.0	40.0	9.9	28.8	12.4	1.0	0.0
15	84	42.9	3.6	53.6	7.1	10.7	3.6	0.0	0.0
16	170.5	63.0	1.8	35.2	45.5	0.0	8.8	8.8	0.0
17	144	41.7	12.5	45.8	12.5	12.5	10.4	0.0	0.0
18	289.5	50.3	1.0	48.7	18.1	0.0	13.0	0.0	0.0
19	218.5	69.8	6.9	23.3	45.8	17.2	0.0	0.0	0.0
20	245	53.5	7.3	39.2	26.7	25.5	0.0	0.0	0.0
21	150.5	70.1	0.0	29.9	0.0	68.1	2.0	0.0	0.0
22	143	67.5	0.0	32.5	2.1	63.3	2.1	0.0	0.0
24	96.5	90.7	0.0	9.3	0.0	90.7	0.0	0.0	0.0
25	279.5	55.5	0.0	44.5	27.7	27.7	0.0	0.0	0.0
26	223	58.3	13.5	28.3	28.0	23.5	0.0	6.7	0.0
27	90.5	96.7	0.0	3.3	0.0	96.7	0.0	0.0	0.0
28	90.5	96.7	0.0	3.3	0.0	96.7	0.0	0.0	0.0
30	183	49.5	1.6	48.9	1.6	47.8	0.0	0.0	0.0
31	186	48.7	0.0	51.3	0.0	48.7	0.0	0.0	0.0
32	369	55.7	16.3	28.0	0.0	27.8	16.9	0.0	0.0
33	187	36.6	8.0	55.3	0.0	35.0	0.0	0.0	0.0
34	337.5	57.3	11.6	31.1	0.0	26.8	18.5	0.0	0.0
42	169.5	56.6	0.0	43.4	8.8	23.9	23.9	0.0	0.0
43	210	51.7	0.0	48.3	7.1	1.4	41.7	0.0	0.0
44	207.5	74.0	14.5	11.6	23.1	7.2	42.2	0.0	0.0
45	182.5	78.6	9.9	11.5	18.1	0.0	35.9	8.2	0.0
46	170.5	61.3	1.8	37.0	45.5	7.0	8.8	0.0	0.0
47	182	80.2	9.9	9.9	56.3	1.6	20.6	0.0	0.0
48	136.5	89.0	4.4	6.6	29.7	2.2	29.7	27.5	0.0
49	127.5	52.9	2.4	44.7	23.5	29.4	0.0	0.0	0.0
50	127.5	50.6	2.4	47.1	16.5	0.0	2.4	0.0	0.0

51	240	45.2	7.5	47.3	0.0	1.3	44.0	0.0	0.0
52	192	54.9	1.6	43.5	0.0	1.6	53.4	0.0	0.0
53	172	48.5	12.2	39.2	1.7	0.0	45.1	0.0	0.0
54	152.5	44.9	11.8	43.3	2.0	0.0	43.0	0.0	0.0
55	399	44.7	28.2	27.1	13.2	3.8	9.4	1.5	0.0
56	384.5	30.3	30.7	39.0	17.0	3.9	0.8	3.9	0.0
63	147.5	65.4	0.0	34.6	2.0	59.3	2.0	0.0	0.0
64	216	42.4	1.4	56.3	13.9	18.8	6.9	0.0	0.0
65	365	49.9	28.8	21.4	4.9	21.2	17.1	0.8	0.0
66	230.5	22.1	33.6	44.3	13.0	2.6	6.5	0.0	0.0
67	250	42.4	7.2	50.4	26.2	0.0	0.0	0.0	0.0
68	129	41.9	2.3	55.8	14.0	0.0	14.0	0.0	0.0
69	136	86.8	2.2	11.0	38.6	2.2	46.0	0.0	0.0

ANNEXE 5 : Résultats de la campagne de terrain : recouvrement (en %) des différentes formes de vie et des types fonctionnels de graminées.

```

def load_data(ids_prairie, root_folder, nb_date, nb_axe, gmm_par = ([],[],[])):
    """
    Load the sample data from files in one sample matrix. Differents combinations
    are stocked in a dictionary.

    Parameters
    -----
    ids_prairie : list of int,
        Id of the meadows to be taken into account in the model
    root_folder : string,
        Path of the folder that contains all the folders of the meadows (named
        by their id)
    nb_date : int,
        The n first dates of temp serie
    nb_axe : int,
        The n first eigen vectors to load
    gmm_par : list of list ``[[idx_meadows], [nb_opt_clust], [idx_clust]]``,
        Contains the meadows for which the optimal number of cluster is different
        from 1. Defaults to none. ::
        idx_meadows : list of int,
            Id of the meadows
        nb_opt_clust : list of int,
            Number of clusters optimal
        idx_clust : list of int,
            Index of the cluster to take into account for the model
            (generally because it's majoritary)

    Returns
    -----
    dict_expl_var : dict
        Dictionary that contains the differents combination of the explicative
        variable

    See Also
    -----
    load_expl_var()
    """

```

(a)

load_data

Definition : load_data(ids_prairie, root_folder, nb_date, nb_axe, gmm_par=([],[],[]))

Type : Present in Scripts.function_glm module

Load the sample data from files in one sample matrix. Differents combinations are stocked in a dictionary.

Parameters

ids_prairie : list of int,
Id of the meadows to be taken into account in the model

root_folder : string,
Path of the folder that contains all the folders of the meadows (named by their id)

nb_date : int,
The n first dates of temp serie

nb_axe : int,
The n first eigen vectors to load

gmm_par : list of list [[idx_meadows], [nb_opt_clust], [idx_clust]].

Contains the meadows for which the optimal number of cluster is different from 1. Defaults to none.

```

idx_meadows : list of int,
    Id of the meadows
nb_opt_clust : list of int,
    Number of clusters optimal
idx_clust : list of int,
    Index of the cluster to take into account for the model
    (generally because it's majoritary)

```

Returns

dict_expl_var : dict
Dictionary that contains the differents combination of the explicative variable

See Also

load_expl_var()

(b)

ANNEXE 6 : Exemple de *docstring* d'une fonction développée pendant le stage. Dans le code (a) et formatée automatiquement pour les utilisateurs (b).

ANNEXE 7 : Code développé pour une chaîne de traitement de classification non supervisée de type GMM. La première fonction fait appel à des fonctions qui sont présentées à la suite.

```
# -*- coding: utf-8 -*-
"""
Created on Wed Apr  1 13:58:20 2015

@author: mlang
"""

import glob
from whittaker_script import whittakerSmoother
import scipy as sp
from osgeo import gdal
from sklearn.cluster import KMeans
from sklearn import mixture
import matplotlib.pyplot as plt
import os
import datetime as dt
from time import strftime
from function_data_raster import write_data, open_data
import otbApplication as otb

def gmm_classification(*args,**kwargs):
    """
    This function perform multiple gmm classications from an image and a sample
    image, depending on the number of clusters wanted. Only the pixels
    different from 0 on the sample image will be classified. If specified,
    a whittaker smoothing application can be applied and the
    temporal profile of cluster means can be plotted. Also, for each
    classification, the means, the covariances, the labels, the weighs, the
    responsabilites and the AIC and BIC can be saved for each classification as
    npdy file.

    Parameters
    -----
    name_file : string
        Path of the image file
    roi_file : string
        Path of the sample image file
    out_folder : string
        Path of the out folder
    rg_cluster : list of integer
        Range of different cluster numbers that should be classified, e.g. if
        rg_cluster = range(2,11), 9 classifications will be performed for a
        number of clusters from 2 to 10 (included)
    bands : list of integer, optional, defaults to None
        The bands of the raster_name file whose value should be extracted.
        Indexation starts at 0. By defaults, all the bands will be extracted
    value_to_extract : float, optional, defaults to None
        If specified, the pixels classified will be only those whose label is
        equal this value. By, defaults all the pixels different from zero are
        extracted.
    out_filename : string, optional
        First part of the classifications name. It'll end by
        _{i}clust_{covtype}.tif where i is the number of clusters and covtype
        is the covariance type. Default to 'gmm_classification'.
    out_plotname : string, optional
        First part of the temporal profile plot name. It'll end by
        _{i}clust_{covtype}.jpg where i is the number of clusters and covtype
        is the covariance type. Default to 'gmm_temporal_profile'.
    no_data : list of integer, optional
        List of no data values that should be set into nan value. Default to
    """
```

```

None.
smooth : bool, optional
    If true, the data ara smoothed. This operation permit to fill the
    nodata values by applying a whittaker smoother. Defaults to False.
l_smooth : int, optional
    Smoothing parameter.Default to 1000.
d_smooth : int, optional
    Order of divided differences. Default to 2.
covariance_type : string, ``{'spherical','tied','diag','full'}``,optional
    String describing the type of covariance parameters to use. ::

        Shape of covariance matrix          covariance_type
        (n_components, n_features)          if 'spherical',
        (n_features, n_features)           if 'tied',
        (n_components, n_features)         if 'diag',
        (n_components, n_features, n_features) if 'full'
    Defaults to 'diag'.
n_init : int, optional
    Number of initializations to perform. The best results is kept.
    Defaults to 1.
save : list of string, optional
    List that contains the parameters of the classification to be saved in
    the npy format. The parameters are the followings ::
        mean      : mean of each cluster
        covars    : covariance matrix
        labels    : labels of the pixels classified
        weights   : proportions of each cluster
        logprob   : logprob of each cluster
        respons   : responsibilities of each cluster
        aic       : aic of the classification
        bic       : bic of the classification
        all       : if specified, all the parameters are saved
    By defaults, none of the parameters are saved.
plot_profile : bool, optional
    If true, the temporal profile of the cluster means is plotted in
    function of the dates. Defaults to True.
write_classif : bool, optional
    If true, the classification images are saved in a 'tif' file. Defaults
    to True.
verbose : bool, optional
    If true, the verbose mode is activate. Defaults to False
visible : bool, optional
    If false, the plot wont be displayed, only saved. Defaults to True

Returns
-----
aic : list of float
    List of the aic of each classification
bic : list of float
    List of the bic of each classification
"""

#Initialize the parameters
name_file = args[0]
roi_file = args[1]
out_folder = args[2]
rg_cluster = args[3]
gmm_args= {}
gmm_args['covariance_type'] = kwargs.get('covariance_type','diag')
gmm_args['n_init'] = kwargs.get('n_init',1)
val_to_extract = kwargs.get('value_to_extract', None)
bands = kwargs.get('bands', None)

```

```

out_plotname = kwargs.get('out_plotname','gmm_temporal_profile')
classif_name = kwargs.get('out_filename','gmm_classification')
smooth = kwargs.get('smooth', False)
plot_profile = kwargs.get('plot_profile', True)
write_classif = kwargs.get('write_classif',True)
v = kwargs.get('verbose',False)
visible = kwargs.get('visible',True)
save = kwargs.get('save', [])

#initialize the dictionary that will contain the output parameters to save
#and their path
params = parameter_to_save('gmm', gmm_args['covariance_type'])

#Extract the sample matrix and some parameters from the image
out = get_samples_from_roi(name_file,roi_file,
                           value_to_extract = val_to_extract,bands = bands)

X = out[0]
coord = out[2]
nr = out[3]
nc = out[4]
GeoTransform = out[6]
Projection = out[7]
if v :
    print 'Sample matrix extracted'

#Define the nodata values into nan
if kwargs.get('no_data'):
    for i in kwargs['no_data']:
        t = sp.where(X == i)
        X[t] = sp.nan

#Extracts the dates
dates = extract_date()
dd = date_as_day(dates)
if bands: #extract the bands (ie: dates) of interest
    dd = dd[bands]
    dates = dates[bands]

if v:
    print 'Dates extracted'
###Smooth the data if specified
if smooth:
    #Initialize parameters
    l_smooth = kwargs.get('l_smooth',1000)
    d_smooth = kwargs.get('d_smooth',2)
    nb_rows = X.shape[0]

    #Perform the smoothing
    for i in range(nb_rows):
        X[i,:], ocv = whittakerSmoother(dd, X[i,:], l_smooth, d_smooth)

    if v:
        print 'Data smoothed'
#initialize list of that'll contain information-criteria
aic = []
bic = []
for i in rg_cluster: #for each number of cluster
    #perform the classification and extract parameters
    gmm_args['n_components'] = i
    classifier = mixture.GMM(**gmm_args)
    classifier_fit = classifier.fit(X)
    params['mean'][1] = classifier_fit.means_

```

```

params['covars'][1] = classifier_fit.covars_
params['labels'][1] = classifier.predict(X) + 1
params['weights'][1] = classifier.weights_
aic.append(classifier_fit.aic(X))
bic.append(classifier_fit.bic(X))
params['aic'][1] = aic[-1]
params['bic'][1] = bic[-1]
params['logprob'][1], params['respons'][1] = classifier.score_samples(X)

if v:
    print 'Classification at {} cluster done'.format(i)

#save the files if indicated
if 'all' in save: #if all the parameters are to save
    for value in params.values():
        sp.save(os.path.join(out_folder,value[0].format(i)),
                sp.asarray(value[1]))
        if v :
            print 'Data saved'
elif save: #otherwise only the ones that are specified
    for par in save :
        sp.save(os.path.join(out_folder,params[par][0].format(i)),
                sp.asarray(params[par][1]))
        if v :
            print 'Data saved'

#plot the temporal profiles if indicated
if plot_profile:
    plotname = out_plotname + '_{}_clust_'.format(i)+ \
                gmm_args['covariance_type'] + '.jpg'
    plotname = os.path.join(out_folder,plotname)
    extract_temp_profile_cluster_ndvi(i, params['mean'][1], dates,
                                     file_name = plotname,
                                     classif_type = ['gmm',\
                                     gmm_args['covariance_type']])
    if not visible: #close the windows if specified
        plt.close('all')
    if v:
        print 'temp profile for {} clusters drawn'.format(i)

#Write the classification file if indicated
if write_classif:
    out_filename = classif_name + '_{}_clust_'.format(i) + \
                gmm_args['covariance_type'] + '.tif'
    out_filename = os.path.join(out_folder,out_filename)
    build_img(nr, nc,1, coord, params['labels'][1], out_filename,
             GeoTransform, Projection)
    if v:
        print 'Classification image for {} clusters written'.format(i)

#save all the aic and bic in a row if specified
if 'aic' in save or 'all' in save :
    name_aic = 'gmm_aic_serie_' + gmm_args['covariance_type'] + \
                '_clust{}_to{}'.format(rg_cluster[0],rg_cluster[-1])
    sp.save(os.path.join(out_folder,name_aic),sp.asarray(aic))
if 'bic' in save or 'all' in save :
    name_bic = 'gmm_bic_serie_' + gmm_args['covariance_type'] + \
                '_clust{}_to{}'.format(rg_cluster[0],rg_cluster[-1])
    sp.save(os.path.join(out_folder,name_bic),sp.asarray(bic))

return aic, bic

```

```

def parameter_to_save(classifier, covar_type = ''):
    """
    This fonction initialize a dictionnary that contains in keys the names of
    the output parameters, and in values their path, and their value.

    Parameters
    -----
    classifier : string : ``{'gmm','kmeans'}``
        The kind of classifier used.
    covar_type : string, optional, defaults to None.
        The covariance type for a gmm classification. It is added in the path
        name of the parameters.

    Returns
    -----
    params : dict
        Dictionnary that contains the following caracteritics of the
        parameters : ::
            'name of the parameter' : [path, value of the parameter]
        The differents parameters are : ::

            Name          Description
            ----          -
            means       : mean of each cluster
            covars      : covariance matrix
            labels      : labels of the pixels classified
            weights     : proportions of each cluster
            logprob     : logprob of each cluster
            respons     : responsibilities of each cluster
            aic         : aic of the classification
            bic         : bic of the classification
            inertia     : inertia of the classification

    """
    params = {}
    params['mean'] = [None, None]
    params['covars'] = [None, None]
    params['labels'] = [None, None]
    params['weights'] = [None, None]
    params['logprob'] = [None, None]
    params['respons'] = [None, None]
    params['aic'] = [None, None]
    params['bic'] = [None, None]
    params['inertia'] = [None, None]

    for k,v in params.items():
        v[0] = classifier + '_' + k + '_{}cluster' + covar_type

    return params

```



```

def get_samples_from_roi(raster_name,roi_name, value_to_extract = None,
                        bands = None):
    '''
    The function get the set of pixel given the thematic map. Both map should
    be of same size.

    Parameters
    -----
    raster_name : string
        The name of the raster file, could be any file that GDAL can open
    roi_name : string
        The name of the thematic image: each pixel whose values is greater
        than 0 is returned
    value_to_extract : float, optional, defaults to None
        If specified, the pixels extracted will be only those which are equal
        this value. By, defaults all the pixels different from zero are
        extracted.
    bands : list of integer, optional, defaults to None
        The bands of the raster_name file whose value should be extracted.
        Indexation starts at 0. By defaults, all the bands will be extracted.

    Returns
    -----
    X: ndarray
        The sample matrix. A nXd matrix, where n is the number of referenced
        pixels and d is the number of variables. Each line of the matrix is a
        pixel.
    Y: ndarray
        the label of the pixel
    t: tuple
        tuple of the coordinates in the original image of the pixels
        extracted
    nr : int
        Number of rows of the roi image
    nc : int
        Number of columns of the roi image
    d : int
        Number of band of the image, or the number of bands specified in the
        bands input parameter
    GeoTransform : Gdal Geotransform type
        Gdal Geotransform parameter
    Projection : Gdal Projectiontype
        Gdal Projection parameter
    '''

    ## Open data
    raster = gdal.Open(raster_name,gdal.GA_ReadOnly)
    if raster is None:
        print 'Impossible to open '+raster_name
        exit()
    roi = gdal.Open(roi_name,gdal.GA_ReadOnly)
    if roi is None:
        print 'Impossible to open '+roi_name
        exit()

    if (raster.RasterXSize != roi.RasterXSize) or \
        (raster.RasterYSize != roi.RasterYSize):
        print 'Images should be of the same size'
        exit()

    ## Get some information
    nc = roi.RasterXSize      #number of columnq

```

```

nr = roi.RasterYSize      #number of rows
d = raster.RasterCount  #number of bands
GeoTransform = roi.GetGeoTransform()
Projection = roi.GetProjection()
if not bands:
    bands = range(d)
else:
    d = len(bands)

## Initialize the output
ROI = roi.GetRasterBand(1).ReadAsArray()
if value_to_extract:
    t = sp.where(ROI == value_to_extract)
else:
    t = sp.nonzero(ROI) #coord of where the samples are

Y = ROI[t].reshape((t[0].shape[0],1)).astype('uint8')
del ROI
roi = None # Close the roi file

try:
    X = sp.empty((t[0].shape[0],d))
except MemoryError:
    print 'Impossible to allocate memory: roi too big'
    exit()

## Load the data

for i in bands:
    temp = raster.GetRasterBand(i+1).ReadAsArray()
    X[:,i] = temp[t]
del temp
raster = None # Close the raster file

return X, Y, t, nr, nc, d, GeoTransform, Projection

def extract_date():
    """
    Function/script that returns an array of dates objects from a list of
    filenames. Each filename contains a date that is transformed. The function
    is more a script than a function because many parameters inside the corpse
    depends on the filenames.

    Returns
    -----
    dates : array of date
            Array of date object
    """
    #extract the list of filename and sort it
    list_name =
glob.glob('/home/mlang/Documents/temp_SIG/Stage/data/Formosat/MS_2013/*8m.tif')
    list_of_dates=[]

    #extraction of the date and creation of date objects list
    #extraction of the date and creation of date objects list
    for filename in list_name:
        date = filename.split(os.sep)[9].split('_')[1]

list_of_dates.append(dt.date(int(date[0:4]),int(date[4:6]),int(date[6:8])))
list_of_dates.sort()

#Convert the list into an array

```

```

dates = sp.asarray(list_of_dates)

return dates

def date_as_day(dates):
    """
    This function returns an array of numbers of day since January 1 st from a
    array of date objects, e.g. for the 01-15-2013 the function returns 15.

    Parameters
    -----
    dates : array of date
        Array of the dates to count the days on

    Returns
    -----
    dates_asday : array of int
        array of the numbers of day since January 1st
    """
    #Initialization of the array
    dates_asday = sp.zeros(len(dates),dtype='int')

    #Fill the array with the number of day
    for i,date in enumerate(dates):
        dates_asday[i]=int(date.strftime('%j'))

    return dates_asday

def extract_temp_profile_cluster_ndvi(nb_cluster, cluster_centers, dates,
                                     title = None, file_name = None,
                                     classif_type = ['', ''], label = '',
                                     loc = 0):
    """
    """

    if nb_cluster <= 9:
        colors = ['b','g','r', 'c', 'm','y', 'k','0.75','#ff9a4e']
    else:
        colors = sp.rand(nb_cluster,3)
    plt.figure()
    #plt.subplot(111)
    #plt.rc('axes', color_cycle=['r', 'g', 'b', 'y'])
    for cluster in range(nb_cluster):
        serie = []
        serie = cluster_centers[cluster,:]
        if label:
            lab = label[cluster]
        else:
            lab = 'cluster n ' + str(cluster + 1)
        plt.plot_date(dates,serie,ls='solid', marker='o',
                     color= colors[cluster],
                     label = lab)

    locs, labels = plt.xticks()
    plt.setp(labels, rotation=25)
    plt.legend(fontsize = 'xx-small', loc = loc)
    plt.ylabel('NDVI')
    if title :
        plt.title(title)
    else:

```

```

plt.title('Temp profile of {} {} unsupervised classification\nfor {}
clusters'.format(classif_type[0],classif_type[1],nb_cluster))
if file_name :
    plt.savefig(file_name, format = 'jpg')

def build_img(nb_rows, nb_col,d, coord, labels, out_filename, GeoTransform,
             Projection, dt = 'int16'):
    """
    The fonction rebuild a classified image from the labels of the classified
    pixels and their initial coordinates.

    Parameters
    -----
    nb_rows : int
        Number of rows of the original image
    nb_col : int
        Number of columns of the original image
    d : int
        Third dimension of the image to build.
    coord : tuple of ndarrays
        Tuple of the coordinates of the pixels classified
    labels : array
        1d matix of pixels labels
    out_filename : string
        Path of the output filename
    GeoTransform : Gdal Geotransform type
        Gdal Geotransform parameter
    Projection : Gdal Projectiontype
        Gdal Projection parameter
    dt : numpy data_type, optional
        Numpy data type parameter. Defaults to 'int16'
    """
    #initialization of the array
    img = sp.zeros((nb_rows,nb_col,d), dt)

    #Check if the label matrix does have the proper dimensions
    labels = labels.reshape((labels.shape[0],1))\
        if len(labels.shape) == 1 \
        else labels

    for i in range(d):
        #fill the matrix with the labels
        img[coord[0],coord[1],i] = labels[:,i]

    #the function write_data doesn't work in the case of an array representing
    #1 band but with three dimension (with actually only two useful). In this
    #case, the dimension is changed from 3 to 2.
    img = img[:, :, 0] if d == 1 else img

    #write the file
    write_data(out_filename,img,GeoTransform,Projection)

```

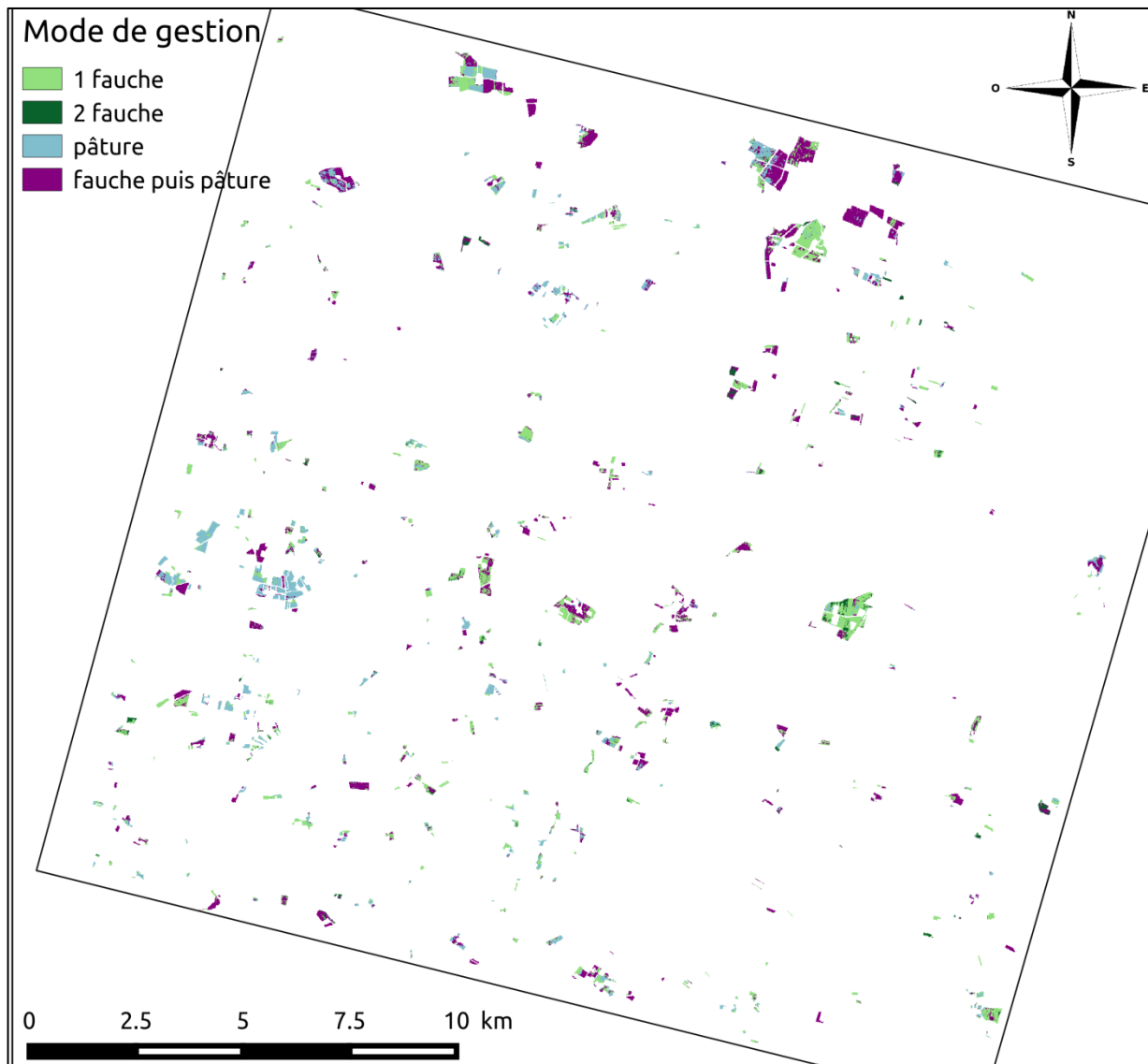
ANNEXE 8 : Matrices de confusion des classifications supervisées sur les classe de précocité (a), les classes de potentiel de productivité (b), les classes de richesse en forme de vie (c), et les classes de précocité et potentiel de productivité (d). Les résultats sont normalisés par rapport aux données terrain (moyennes de 30 itérations, avec l'écart-type associé entre parenthèse).

(a) Précocité		Classification			
		Très précoces	Précoces	Tardives	Précision réalisateur
Terrain	Très précoces	0.36	0.63	0.01	36%(±1)
	Précoces	0.11	0.84	0.05	84% (±1)
	Tardives	0.12	0.63	0.25	25% (±2)
	Précision utilisateur	64% (±2)	57% (±1)	57% (±4)	
Précision totale : 0.58 (±1) Kappa : 0.24 (±0.01)					

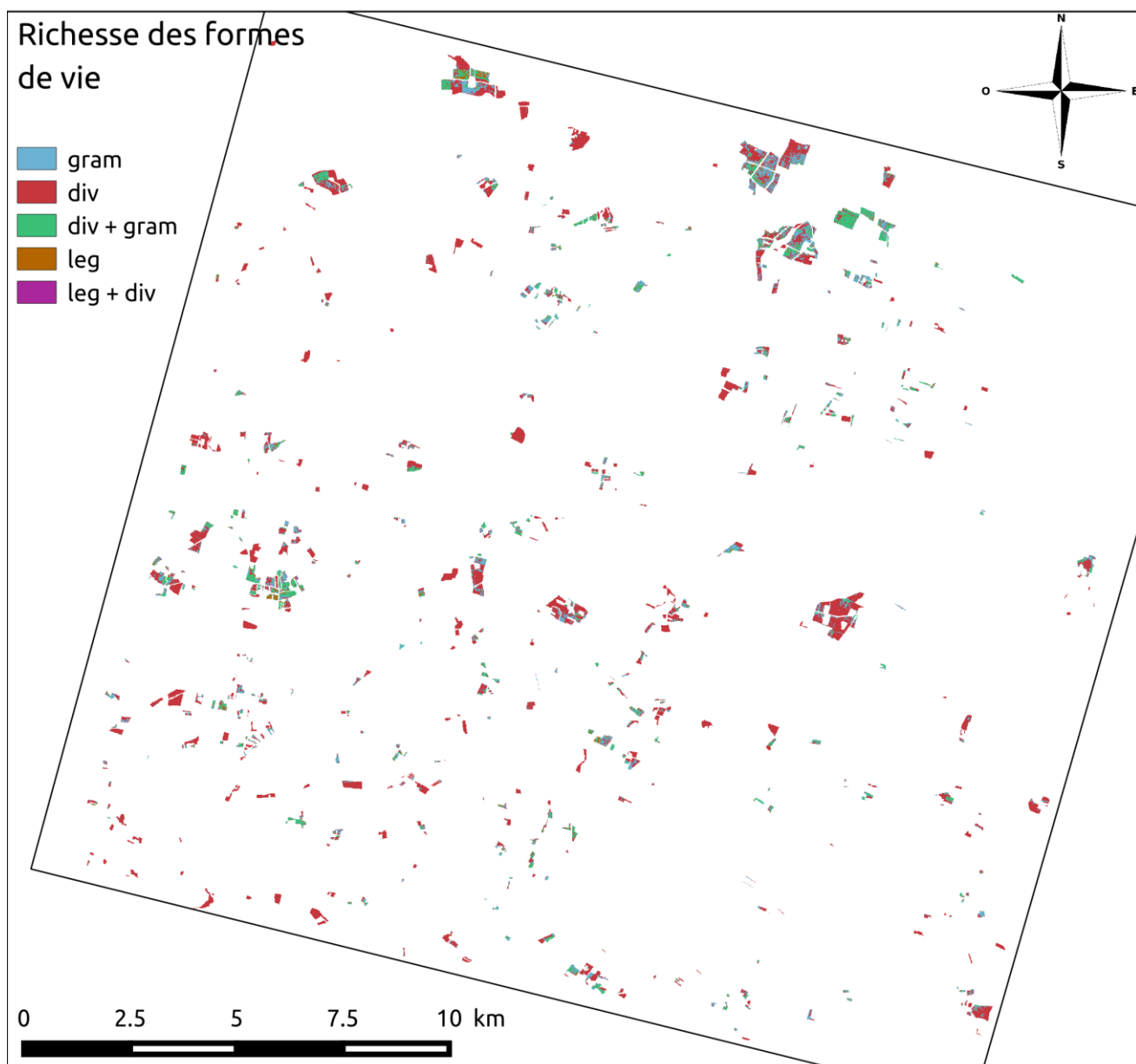
(b) Potentiel de productivité		Classification			
		Elevé	Normale	Faible	Précision réalisateur
Terrain	Elevé	0.72	0.28	0.00	72%(±1)
	Normale	0.13	0.87	0.00	87% (±2)
	Faible	0.00	1.00	0.00	0% (±0)
	Précision utilisateur	96%(±1)	38% (±1)	0% (±0)	
Précision totale : 73% (±1) Kappa : 0.47 (±0.01)					

(c) Richesse en formes de vie		Classification					Précision réalisateur
		Gram	Div	Gram + Div	Leg	Div + Leg	
Terrain	Gram	0.18	0.38	0.30	0.14	0.00	18% (±2)
	Div	0.24	0.59	0.14	0.03	0.00	59% (±1)
	Gram + Div	0.19	0.01	0.72	0.08	0.00	72% (±6)
	Leg	0.11	0.05	0.18	0.67	0.00	67% (±6)
	Div + Leg	0.13	0.71	0.14	0.03	0.00	0% (±0)
	Précision utilisateur	13% (±1)	66% (±1)	9%(±1)	22%(±3)	0%(±0)	
Précision totale : 42% (±1) Kappa : 0.37 (±0.08)							

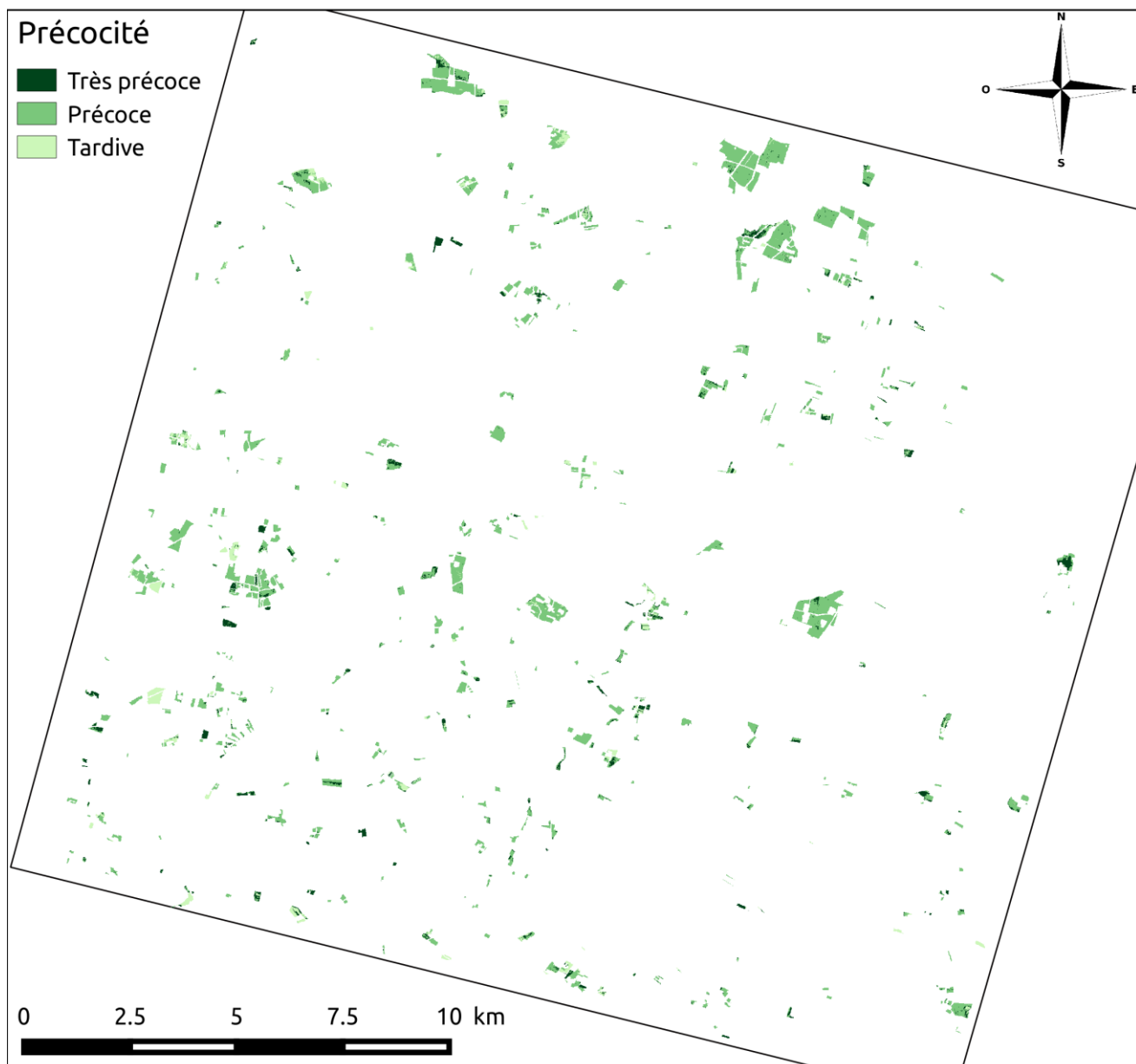
(d) Potentiel de productivité et phénologie		Classification							
		Prod3 – Prec3	Prod3 – Prec2	Prod3 – Prec1	Prod2 – Prec3	Prod2 – Prec2	Prod2 – Prec1	Prod1- Prec3	Précision réalisateur
Terrain	Prod3 – Prec3	0.52	0.40	0.00	0.08	0.00	0.00	0.00	52%(±2)
	Prod3 – Prec2	0.12	0.56	0.04	0.25	0.01	0.02	0.00	56%(±1)
	Prod3 – Prec1	0.16	0.31	0.29	0.14	0.01	0.09	0.00	29% (±2)
	Prod2 – Prec3	0.00	0.20	0.00	0.72	0.04	0.04	0.01	72% (±3)
	Prod2 – Prec2	0.00	0.34	0.00	0.43	0.19	0.04	0.00	19% (±3)
	Prod2 – Prec1	0.00	0.07	0.00	0.15	0.00	0.78	0.00	78% (±3)
	Prod1- Prec3	0.00	0.06	0.00	0.85	0.03	0.06	0.01	1% (±1)
	Précision utilisateur	62% (±2)	60% (±1)	62% (±4)	28% (±2)	53% (±9)	42% (±5)	2% (±4)	
Précision totale : 51% (±1) Kappa : 0.41 (±0.05)									



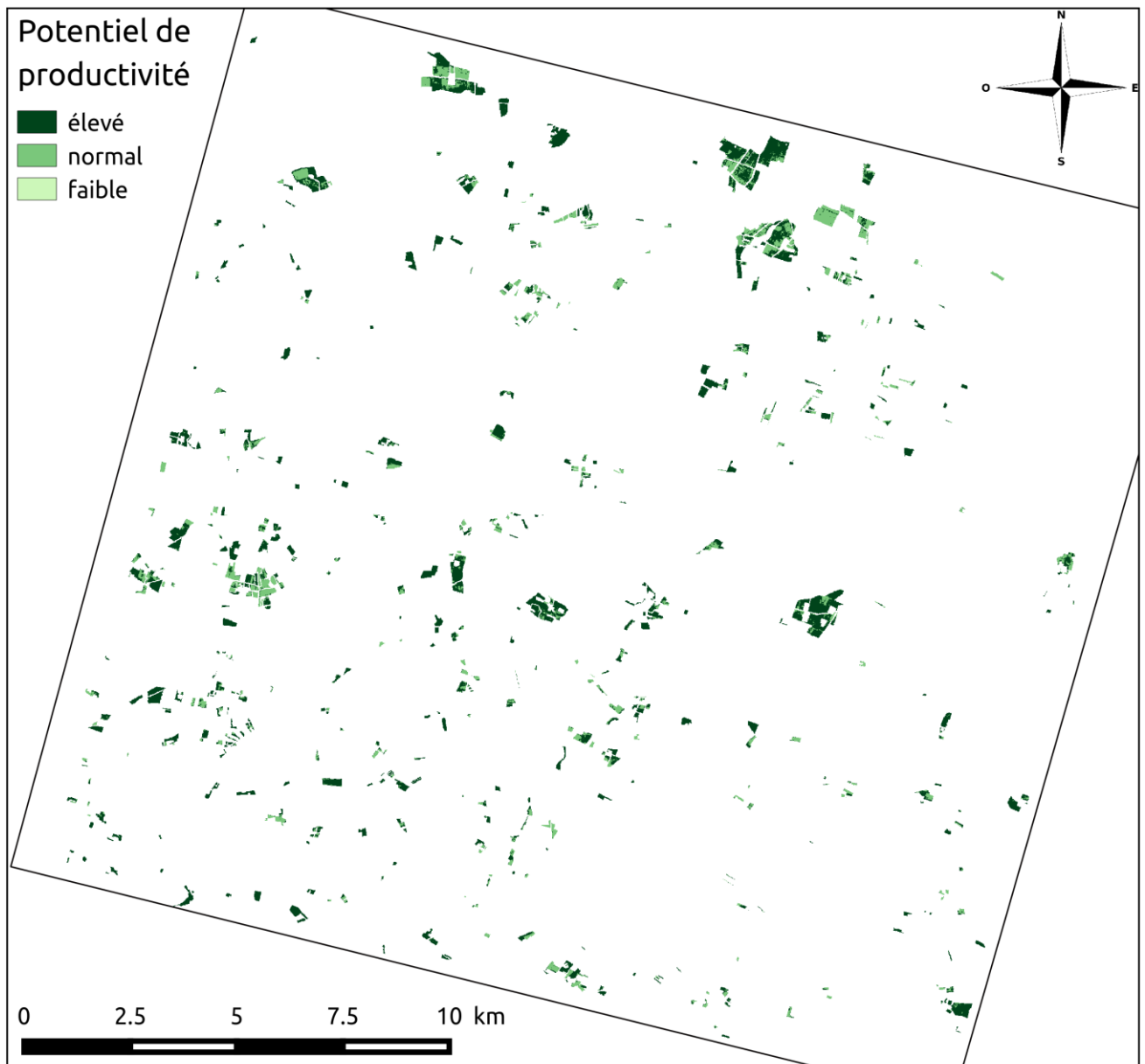
ANNEXE 9 : Classification supervisée des prairies selon leur mode de conduite.



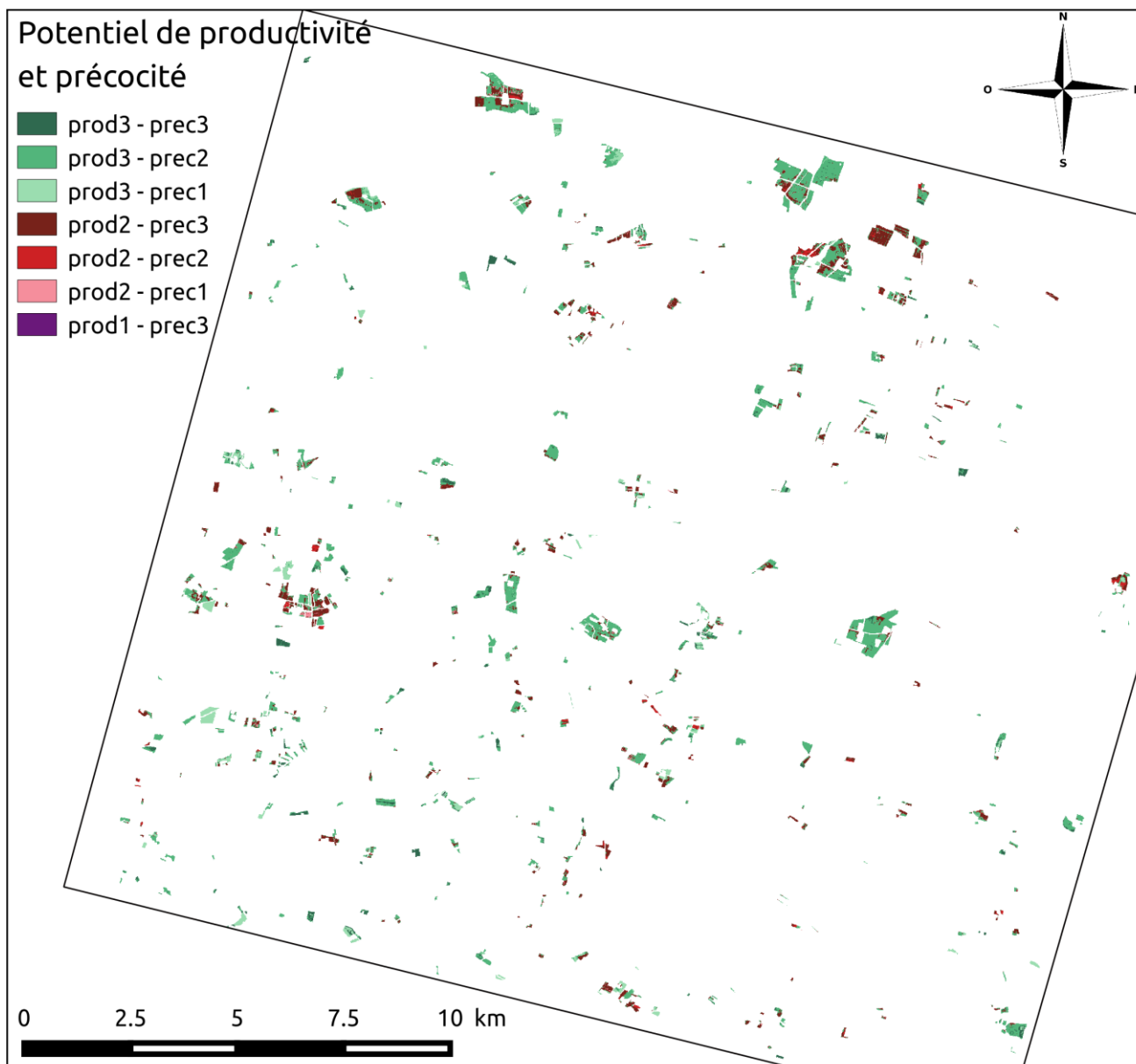
ANNEXE 10 : Classification supervisée des prairies selon leur richesse en formes de vie.



ANNEXE 11 : Classification supervisée des prairies selon leur précocité.



ANNEXE 12 : Classification supervisée des prairies selon leur potentiel de productivité



ANNEXE 13 : Classification supervisée des prairies selon leur précocité et leur potentiel de productivité

ANNEXE 14 : Coefficient de détermination R^2 obtenus pour avec des modèles linéaires réalisés avec l'algorithme de Lasso (a), de Ridge (b) et OMP (c).

(a) Lasso		Combinaisons de variables explicatives						
		moyenne	variance	coord	m+v	m+c	v+c	m+v+c
Variable à expliquer	% gram	0.12	0.06	0.29	0.12	0.00	0.13	0.00
	% leg	0.00	0.00	0.03	0.00	0.00	0.00	0.00
	% div	0.10	0.10	0.00	0.00	0.00	0.00	0.00
	% A	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	% B	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	% b	0.17	0.00	0.16	0.18	0.33	0.00	0.41
	% C	0.09	0.00	0.00	0.09	0.09	0.00	0.09

(b) Ridge		Combinaisons de variables explicatives						
		moyenne	variance	coord	m+v	m+c	v+c	m+v+c
Variable à expliquer	% gram	0.13	0.05	0.29	0.09	0.25	0.21	0.13
	% leg	0.00	0.00	0.01	0.23	0.00	0.01	0.00
	% div	0.08	0.10	0.00	0.11	0.10	0.19	0.13
	% A	0.16	0.06	0.00	0.04	0.01	0.00	0.01
	% B	0.22	0.00	0.20	0.10	0.27	0.08	0.44
	% b	0.14	0.00	0.14	0.00	0.30	0.15	0.36
	% C	0.15	0.03	0.00	0.17	0.10	0.00	0.11

(c) OMP		Combinaisons de variables explicatives						
		moyenne	variance	coord	m+v	m+c	v+c	m+v+c
Variable à expliquer	% gram	0.10	0.06	0.23	0.10	0.10	0.09	0.10
	% leg	0.16	0.01	0.18	0.27	0.07	0.20	0.27
	% div	0.12	0.10	0.03	0.12	0.12	0.08	0.12
	% A	0.16	0.03	0.03	0.06	0.06	0.03	0.06
	% B	0.02	0.00	0.09	0.02	0.09	0.09	0.09
	% b	0.17	0.01	0.18	0.09	0.18	0.18	0.39
	% C	0.15	0.04	0.01	0.15	0.15	0.04	0.15

