Multimodel Evaluation of Nitrous Oxide Emissions From an Intensively Managed Grassland

Kathrin Fuchs, Lutz Merbold, Nina Buchmann, Daniel Bretscher, Lorenzo Brilli, Nuala Fitton, Cairistiona Topp, Katja Klumpp, Mark Lieffering, Raphaël Martin, et al.

AGU100 ADVANCING EARTH AND SPACE SCIENCE

JGR Biogeosciences

RESEARCH ARTICLE
10.1029/2019JG005261

Multimodel Evaluation of Nitrous Oxide Emissions From an Intensively Managed Grassland

Kathrin Fuchs[1,2], Lutz Merbold[1,3], Nina Buchmann[1], Daniel Bretscher[4], Lorenzo Brilli[5,6], Nuala Fitton[7], Cairistiona F. E. Topp[8], Katja Klumpp[9], Mark Lieffering[10], Raphaël Martin[9], Paul C. D. Newton[10], Robert M. Rees[8], Susanne Rolinski[11], Pete Smith[7], and Val Snow[12]

[1]Department of Environmental Systems Science, Institute of Agricultural Sciences, ETH Zurich, Zürich, Switzerland, [2]Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research - Atmospheric Environmental Research, Garmisch-Partenkirchen, Germany, [3]Mazingira Centre, International Livestock Research Institute, Nairobi, Kenya, [4]Agroscope Research Station, Climate and Air Pollution, Zürich, Switzerland, [5]CNR-Ibimet, National Research Council-Institute of Biometeorology, Firenze, Italy, [6]DAGRI, University of Florence, Firenze, Italy, [7]Institute of Biological and Environmental Sciences, School of Biological Sciences, University of Aberdeen, Aberdeen, UK, [8]Department of Agriculture, Horticulture and Engineering Sciences, Scotland's Rural College, Edinburgh, UK, [9]INRA, VetAgro Sup, UCA, UMR Écosystème Prairial (UREP), Clermont-Ferrand, France, [10]AgResearch Grasslands Research Centre, Palmerston North, New Zealand, [11]Potsdam Institute for Climate Impact Research, Potsdam, Germany, [12]AgResearch-Lincoln Research Centre, Lincoln, New Zealand

Key Points:
- Biogeochemical models are useful to assess N_2O fluxes from grasslands but still need validation against in situ measurements
- DayCent performed best for annual N_2O emissions while APSIM best represented daily N_2O emissions, particularly after fertilizer application
- The ensemble-average improved the estimated N_2O emissions compared to the IPCC-Swiss estimate

Supporting Information:
- Supporting Information S1

Correspondence to:
K. Fuchs,
kathrin.fuchs@posteo.de

Citation:
Fuchs, K., Merbold, L., Buchmann, N., Bretscher, D., Brilli, L., Fitton, N., et al. (2020). Multimodel evaluation of nitrous oxide emissions from an intensively managed grassland. *Journal of Geophysical Research: Biogeosciences*, *125*, e2019JG005261. https://doi.org/10.1029/2019JG005261

Received 16 MAY 2019
Accepted 30 NOV 2019
Accepted article online 18 DEC 2019

Author Contributions:
Conceptualization: Kathrin Fuchs, Lutz Merbold, Mark Lieffering, Paul C. D. Newton, Robert M. Rees, Susanne Rolinski, Pete Smith, Val Snow
Data curation: Kathrin Fuchs, Raphaël Martin
Formal analysis: Kathrin Fuchs, Lorenzo Brilli, Nuala Fitton, Katja Klumpp, Raphaël Martin, Val Snow
Funding acquisition: Lutz Merbold, Nina Buchmann
Methodology: Kathrin Fuchs, Lutz Merbold, Susanne Rolinski
(continued)

Abstract Process-based models are useful for assessing the impact of changing management practices and climate on yields and greenhouse gas (GHG) emissions from agricultural systems such as grasslands. They can be used to construct national GHG inventories using a Tier 3 approach. However, accurate simulations of nitrous oxide (N_2O) fluxes remain challenging. Models are limited by our understanding of soil-plant-microbe interactions and the impact of uncertainty in measured input parameters on simulated outputs. To improve model performance, thorough evaluations against in situ measurements are needed. Experimental data of N_2O emissions under two management practices (control with typical fertilization versus increased clover and no fertilization) were acquired in a Swiss field experiment. We conducted a multimodel evaluation with three commonly used biogeochemical models (DayCent in two variants, PaSim, APSIM in two variants) comparing four years of data. DayCent was the most accurate model for simulating N_2O fluxes on annual timescales, while APSIM was most accurate for daily N_2O fluxes. The multimodel ensemble average reduced the error in estimated annual fluxes by 41% compared to an estimate using the Intergovernmental Panel on Climate Change (IPCC)-derived method for the Swiss agricultural GHG inventory (IPCC-Swiss), but individual models were not systematically more accurate than IPCC-Swiss. The model ensemble overestimated the N_2O mitigation effect of the clover-based treatment (measured: 39–45%; ensemble: 52–57%) but was more accurate than IPCC-Swiss (IPCC-Swiss: 72–81%). These results suggest that multimodel ensembles are valuable for estimating the impact of climate and management on N_2O emissions.

Plain Language Summary We tested the performance of three dynamic simulation models against measured nitrous oxide (N_2O) fluxes and its driver variables for a Swiss grassland. We showed that DayCent performed best in the prediction of annual N_2O emissions but was outperformed by APSIM for daily N_2O emissions. We identified particular strengths and weaknesses of each model. Further, we compared the individual models against the N_2O flux estimate made with the Intergovernmental Panel on Climate Change (IPCC)-derived method for the Swiss agricultural greenhouse gas inventory (IPCC-Swiss). Most individual models were worse than IPCC-Swiss but the mean of all model predictions was much better than IPCC-Swiss. All methods overestimated the N_2O mitigation effect of a clover-based N_2O mitigation. IPCC-Swiss was worst and the model ensemble was best at estimating the effects of the mitigation. The findings highlight that using multiple models in an ensemble is beneficial for assessing management and climate impacts on N_2O emissions.

1. Introduction

Nitrous oxide (N_2O) concentrations in the atmosphere impacts the Earth's system in two ways: first by its global warming effect as the third most important greenhouse gas (Intergovernmental Panel on Climate Change, 2013) and second as the most important substance contributing to stratospheric

Project administration: Val Snow
Software: Mark Lieffering
Supervision: Lutz Merbold, Nina Buchmann, Val Snow
Validation: Kathrin Fuchs, Daniel Bretscher, Lorenzo Brilli, Nuala Fitton, Katja Klumpp, Raphaël Martin, Val Snow
Visualization: Kathrin Fuchs
Writing - original draft: Kathrin Fuchs
Writing – review & editing: Kathrin Fuchs, Lutz Merbold, Nina Buchmann, Daniel Bretscher, Lorenzo Brilli, Nuala Fitton, Cairistiona F. E. Topp, Katja Klumpp, Paul C. D. Newton, Robert M. Rees, Susanne Rolinski, Pete Smith, Val Snow

ozone depletion during the 21st century (Ravishankara et al., 2009; United Nations Environment Programme (UNEP), 2013). Consequently, N_2O emissions and other greenhouse gases (GHGs) need to be reduced in order to reach the climate goal of limiting global temperature rise during this century to below 2 °C compared to preindustrial levels (The Paris Agreement; United Nations Framework Convention on Climate Change, 2015) and to avoid undermining the achievements of the Montreal protocol (United Nations Environment Programme, 1987), which was successfully implemented to avoid ozone-depleting substances.

Agriculture is the dominant source of global anthropogenic N_2O emissions, contributing approximately 66% (3.8–6.8 Tg N_2O-N $year^{-1}$, UNEP, 2013). Agricultural N_2O emissions mainly originate from nitrogen (N) in mineral and organic fertilizers, and crop residues and through enhanced organic matter mineralization following soil cultivation (tillage) (Davidson, 2009; Mosier et al., 1998; UNEP, 2013). Grasslands cover more than one quarter of the terrestrial surface and comprise about 70% of agricultural lands (Food and Agriculture Organization, 2013; White et al., 2000). Agroecosystems can act as a potent GHG sink via carbon (C) sequestration, reflected in large C stocks accumulated in grasslands soils (Conant et al., 2005, 2017). However, the C and N fluxes largely depend on grassland management (Soussana et al., 2010), for example, the application of mineral fertilizers and manures to grassland and grassland tillage. As a response to agricultural practices, European grasslands currently are not a significant sink in terms of total GHGs (-14 ± 18 g C m^{-2} $year^{-1}$; Schulze et al., 2009), as CO_2 losses via respiration and N_2O losses from the soil offset photosynthetic CO_2 uptake. Intensively managed mown grasslands have been shown to act as a net source of GHGs to the atmosphere when accounting for on- and off-site emissions (Soussana et al., 2010). However, for assessing GHG mitigation options a thorough understanding of management effects on GHG emissions is necessary.

Emissions of N_2O from agricultural soils occur due to microbial processes, most importantly during nitrification and denitrification, particularly the latter. Nitrification is the oxidation of ammonium (NH_4^+) to nitrate (NO_3^-) via several intermediates under aerobic conditions, with N_2O as a by-product. Denitrification is the reduction of NO_3^- to dinitrogen (N_2) under anaerobic conditions, with N_2O as an intermediate substance, which only under complete anoxic conditions is reduced further to N_2 (Baggs, 2008; Butterbach-Bahl et al., 2013; van Groenigen et al., 2015). Known drivers of these processes are available NH_4^+ and NO_3^-, labile organic C as substrate for heterotrophic microorganisms, soil temperature, soil water content, and soil oxygen concentration, and soil pH (Ball, 2013; Blagodatsky & Smith, 2012; Butterbach-Bahl et al., 2013; Hörtnagl et al., 2018). Up to date we lack the ability to explain the variation in observed N_2O emissions with known environmental drivers, reflecting our yet limited ability to comprehensively measure and understand N cycling processes and their interactions (Butterbach-Bahl et al., 2013; Kuypers et al., 2018). High variability of N_2O fluxes in time and space (Cowan et al., 2015; Groffman et al., 2009) makes the bottom-up estimation of national emissions problematic, resulting in large uncertainties (Reay et al., 2012). Due to the lack of easily accessible and reliable alternatives, national N_2O emission inventories are mostly based on simple emission factors (EFs) as used in the Intergovernmental Panel on Climate Change (IPCC) Tier 1 approach (Intergovernmental Panel on Climate Change, 2008). This approach takes into account N inputs from fertilizers, crop residues, mineralization, atmospheric deposition, and urine and dung deposited by grazing animals but neglects any site-specific effects, for example, climatic conditions and/or soil properties. It may be suitable for estimating total national emissions, integrating fluxes over a large area and over a long time period. However, this approach is not intended for use at specific sites and can lead to large deviations from the measured N_2O fluxes. As a more sophisticated method, models can be used to simulate influencing factors that are then used to determine N_2O emissions (Tier 2) or to directly simulate N_2O emissions on a regional scale and with high temporal resolution (Tier 3). Still, the use of models for emissions inventories remain scarce (Environmental Protection Agency, 2019) and the added value compared with IPCC Tier 1 approaches needs to be shown.

Process-based biogeochemical models provide an opportunity to scale up N_2O flux estimates based on process equations reflecting a simplified synthesis of the currently available process knowledge in C and N cycling and their drivers. However, it remains a challenge for state-of-the-art process-based biogeochemical models to accurately represent interannual and intraannual patterns in N_2O emissions with sufficient

accuracy and certainty (Ehrhardt et al., 2018; Zimmermann et al., 2018) for implementation in policy. Studies validating simulated N_2O emissions were commonly based on static chamber measurements, typically restricted to a few events during the growing season on a plot scale, and concurrent meteorological conditions (Zimmermann et al., 2018). In order to go beyond current assessments and obtain a comprehensive and robust assessment of model performance, new validation exercises need to cover longer time spans entailing a wide variability in meteorological conditions and management activities (such as fertilization, harvest, grazing, and resowing), a much higher temporal resolution compared to typical manual chamber measurements, and be at the ecosystem scale (>1 ha) rather than the plot scale (less than a few square meters).

Dynamics in N_2O emissions, particularly the sporadic nature of peak emissions, require high-resolution flux data in order to capture emission peaks ("hot moments"), induced by management and environmental conditions (Barton et al., 2015; Fuchs et al., 2018; Groffman et al., 2009; Hörtnagl et al., 2018). Eddy covariance (EC) measurements are continuous in time, measure a spatial average at the ecosystem scale, and allow for high confidence that the measurements represent the field-scale fluxes at the soil-atmosphere interface (Finkelstein & Sims, 2001; Rannik et al., 2016). N_2O flux measurements covering four years were acquired with the EC technique at an intensively managed Swiss grassland site (Fuchs et al., 2018) providing a unique opportunity to validate process-based models on daily, weekly, monthly, and annual time scales. Here we applied three process-based models with some variants, resulting in five sets of outputs in total and compared the model outputs against in situ observations. The model selection includes DayCent, which has already been applied for national N_2O reporting (Environmental Protection Agency, 2019); APSIM, which is widely applied for different ecosystems; and PaSim, which has been developed particularly for grasslands.

Combining diverse models, which have different strengths and weaknesses, in a multimodel ensemble promises to increase the accuracy and reliability of the results. Models represent ecosystem fluxes as a set of equations and thereby inevitably introduce inaccuracies that propagate and upscale to inaccuracies in model outputs. The idea of the multimodel ensemble concept is to account for this inherent model uncertainty by applying several skillful and independent models to better cover the whole space of possible outputs. The resulting multimodel ensemble average is subsequently improved because of error cancelation, which results in increased consistency and reliability (Hagedorn et al., 2005). Since a few recent studies of biogeochemical models have shown the ensemble estimates to be more accurate than individual model results (Asseng et al., 2013; Ehrhardt et al., 2018; Wallach et al., 2018), the use of ensemble averages is a promising option for Tier 3 estimates.

Our specific objectives were as follows:

1. To test the performance of the multimodel ensemble to simulate N_2O emissions;
2. To quantify differences between modelled and measured N_2O emissions with respect to cumulative daily/weekly/monthly/annual fluxes and identify periods (i.e., management events, meteorological conditions) of coherences and periods of discrepancies between modelled and measured N_2O emissions; and
3. To assess the performance of each model in representing key variables driving N_2O emissions (i.e., soil temperature, soil water content, NH_4^+, and NO_3^- concentrations) and to reveal the dependence of performance in N_2O emissions on the drivers' performances.

We hypothesized models as follows:

1. To perform better than the IPCC-Swiss estimate since models include additional information (i.e., variability in soil meteorology and management);
2. To perform best following recent fertilizer application events, because then emissions are largely driven by external N inputs; and
3. To simulate N_2O emissions most accurately during periods when driver variables such as soil temperature, soil water content, ammonium, and nitrate concentrations were modelled accurately, because those control N_2O emissions

Our findings will thus assist in diagnosing potential causes of discrepancies and to specify conditions for which improvements in the models and/or data collection are needed.

2. Materials and Methods

2.1. Site conditions and Management

The Chamau field site (Swiss FluxNet site code: CH-CHA) is a temperate grassland located on the Swiss Plateau 30 km southwest of Zurich (47°12′36.8″N, 8°24′37.6″E, 393 m above sea level), characterized by average annual temperature of 9.8 °C and 1,179 mm precipitation (Gilgen & Buchmann, 2009; based on data from the MeteoSwiss station Cham). The soil is a Cambisol/Gleysol with a bulk density in 0–0.2 m depth between 0.9 and 1.3 g cm^{-3} (Roth, 2006), and pH of 6.5 (in 2014; Labor Ins AG, Kerzers, Switzerland). The site has been a permanent grassland since at least 2002, with the latest restoration in 2012 (Merbold et al., 2014), when it was resown with perennial ryegrass (*Lolium perenne*), common meadow grass (*Poa pratensis*), red fescue (*Festuca rubra*), timothy (*Phleum pratense*), white clover (*Trifolium repens*), and red clover (*Trifolium pratense*). Besides these sown species, dandelion (*Taraxacum officinale*), and rough meadow grass (*Poa trivialis*) occur.

In 2015, the site was divided into two adjacent grassland parcels (Fuchs et al., 2018; Parcel A of 2.2 ha and Parcel B of 2.7 ha). The conventional management (Control treatment: Ctr; Parcel A) consisted of four to six harvests (mown, used as silage or hay) and a subsequent application of fertilizer in form of liquid slurry three to seven days after mowing. Typical annual slurry applications at the site were 266 ± 75 kg N ha^{-1} $year^{-1}$ (average ± SD over the 11 years 2003–2014). In the years 2015 and 2016, we tested an N_2O mitigation option (Parcel B; Clover treatment: Clo), that is, oversowing with clover in order to increase biologically fixed N (BFN) whilst omitting fertilization (Fuchs et al., 2018). Oversowing with *Trifolium pratense L.* and two varieties of *Trifolium repens L.* was carried out in spring each year to increase the proportion of clover. The land owner carried out the oversowing by harrowing to 0.01 m depth and sowing on top of the existing vegetation with the purpose of depressing herbs and shifting species composition to a grass-clover mixture with a higher proportion of clover than on the control. Furthermore, the parcels were occasionally grazed, mostly during winter. Detailed management information and soil characteristics are given in Table S1 in the supporting information and in Fuchs et al. (2018).

2.2. Eddy Covariance Flux, Meteorological, and Soil Measurements

We continuously measured greenhouse gas exchange (CO_2, N_2O, CH_4, and H_2O) at a tower located at the boundary between the two parcels using the EC technique (Aubinet et al., 2012; Eugster & Merbold, 2014) during the four years presented in this study (2013–2016). In the EC technique the gas flux is calculated from the covariance of the vertical wind velocity with the respective gas concentration. Due to the tower location at the parcel boundary, the two prevailing wind directions cause the fetch of the EC measurements being most of the time either in one or the other parcel. Details of the eddy covariance measurements and flux postprocessing and the attribution of the flux to the two parcels are described in Fuchs et al. (2018).

Observations of air temperature and relative humidity (2 m height; Hydroclip S3 sensor, Rotronic AG, Switzerland), components of the radiation balance (2 m height; CNR1, Kipp & Zonen B.V., Delft, The Netherlands), and precipitation (1 m height; tipping bucket rain gauge model 10116, Toss GmbH, Potsdam, Germany) were acquired from the tower located between both parcels. Soil microclimatic variables were continuously measured next to the tower, including volumetric soil water content (at 0.04 and 0.15 m depth; ML2x sensors, Delta-T Devices Ltd., Cambridge, UK) and soil temperature (at 0.05, 0.10, and 0.15 m depth; TL107 sensors, Markasub AG, Olten, Switzerland). Soil temperature and soil water content from the measurements near the tower were used for both parcels. While soil temperature at 0.1 m depth was available for the study period, soil water content in 0.1 m depth was not continuously available due to sensor failure, and therefore the average from sensors in 0.04 and 0.15 m depth was used for this analysis.

2.3. Models and Model Variants

We used three process-based models: APSIM (Holzworth et al., 2014, in two variations), DayCent (Parton et al., 1998; Del Grosso et al., 2001, in two variations) and PaSim (Riedo et al., 2000). We calibrated models previous to this study (corresponding to stage 5 in Ehrhardt et al., 2018) using site data from 2010–2012. For model descriptions, see Ehrhardt et al. (2018) and Fitton et al. (2019). In addition, we applied the DayCent and APSIM models to a nearby site to validate their estimation of grassland yield and N fixation (Fitton et al., 2019).

DayCent is based on the Century model (Parton et al., 1998) but uses a daily time step (Del Grosso et al., 2001). We applied the two variants DayCent v4.5 2010 (here DC1) and DayCent v4.5 2013 (here DC2). These variants differ in their calculations of solar radiation and in their calculations of maintenance and growth respiration. The later version also includes the simulation of freeze-thaw events. DayCent includes four main submodels, which are (1) the plant growth submodel for calculating biomass production and allocating net primary production to the plant pools, (2) the soil organic matter submodel for simulating decomposition of dead plant material (litter) and soil organic matter and allocating soil carbon to three soil organic carbon pools and the litter pool, (3) the soil water submodel for the water flow between different layers, and (4) the trace gas flux submodel for gaseous emissions.

PaSim (Calanca et al., 2007) is a pasture model simulating water, C and N cycling in grassland at a subdaily time step, here aggregated to daily outputs. Different modules are responsible for microclimate, soil, vegetation, herbivores, and management. C from photosynthesis and N from soil and fixation are allocated dynamically to one root and three shoot compartments.

APSIM (v7.10 r4162), the Agricultural Production Systems sIMulator (Holzworth et al., 2014), was used in two variants (AP1 and AP2), which differ in their soil water modules. The SWIM water module uses the Richards equation (here AP1; Huth et al., 2012), while the Soil Water (SoilWat) module is capacitance based (here AP2; Probert et al., 1998). N_2O emissions are known to be very sensitive to soil water content so the variation in soil water model, while keeping other aspects constant, was deliberately introduced to understand if one water model was better than the other. The AgPasture module (Li et al., 2011) was used for pasture growth, with allocations of N reserves according to Vogeler and Cichota (2016) and the Penman-Monteith equation designed for intermingled canopies (Snow & Huth, 2004). The SoilN module was applied for soil organic matter and nitrogen transformations (Probert et al., 1998).

2.4. Model Input Data and Model Setup

Modeling groups received detailed or specific management data (amount of N, type of management; see Table S1) for the 4-year observation period (2013 to 2016) as well as the site history (since 2002), including the information on the tillage and sowing operations for the regrassing in 2012. Climate data from the field site were used as model input for 2010–2016, while historical data before 2010 were used from AgMERRA (Ruane et al., 2015) in case of spin-up of models. Input data in daily time steps were mean, minimum, and maximum air temperature, total precipitation, average wind speed, average global radiation, average relative humidity, and average dewpoint temperature. These variables were mainly directly measured at the station (74–95% of the days, depending on the variable, see Table S2 for details). Data from the proximate meteorological station Cham were acquired from the Swiss meteorological service MeteoSwiss (https://gate.meteoswiss.ch/idaweb) and used for gap filling if available. For the rare cases when neither the original value from the Chamau nor a value from Cham were available, the mean value over all seven years (2010–2016) at the day of the year (DOY) was used, which was only the case on 12 days for shortwave radiation and on 36 days for relative humidity (Table S2). We used the simplest form of a multimodel ensemble and merged the individual model outputs each day with equal weights.

2.5. Statistical Analysis

We used the term "background fluxes" for N_2O fluxes beneath the threshold of 1.2 mg N_2O-N m^{-2} day^{-1}. This threshold corresponded to the mean of N_2O fluxes during the months October and March at Chamau, reflecting N_2O fluxes at the start and end of the growing season (Fuchs et al., 2018) when no management event took place. The threshold corresponded well with the background fluxes reported for another Swiss grassland site (Neftel et al., 2007).

We defined an N_2O emission "peak" as N_2O emissions exceeding background emissions. We distinguished "peaks after management", which were defined as peaks finishing ≤14 days after a management event (e.g., fertilizer application, harvest, and grazing) and all other peaks as "peaks not directly linked to management".

As a measure of the error in estimated values, the root-mean-square error was used (Bennett et al., 2013).

$$\mathrm{RMSE} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(S_i - O_i)^2} \quad (1)$$

S_i denotes the simulated value and O_i the observed value at index i, and n is the number of observed values (Bennett et al., 2013). With the same notation, the bias is defined as

$$\mathrm{Bias} = \frac{1}{n}\sum_{i=1}^{n}(S_i - O_i) \quad (2)$$

A positive bias indicates an overestimation; a negative bias indicates an underestimation by the model simulations. Relative RMSE and relative bias were calculated by further dividing by the mean of all observations $\overline{O}$. When analyzing daily values, the analysis reflects only the days of directly comparable observed and simulated fluxes, while few days were omitted when no measurements were available.

RMSE95 and Bias95 indicate the 95% confidence intervals for RMSE and Bias. RMSEs larger than the RMSE95 and Biases outside the Bias95 confidence interval indicate significant differences between simulated and observed values (Smith & Smith, 2007).

$$\mathrm{RMSE95} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{SE}_i \times t_m)^2} \quad (3)$$

The RMSE95 uses the standard error of the ith measurements (SE_i) and the value of the t statistics for m replicates.

In order to assess potential time offsets, that is, a delayed response to observed peak N_2O emissions by models, we tested if the RMSE was lower for lagged model outputs compared to the outputs at lag 0. The RMSE at lag l is calculated as the RMSE of observations with a delayed time series by lag l:

$$\mathrm{RMSE}_l = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(S_{i+l} - O_i)^2} \quad (4)$$

If the modeled variable lags behind the observed variable, the RMSE will be lower at a lag >0 (lag 0 corresponds to the unshifted time series). Negative lags were unimportant here because no model showed too early model responses in N_2O emissions. Time lags were investigated by shifting the time series 1–10 days.

We used the Nash-Sutcliffe model efficiency NSE (Nash & Sutcliffe, 1970) to assess model performance in comparison with the measured site mean (McCuen et al., 2006; Nash & Sutcliffe, 1970):

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(S_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \quad (5)$$

A NSE of 1 would result for identical simulated and observed values; a positive NSE implies that the simulated values are better estimates than the mean of all observations across the simulation period.

We investigated all fertilization events during the 4-year observation period comparing the development of cumulative N_2O fluxes of observations and simulations for 14 days after a fertilizer amendment. The 2-week period has previously been identified as a general time span within which N_2O emissions return to prefertilization values (e.g., Bowatte et al., 2018). At the Chamau site N_2O fluxes typically decayed within less than a week (Fuchs et al., 2018; Hörtnagl et al., 2018); thus, the chosen time interval was a conservative choice to include potentially lagged simulated N_2O emission peaks.

To analyze potential effects of model performance in simulating driver variables on the performance of simulated N_2O fluxes, the average deviations of the simulations from the observed values (bias in soil water content (ΔSWC) and soil temperature (ΔTS)) were calculated per model for each 14-day postfertilization period. Similarly, the deviation of observed and simulated cumulative N_2O emissions (ΔN_2O) for 14 days after fertilizer application was calculated per model and event. The deviations in the different driver variables are potentially affecting deviations in cumulative N_2O emissions. Thus, in order to assess the relationship between deviations in driver variables and deviations in N_2O emissions, we performed a multiple linear

regression analysis to detect the effect of overestimated SWC or TS (and their interaction) in a systematic way.

For giving reference to the actual soil climatic conditions, the driver variables soil temperature and soil water contents were classified according to the following scheme: For soil temperatures, average values <10 °C were classified as cold, 10–15 °C as fresh, 15–20 °C mild, and > 20 °C as warm. For soil water contents, conditions <40% SWC were classified as dry, 40–45% as moderately moist, 45–50% as moist, and >50% as extremely moist. We used daily values for our analysis even though a 2–3 days moving average might be conceptually more appropriate, since a moving average resulted in little changes in the validation result.

2.6. Estimates Using the IPCC-Swiss Method

N_2O emissions were calculated according to the current IPCC Guidelines for National Greenhouse Gas Inventories (Intergovernmental Panel on Climate Change, 2008) adapted to the specific case of Switzerland as described in FOEN (2018):

$$F_{N_2O-Direct} = (F_{SN} + F_{ON} + F_{CR} + F_{SOM})*EF_1 + F_{PRP,CPP}*EF_{3PRP,CPP} + F_{PRP,SO}*EF_{3PRP,SO} \quad (6)$$

$$F_{N_2O-Indirect} = F_{AD}*EF_4 \quad (7)$$

where F_{SN} and F_{ON} are the amendments of total synthetic and organic N to the soil, F_{CR} the N from crop residues, F_{SOM} the N from mineralization, and F_{AD} atmospheric deposition. $F_{PRP,SO}$ and $F_{PRP,CPP}$ are the amounts of N deposited during grazing by sheep and cattle, respectively. EF_1 and $EF_{3PRP,CPP}$ correspond to the Tier 1 IPCC emission factors for direct soil emissions and urine and dung deposited by grazing sheep and cattle, respectively, that is, 0.01 and 0.02. F_{AD} reflects atmospheric N deposition. The parameters F_{SN} and F_{ON} were the applied fertilizer amounts. F_{CR}, F_{SOM},F_{AD}, $F_{PRP,SO}$, and $F_{PRP,CPP}$ were calculated per parcel and year using the calculation schemes of the Swiss GHG inventory for intensive meadows (FOEN, 2018). F_{CR} was the standard inventory yield of intensive meadows multiplied by 2.4% N content and the fraction of residuals left on the field, assumed as 15% of the yields (FOEN, 2018). F_{SOM} was calculated as the inventories C mineralization rate, divided by the C/N ratio of 9.8 (Leifeld et al., 2007). For atmospheric deposition (F_{AD}) a value of 33.8 kg N ha^{-1} $year^{-1}$ was used from modelled estimates of local deposition (Rihm & Achermann, 2016). $F_{PRP,SO}$ and $F_{PRP,CPP}$ were calculated from livestock numbers, the duration of grazing, and the livestock specific nitrogen excretion rates (N_{ex}), which were adopted from the Swiss GHG inventory, that is, 111 kg N $head^{-1}$ $year^{-1}$ for mature dairy cattle and 8.4 kg N $head^{-1}$ $year^{-1}$ for sheep (FOEN, 2018). Total N_2O emissions were calculated by adding direct and indirect emission estimates at the site.

2.7. Uncertainties in N_2O Flux Observations

The source of uncertainties was minimized by using a dataset of high temporal resolution. Nevertheless, EC measurements are subject to uncertainties, for example, due to the random sampling error, footprint variability, instrument noise, or lateral fluxes. Despite uncertainties in measured annual N_2O fluxes of $\pm$0.043 to $\pm$0.200 g N_2O-N m^{-2} $year^{-1}$, most simulated annual N_2O fluxes showed significant biases, that is, biases larger than the measurement uncertainties (Table 1).

3. Results

3.1. Simulated and Observed N_2O Fluxes

3.1.1. Interannual Variability and Model Performance

Measured annual N_2O emissions for the fertilized treatment (Ctr) were significantly higher (0.511 g N_2O-N m^{-2} $year^{-1}$) compared to the unfertilized treatment (Clo) across all years (0.301 g N_2O-N m^{-2} $year^{-1}$; $p < 0.05$) (Table 1). The same pattern was depicted well in all models. However, annual N_2O emissions were underestimated (ensemble mean; E-Mean), particularly during the year with cattle summer grazing (2014 in Ctr) and during the moist year 2016 in both parcels.

Annual simulations of N_2O emissions of the model ensemble outperformed the IPCC-Swiss estimate, shown by a lower RMSE compared to the $RMSE_{IPCC}$ (Figure 1 and Table S3). While some models (PaSim and DC2) overestimated N_2O emissions at annual timescales, others (AP1, AP2, and DC1) underestimated annual N_2O emissions. DayCent (both versions) performed particularly well in simulating annual N_2O emissions,

Table 1
Annual Aggregates of Measured and Simulated N_2O Fluxes From DayCent Variants DC1, DC2, PaSim, APSIM variants AP1, AP2, and the IPCC Tier 1 Estimate, the Ensemble Median (E-median), and the Ensemble Mean (E-Mean; Both in g N_2O-N m^{-2} $year^{-1}$)

Parcel	Year	Treatment	Measured (±SE)	DC1	DC2	PaSim	AP1	AP2	IPCC	E-Mean	E-Median
A	2013	Ctr	0.535 (±0.097)	0.287	0.610	0.743	0.292	0.365	0.276	0.455	0.362
A[a]	2014	Ctr	0.559 (±0.121)	0.389	0.634	0.431	0.183	0.149	0.377	0.374	0.274
A	2015	Ctr	0.393 (±0.089)	0.354	0.425	0.972	0.418	0.429	0.361	0.485	0.406
A	2016	Ctr	0.594 (±0.200)	0.262	0.528	0.658	0.238	0.283	0.253	0.402	0.319
B	2013	Ctr	0.492 (±0.104)	0.349	0.601	0.982	0.352	0.421	0.303	0.516	0.402
B	2014	Ctr	0.493 (±0.119)	0.391	0.596	0.715	0.188	0.178	0.353	0.412	0.321
B[b]	2015	Clo	0.217 (±0.043)	0.157	0.319	0.311	0.153	0.148	0.070	0.210	0.171
B	2016	Clo	0.385 (±0.095)	0.087	0.402	0.261	0.066	0.069	0.072	0.192	0.131

Note. Parcel A was fertilized in all years 2013–16, while Parcel B was fertilized only in 2013–2014 (referred to as fertilized control treatment-years Ctr) and was subject to the unfertilized clover treatment (Clo) during 2015–2016. SE = standard error.
[a]Parcel A was grazed with cattle for 36 days during the season, replacing two cuts. [b]Parcel B was grazed with sheep for eleven days during the growing season, replacing one cut. Other than that, only winter grazing took place.

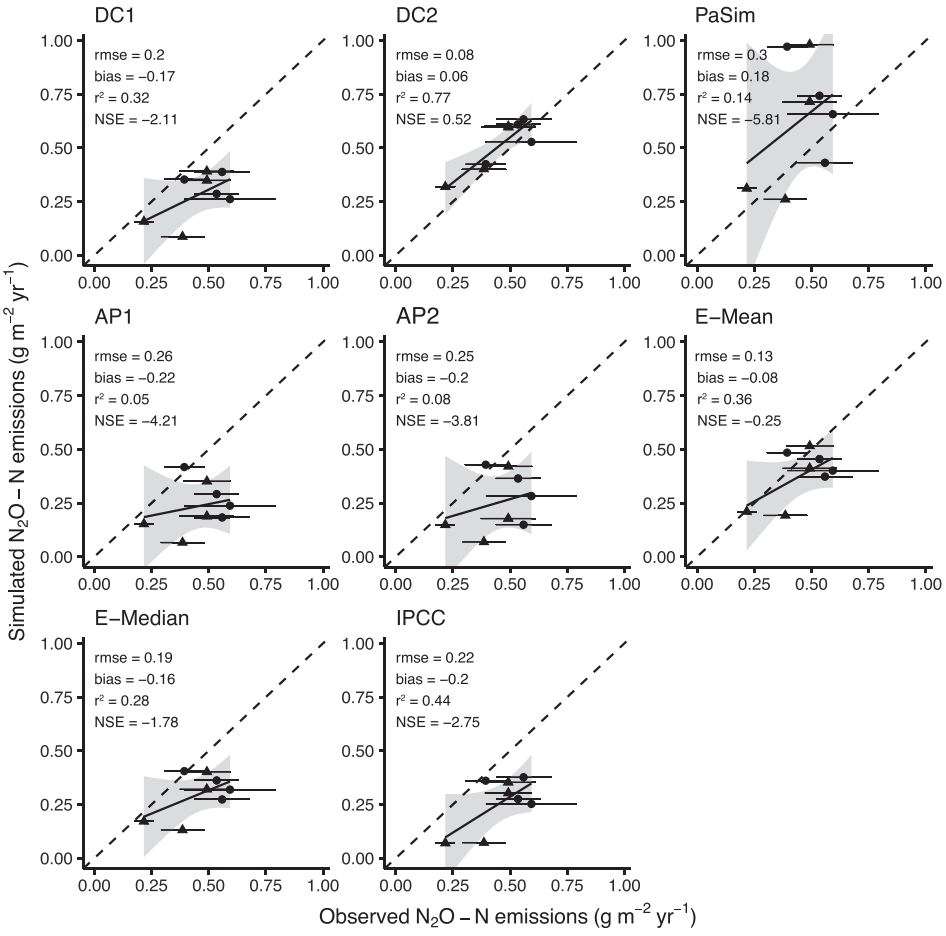

Figure 1. Annual values of observed (horizontal axes) versus simulated N_2O emissions (vertical axes) for models DC1, DC2, PaSim, AP1, AP2, the ensemble mean (E-Mean), the ensemble median (E-Median), and for the IPCC-Swiss estimate in Parcel A (circles) and Parcel B (triangles), with the horizontal whiskers representing the measurement uncertainties (±1 SE). The dashed lines indicate the 1:1 lines, and the solid lines display the linear regression line between observed and simulated N_2O fluxes and the grey shading depicts the regressions' 95% confidence interval. Bias and RMSE are given in g N m^{-2} $year^{-1}$.

indicated by a low RMSE (for DC1 0.20 and for DC2 0.08 g N_2O-N m^{-2} $year^{-1}$). The DayCent variant DC1 underestimated N_2O emissions by 38%, while DC2 slightly overestimated annual N_2O emissions (+12%). DayCent showed a regression slope of observed versus simulated values close to 1 ($slope_{DC1}$ = 0.9; $slope_{DC2}$ = 0.7), representing the interannual variability in N_2O emissions (Figure 1). Interannual variability of N_2O emissions was simulated comparably well by PaSim ($slope_{PaSim}$ = 1.1). However, PaSim overestimated annual N_2O emissions by 38% with highest RMSE ($RMSE_{PaSim}$ = 0.33 g N_2O-N m^{-2} $year^{-1}$; Figure 1). In contrast, APSIM showed a regression slope <1 and generally underestimated annual N_2O emissions ($bias_{AP1}$ = −48% and $bias_{AP2}$ = −44%), which represents a higher bias than for the IPCC-Swiss estimate of 44% (Figure 1 and Table S3). In summary, the multimodel ensemble average (E-mean) largely improved the accuracy compared to the IPCC-Swiss estimate. The error of the ensemble mean was 41% lower than the error in the IPCC-Swiss estimate ($RMSE_{E\text{-}Mean}$ = 0.13 versus $RMSE_{IPCC}$ = 0.22 g N_2O-N m^{-2} $year^{-1}$). Only one model (DC2) was more accurate than the ensemble mean.

3.1.2. Model Performance (Intraannual Variability: Seasonal, Monthly, Weekly, and Daily N_2O Fluxes)

Fluxes were characterized by low winter (December–February, DJF) N_2O emissions (0.6 mg N_2O-N m^{-2} day^{-1}) and significantly higher N_2O emissions during the growing season (March–May, MAM: 1.2; June–August, JJA: 2.6; September–November, SON: 1.2; all in mg N_2O-N m^{-2} day^{-1} in Ctr; Table S4). This pattern coincided not only with higher temperatures during these months but also with N inputs via fertilization, which were linked to the respective season (Figure 2). In the nonfertilized Clo treatment N_2O emissions were generally lower during all seasons (DJF: 0.4; MAM: 0.4; SON: 0.7; all in mg N_2O-N m^{-2} day^{-1}; Table S4), reaching up to 1.6 mg N_2O-N m^{-2} day^{-1} on average during summer months (JJA).

Winter N_2O fluxes (DJF) were consistently underestimated by all models (Figure 3 and Table 2), with most models having a stronger bias in DJF compared to all other seasons (Table 2), while RMSE was low in winter compared to other seasons due to less variability in N_2O emissions. Still, early 2013 and winter 2013/2014 N_2O emission peaks occurred, but no model represented these observed peaks, which appeared without previous management activities. Springtime (MAM) N_2O fluxes were underestimated by APSIM and DayCent but overestimated by PaSim (Figure 3). The response of N_2O fluxes to fertilizer events in springtime was often not simulated well. For instance, only one model (AP2) simulated the peak after the first fertilizer amendment in 2013 (Figure 3). Summer (JJA) N_2O fluxes showed the lowest absolute bias, indicating that their overall magnitude was well represented (Table 2; except PaSim). However, summer N_2O estimates showed the highest RMSE compared to other seasons (Table 2; except PaSim), reflecting the challenge of simulating the dynamics of daily N_2O fluxes. During autumn (SON), PaSim and DC2 overestimated N_2O emissions, while all other models underestimated them (Figure 3), similarly to springtime N_2O simulations. The deviations in both directions compensated each other, resulting in an improved ensemble mean compared to individual model estimates in summer and autumn (Table 2).

Relative bias and relative RMSE were higher in the Clo treatment compared to the Ctr, reflecting that per unit of seasonal N_2O emission, the N_2O fluxes in the Clo treatment (low N input, biologically fixed N) were more difficult to predict. The absolute RMSE was lower for all models in the Clo treatment compared to the Ctr, reflecting that the Ctr typically showed larger variability in N_2O fluxes and therefore was more challenging for the models (Table 2).

A good performance in annual cumulative N_2O emissions in several models (DC1 and DC2) did not always coincide with a good representation on the weekly and daily timescale (Figures 2 and 3). Even if the annual N_2O emissions were well represented, the lack of coincidence in time led to higher errors (RMSE and RRMSE) at shorter time scales, for example, daily, weekly, and monthly estimates compared to annual estimates (Table S3). For instance, DayCent performed well for annual cumulative N_2O fluxes but did not pick up the measured peak N_2O emissions in their magnitude at the time of occurrence (Figure 2) and instead estimated rather steady N_2O emissions.

PaSim produced a large RMSE across timescales, largely caused by the strong positive bias, while the slope of observed versus simulated values was close to 1 across time scales (Table S3). The two APSIM variants performed best across models for simulating the variability in weekly and daily N_2O fluxes (minimum RMSE, Table S3) and comparably to DayCent for monthly aggregates.

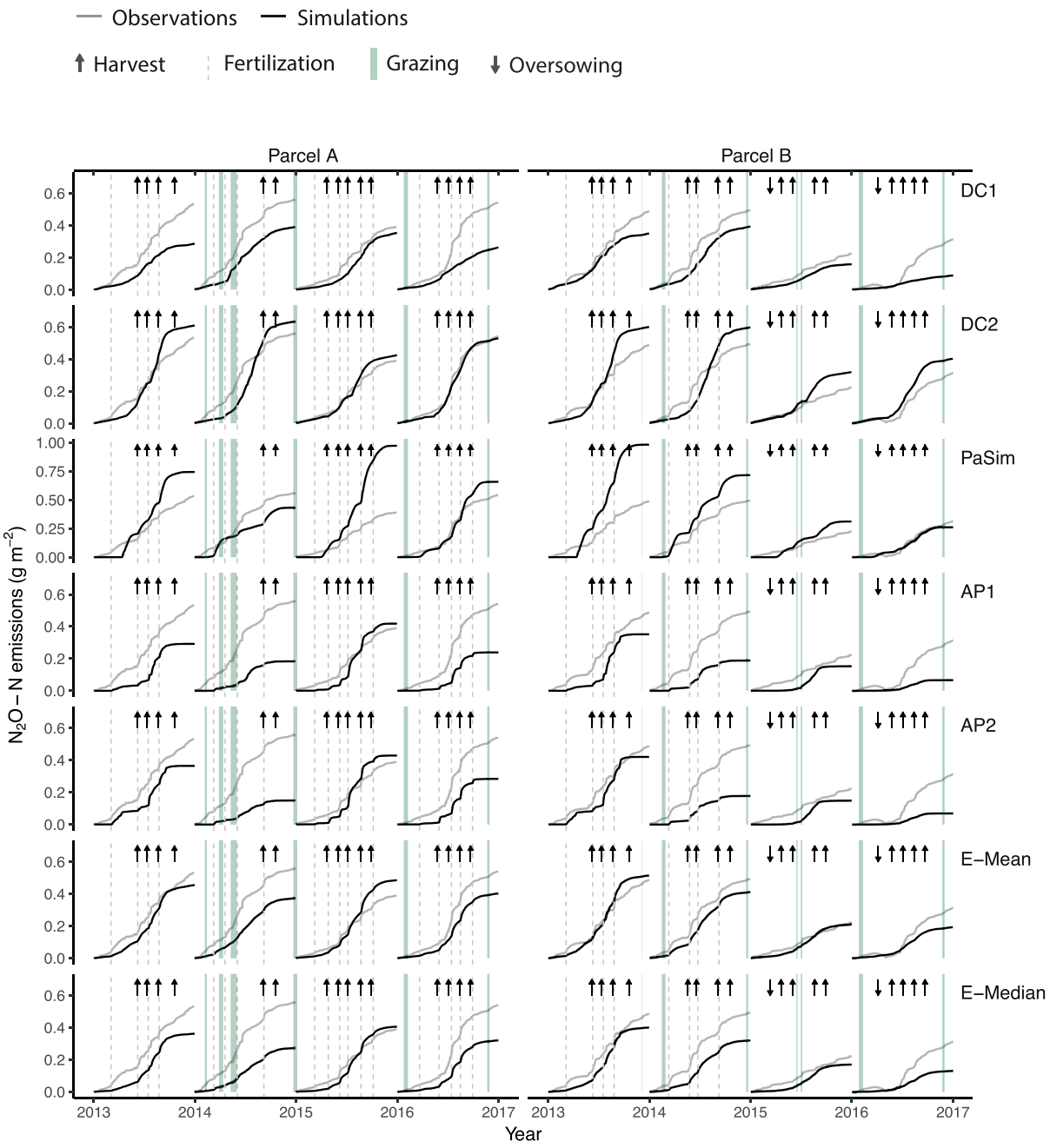

Figure 2. Cumulative daily time series of simulated (black) and observed (grey) N_2O fluxes for Parcel A (left) and Parcel B (right). Dashed lines indicate fertilization events; usually in form of liquid slurry, except for Parcel A in 2014 after both grazing events when mineral fertilizer in form of calcium-ammonium-nitrate was applied (see also Table S1 for detailed management information). Upward arrows indicate the moment of harvest and downward arrows indicate oversowing. Grazing periods are depicted by the green solid background bars. Note that panels differ in their y scale, but the observed (grey) measured fluxes are displayed in all panels as the reference.

To investigate the potential effect of delayed peak N_2O emissions on the model validation, we investigated lagged RMSEs. However, we found no systematic time offsets for shifted model response by 1–10 days. The APSIM variants were, in most cases, estimating the peaks at the correct time (i.e., showing the lowest RMSE for the unshifted time series), while the other models showed minimum RMSEs at randomly varying lags across events.

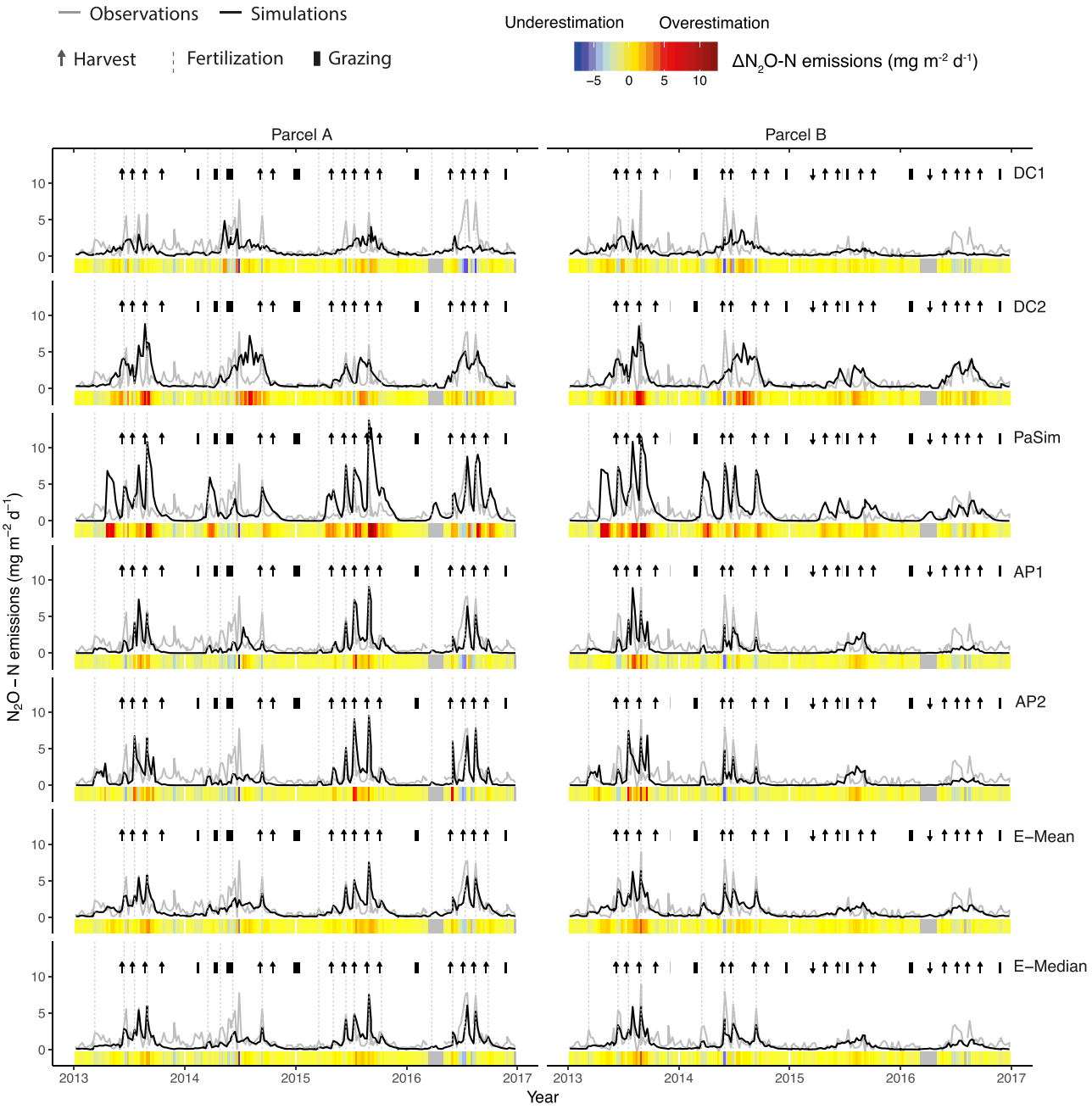

Figure 3. Weekly averages of simulated (black) and observed (grey) N_2O fluxes for Parcel A (left) and Parcel B (right) by models used in this study and the multi-model ensemble (top to bottom). Upward arrows indicate harvest and downward arrows indicate oversowing. Grazing periods are shown as black bars. The weekly bias in N_2O fluxes (ΔN_2O) is displayed as a colored bar at the bottom of each figure, with red indicating an overestimation, blue an underestimation by the respective model, and yellow a bias close to zero (see legend). A grey colored bar indicates periods of missing data.

No individual model estimated more accurate daily values than the average of all observations and achieved a positive NSE on the daily basis. In contrast, the ensemble average achieved an NSE of 0.42, showing that using the ensemble mean largely improved the performance on the daily timescale.

3.1.3. Model Performance in Estimating N_2O Emissions Following Management Events

Our goal was to reveal the strength of each model in estimating N_2O emissions by highlighting the conditions under which they simulated most accurately. Second, we point out the weaknesses distinguishing two types of discrepancies; we call the first one "blind spots" and refer to periods of underestimation,

Table 2
Mean Weekly Bias and Mean Weekly RMSE Per Model Family Derived From Analyses of Variances (Four-Way Analysis of Variance Per Column) Including the Main Effects of Treatment, Year, Season, and Model Variant

		DayCent	APSIM	PaSim	E-Mean	DayCent	APSIM	PaSim	E-Mean
Category	Level	Bias (mg m^{-2} day^{-1})	Bias (mg m^{-2} day^{-1})	Bias (mg m^{-2} day^{-1})	Bias (mg m^{-2} day^{-1})	RMSE (mg m^{-2} day^{-1})	RMSE (mg m^{-2} day^{-1})	RMSE (mg m^{-2} day^{-1})	RMSE (mg m^{-2} day^{-1})
Treatment	Grand mean	−0.15	−0.55	0.53	−0.19	1.28	1.31	1.93	0.99
	Clo	-0.06^{a}	-0.43^{a}	0.00^{a}	-0.19^{a}	0.77^{a}	0.85^{a}	0.91^{a}	0.61^{a}
	Ctr	-0.18^{a}	-0.59^{a}	0.71^{b}	-0.18^{a}	1.45^{b}	1.47^{b}	2.28^{b}	1.11^{b}
Year	2013	-0.15^{a}	-0.41^{b}	0.78^{ab}	-0.10^{ab}	1.47^{b}	1.44^{b}	2.55^{b}	1.19^{a}
	2014	-0.18^{a}	-1.02^{a}	-0.10^{a}	-0.46^{a}	1.49^{b}	1.45^{b}	1.20^{a}	0.98^{a}
	2015	0.06^{a}	0.03^{c}	1.24^{b}	0.20^{b}	0.97^{a}	1.10^{a}	2.33^{ab}	0.87^{a}
	2016	-0.32^{a}	-0.80^{ab}	0.20^{ab}	-0.39^{a}	1.18^{ab}	1.26^{ab}	1.66^{ab}	0.91^{a}
Season	DJF	-0.31^{a}	-0.64^{a}	-0.61^{a}	-0.43^{a}	0.59^{a}	0.81^{a}	0.79^{a}	0.61^{a}
	MAM	-0.26^{a}	-0.67^{a}	0.42^{ab}	-0.28^{ab}	0.99^{ab}	0.97^{ab}	1.80^{ab}	0.74^{a}
	JJA	0.13^{a}	-0.34^{a}	1.07^{b}	0.05^{b}	2.29^{c}	2.26^{c}	2.60^{b}	1.59^{b}
	SON	-0.15^{a}	-0.56^{a}	1.24^{b}	-0.08^{ab}	1.23^{b}	1.20^{b}	2.50^{b}	1.00^{a}
Model	DC1	-0.48^{a}	—	—	—	1.22^{a}	—	—	—
	DC2	0.19^{b}	—	—	—	1.33^{a}	—	—	—
	AP1	—	-0.58^{a}	—	—	—	1.31^{a}	—	—
	AP2	—	-0.53^{a}		—	—	1.32^{a}	—	—
	PaSim	—	—	0.53	—	—	—	1.93	—
	E-Mean	—	—	—	−0.19	—	—	—	0.99

Note. The superscript letters indicate significant differences among levels (TukeyHSD) per factor.

when N_2O emission occur in observations but are absent in model simulations. The other type of error, "phantoms," are periods of overestimations when observed N_2O emissions remain at background levels.

Observed daily N_2O fluxes were characterized by a background N_2O flux (<1.2 mg N_2O-N m^{-2} day^{-1}; see section 2 for the definition) during 64% of the measurement days in the Ctr and 82% of the days in the Clo parcel. DayCent represented most of these background flux days and simulated "phantom" peaks on 18–27% of background days (Table 3). PaSim simulated two thirds of the observed background N_2O flux days adequately, corresponding to phantom peaks on 34% of background days. APSIM showed phantom peaks on 10% of the observed background N_2O flux days. Peak days were represented best by DC2 and PaSim, both correctly predicting almost 60% of the peak days, while APSIM and DC1 represented only one third of the peak days in the control treatment.

The Clo treatment showed fewer peak days (18%) compared to the Ctr (36%), due to the higher management intensity in Ctr. The peak days in Clo in PaSim and DC2 were similarly well represented as in the Ctr, but DC1 and both APSIM variants represented less peak days in Clo (Table 3).

Fertilizer application increased the percentage of correctly simulated peak days (Table 3). However, for harvest or grazing no such pattern occurred. Thus, the peaks associated with fertilizer applications were easier to predict than peaks following grazing or harvest events.

Fertilizer application was mostly followed by high N_2O fluxes in our observations, indicated by an increase in the median N_2O flux across all fertilizer applications (Figure 4), whereas the response of the models to fertilization differed widely, shown by the large differences between the 5% and 95% N_2O flux percentiles (Figure 4). DayCent in particular did not simulate distinct peaks following fertilizer application. PaSim depicted well the onset of the emission peak and simulated peak N_2O fluxes directly after fertilizer application in the correct order of magnitude but with a slightly lower median value than the observations (2.77 mg N_2O-N m^{-2} day^{-1}). However, peaks in PaSim were usually prolonged for several weeks instead of decaying after a few days as in the measurements (Figure 4). This effect led to a large overestimation of annual N_2O fluxes (Figure 1 and Table S3). APSIM represented peak N_2O emissions in many cases at the correct date and with the right decay pattern but underestimated them, indicated by the APSIM median remaining slightly lower that the observed median (Figure 4). Still, due to the overall underestimation of peak N_2O emissions, the cumulative fluxes after events were generally underestimated (Figure S1). N_2O emission pulses that were

Table 3
Number of Correctly Simulated Background Days ("Correct Background") (Threshold <1.2 mg N_2O-N m^{-2} day^{-1}, See Section 2.6) Relative to the Observed Total Number of Background Days; Relative Number of Phantoms, Which Are Peak Days Simulated on Observed Background Days; Number of Undetected Blind Spots, Which Are Not Correctly Detected Peak Days Relative to the Total Observed Peak Days, and Number of Correctly Simulated Peak Days ("Correct Peak") Relative to the Total Number of Observed Peak Days (A) for the Whole Observation Period and for Specific Events Such as (B) Fertilization, (C) Harvest, and (D) Grazing for the Ctr (Left) and the Clo Treatments (Right)

(A) Complete observation period	Control treatment Ctr					Clover treatment Clo				
	DC1	DC2	PaSim	AP1	AP2	DC1	DC2	PaSim	AP1	AP2
Correct background (%)	85	73	66	**90**	89	99	76	78	**93**	92
Phantoms (%)	15	27	34	**10**	11	1	24	22	**7**	8
Blind spots (%)	65	**42**	**42**	70	65	98	**38**	49	89	89
Correct peak (%)	35	**58**	**58**	30	35	2	**62**	51	11	11
(B) Week after fertilization	**Control treatment Ctr**					**Clover treatment Clo**				
	DC1	**DC2**	**PaSim**	**AP1**	**AP2**	**DC1**	**DC2**	**PaSim**	**AP1**	**AP2**
Correct background (%)	**76**	47	27	67	41	—	—	—	—	—
Phantoms (%)	**24**	53	73	33	59	—	—	—	—	—
Blind spots (%)	54	29	**20**	37	26	—	—	—	—	—
Correct peak (%)	46	71	**80**	63	74	—	—	—	—	—
(C) Three days after harvest	**Control treatment Ctr**					**Clover treatment Clo**				
	DC1	**DC2**	**PaSim**	**AP1**	**AP2**	**DC1**	**DC2**	**PaSim**	**AP1**	**AP2**
Correct background (%)	**82**	64	27	**82**	**82**	**94**	65	47	82	82
Phantoms (%)	**18**	36	73	**18**	**18**	**6**	35	53	18	18
Blind spots (%)	68	**24**	41	81	84	100	**0**	73	93	93
Correct peak (%)	32	**76**	59	19	16	0	**100**	27	7	7
(D) Grazing period up to 2 weeks after grazing	**Control treatment Ctr**					**Clover treatment Clo**				
	DC1	**DC2**	**PaSim**	**AP1**	**AP2**	**DC1**	**DC2**	**PaSim**	**AP1**	**AP2**
Correct background (%)	92	92	91	**100**	**100**	88	**92**	84	91	90
Phantoms (%)	8	8	9	**0**	**0**	12	**8**	16	9	10
Blind spots (%)	**67**	69	83	100	97	70	80	**54**	70	65
Correct peak (%)	**33**	31	17	0	3	30	20	**46**	30	35

Note. During the whole observation period we observed 130 peak days and 601 background days in Clo, and 518 peak days and 943 background days in the Ctr. Bold numbers indicate those cases best coinciding with our observations.

observed later, that is, after fertilizer application (e.g., day 10 and day 19), were associated with rewetting of the soil. AP1 represented these rewetting events best (e.g., event of 19 July 2013, Figure S1).

3.2. Model Performance of Driver Variables

Soil temperatures were simulated with high accuracy but were slightly underestimated by all models, ranging from a bias of −1.6 °C (in AP1) to −3.2 ° C (in PaSim; Table 4 and Figure S2). Periods of underestimated soil temperatures (0.1 m depth) were particularly prevalent in winter for DC2 and PaSim, where models predicted frozen soils at 0.1 m depth, while observed soil temperatures were consistently above 0. Soil water content was underestimated by both DayCent variants and less so by PaSim (Table 4 and Figure S3). In contrast, both APSIM variants simulated soil water content well, with only a slight positive bias (Table 4 and Figure S3).

Soil NH_4^+ concentrations were clearly overestimated in DayCent, unbiased in PaSim, and underestimated in APSIM (Table 4 and Figure S4). Accuracies in soil NH_4^+ concentration were lowest in DC2 and PaSim and highest in DC1 and both APSIM variants. Soil NO_3^- concentrations showed larger RMSE compared to NH_4^+ concentrations (Table 4 and Figure S5). While DC1 underestimated NO_3^- concentrations, DC2 overestimated these. All other models showed no bias in NO_3^- concentrations (Table 4). From visual inspection, PaSim and APSIM followed the observed patterns in mineral N best (Figure S5). The amount of N exported via biomass harvest is a major component of the soil N balance and a negative correlation with N_2O emissions could be expected. N exported via harvest was unbiased by DC2, but overestimated by all other models (Table 4 and Figure S6).

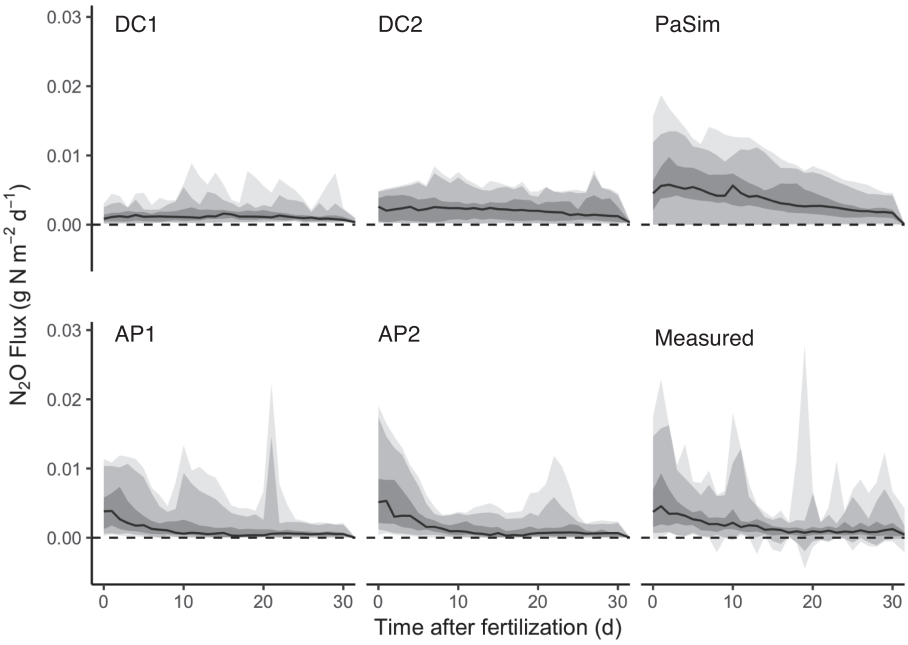

Figure 4. N_2O fluxes after fertilizer applications at day 0. The black line depicts the median across all fertilizer events, the lightest grey shadow depicts the range, the medium grey shadow includes 90% of the observations (5^{th} to 95^{th} percentile), and the dark grey shadow includes 50% (25^{th} to 75^{th} percentile).

3.3. Synthesis: Effect of Model Performance in Simulating Driver Variables on the Performance of Simulating N_2O Fluxes

Knowing whether timespans of high accuracy in N_2O flux estimates coincided with timespans of high accuracy in simulated driver variables of N_2O fluxes reveals if inaccurate N_2O flux estimates might be attributed to weak performance in (at least one of) the previously identified driver variables. We first focused on timespans that are well known for potential high emissions, and thus related the bias in cumulative N_2O fluxes (ΔN_2O) over the 14 days postfertilization period to the bias in soil temperatures (ΔTS) and soil water contents (ΔSWC) (Figure 5).

When analyzed across models, significant increases in ΔN_2O at 14 days after fertilization were found associated with increased ΔTS ($p < 0.001$), larger ΔSWC ($p < 0.05$), and higher biases in NO_3^- concentrations (ΔNO_3^-) ($p < 0.001$). In contrast, no significant relationship between bias in NH_4^+ concentrations (ΔNH_4^+) and ΔN_2O was found.

However, these findings did not hold when the analysis was performed separately for each model. For instance, a more accurate representation of the NO_3^- concentrations at one fertilizer application compared to another event did not significantly improve N_2O estimates. Still, models with accurate soil NO_3^-

Table 4
Model Performance of Simulated Drivers of N_2O Emissions (Soil Temperature TS, soil Water Content SWC, Ammonium NH_4^+, and Nitrate Concentrations NO_3^-) Based on Daily Values During the Full Observation Period

	TS (°C)		SWC (%)		NH_4^+ (kg N ha^{-1})		NO_3^- (kg N ha^{-1})		Harvest N (kg N ha^{-1})	
Model	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
95% CI	±0.8	0.8	±2.5	2.4	±3.5	4.6	±5.7	8.6	±8.1	8.0
DC1	−1.7	2.5	−9.0	13.1	1.3	7.6	−7.7	18.2	45.7	52.3
DC2	−1.8	3.9	−12.3	15.9	10.8	17.5	9.6	23.2	−4.8	26.3
PaSim	−3.2	4.8	−4.3	6.3	2.9	12.1	3.2	17.2	12.6	22.7
AP1	−1.6	2.0	3.5	6.0	−3.6	8.5	−2.0	20.9	21.7	39.3
AP2	−1.5	2.0	2.4	6.5	−4.8	8.3	−0.1	21.1	23.3	36.8

Note. The 95% confidence intervals (95% CI) of bias and RMSE indicate the thresholds beyond which the model's bias or RMSE, respectively, is significant.

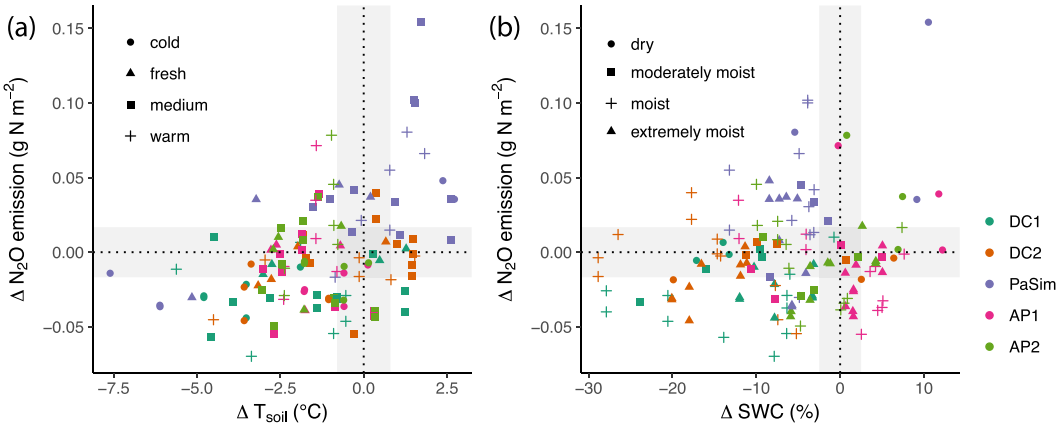

Figure 5. Biases in simulated N_2O fluxes (ΔN_2O) accumulated for 14 days after fertilizer applications per model and event in relation to (a) biases in simulated soil temperatures (ΔT_{soil}) and (b) biases in simulated soil water content (ΔSWC). For soil temperatures, averages of observations <10 °C were classified as cold, 10–15 °C as fresh, 15–20 °C mild, and >20 °C as warm. For soil water content, averages of observations <40% SWC were classified as dry, 40–45% as moderately moist, 45–50% as moist, and >50% as extremely moist.

concentrations achieved lower ΔN_2O compared to models with less accurate soil NO_3^- concentrations. Similar to the analysis across models, for each model a bias in NH_4^+ concentration did not explain a bias in N_2O fluxes for any of the models (not shown), or in other words, a more accurate simulation of NH_4^+ concentrations did not increase the accuracy in N_2O estimates.

In contrast, the performance in simulating soil climatic conditions was related to the performance of ΔN_2O for some of the models. Biases in soil temperature significantly increased ΔN_2O for PaSim (Figure 5) with an overestimation of 0.0108 g N_2O-N m^{-2} per 1 °C in soil temperature (referring to the 14 day cumulative flux). Accordingly, the general overestimation in N_2O flux was reduced per degrees celsius of temperature underestimation during the 2 weeks after fertilization ($p < 0.05$). Underestimated soil temperatures were associated with underestimated N_2O fluxes (0.0047 g N_2O-N m^{-2}; $p < 0.05$) in DC2, but ΔTS did not affect ΔN_2O in DC1. Similarly, in both APSIM variants the bias in cumulative N_2O fluxes 14 days after fertilization was unaffected by ΔTS.

Underestimations of SWC during moist conditions were associated with underestimations in N_2O in DC1 (Figure 5), which was not the case for similar SWC underestimations in DC2. Overestimated N_2O fluxes by PaSim under dry conditions coincided with overestimated SWC. During several extremely moist events, PaSim underestimated SWC and overestimated N_2O emissions (Figure 5).

A general, positive bias in PaSim N_2O emissions of 0.058 N_2O-N g m^{-2} (sum of 14 days) was not explained by biases in microclimate nor systematically related to particular microclimatic conditions. Still, the higher ΔSWC, the more the N_2O flux was overestimated (0.0042 g N_2O-N m^{-2}, $p < 0.05$). In other words, the N_2O flux in PaSim was generally overestimated independent of the driver variables, but this was reduced for events coinciding with an underestimation of SWC, as the SWC effect might compensate and reduce the overestimate in N_2O flux (Figure 5). In summary, multiple regression showed that biased N_2O estimates coincided with biased temperature and or soil water content in several cases (e.g., PaSim). However, in others (e.g., APSIM), they were unaffected by the biases in soil water content and soil temperature.

A temporally explicit analysis showed that a bias in N biomass harvested did not significantly affect the subsequent three weeks' cumulative N_2O fluxes. When analyzing the effects of exported biomass N on mineral N pools, AP1 was the only model, which showed a significant negative effect on NO_3^- concentrations ($p < 0.05$), that is, APSIMs underpredicted N_2O emissions coincided with overestimated exported biomass N.

The dynamics of model performances showed the coincidence or absence of coincidence of N_2O emission and its drivers (soil temperature, soil water content, NO_3^-, and NH_4^+) in relation to the time of the year. This indicated that accurately modeled N_2O emissions did often not necessarily imply accurate

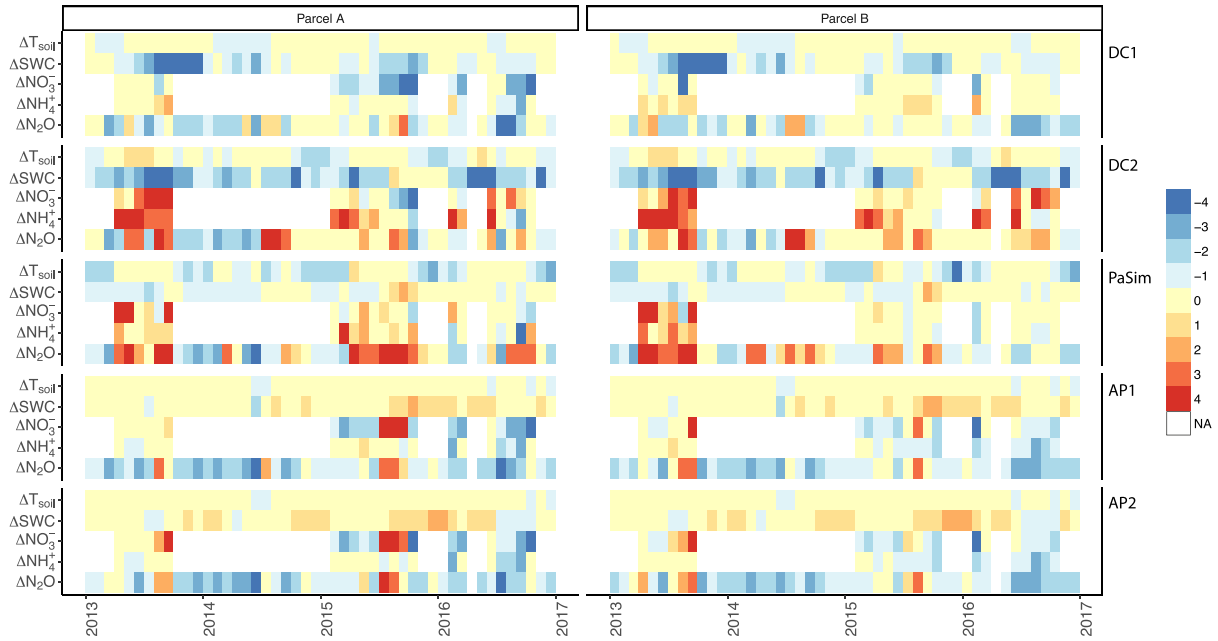

Figure 6. Monthly biases of simulated soil temperature (ΔT_{soil}), soil water content (ΔSWC), nitrate concentration (ΔNO_3^-) and ammonium concentrations (ΔNH_4^+), and N_2O fluxes (ΔN_2O) depicted as a time series. Biases are classified in categories according to their magnitude: For ΔT_{soil}, the categories 0–4 entail biases between −2.5–2.5 °C (0), 2.5–5 °C (1), 5–7.5 °C (2), 7.5–10 °C (3) and >10 °C (4). For ΔSWC, categories 0–4 entail biases of −5–5% (0), 5–10% (1), 10–15% (2), 15–20% (3), and >20% SWC (4). For ΔNO_3^- and ΔNH_4^+, categories 0–4 entail biases of −5–5 (0), 5–10 (1), 10–15 (2), 15–20 (3), and >20 kg N ha^{-1}(4). White fields indicate that no data were available.

simulations of drivers (Figure 6). For instance, in DayCent underestimations of soil water content and overestimations of soil NH_4^+ coincided with accurate N_2O emissions (e.g., DC2 and DC1 in 2015).

4. Discussion

In line with our hypotheses we found that (1) the model ensemble performed better than the IPCC-Swiss estimate, but only DayCent clearly outperformed the IPCC as an individual model on the annual time scale, (2) models followed emission peaks better after recent fertilizer application events compared to peaks in the whole observation period; (3) model's performances in N_2O emissions could partly be related to their performance in N_2O driver variables. However, in many cases a straightforward coincidence in time was not achieved, which might be due to several interacting sources of uncertainties and compensatory effects. We have highlighted different strengths and weaknesses in each model and found these to differ for each model, which are discussed in the following paragraphs.

4.1. Discussion of the Validation Results in the Context of Other Studies

Previous studies have evaluated the models DayCent, PaSim, and APSIM with N_2O flux data from lab experiments and on grassland sites under different climatic and soil conditions (Abdalla et al., 2010; Ehrhardt et al., 2018; Giltrap et al., 2015; Khalil et al., 2016; Y. Li et al., 2005; W. J. Parton et al., 2001; Stehfest & Müller, 2004; Thorburn et al., 2010; Xing et al., 2011). While some studies focused on the performance of the average N_2O flux estimates across sites (Ehrhardt et al., 2018), most evaluated the dynamics of N_2O fluxes, usually using manual chamber measurements (Abdalla et al., 2010; Giltrap et al., 2015; Khalil et al., 2016; Stehfest & Müller, 2004; Zimmermann et al., 2018). For instance, Fitton et al. (2014) showed relative biases for DayCent for three UK sites ranging from −84% to +10%. Further, Zimmermann et al. (2018) published results from Irish grassland sites with relative biases between −5% and 88% in DayCent, and further for models not used here with similar to wider biases of −116–71% in DNDC 9.4, −48–87% in DNDC 9.5, and −1,395–40% in ECOSSE. Our relative biases at Chamau (DayCent −35% to +15%, PaSim +41% and APSIM −47 to −43%) compared well with findings from other grassland sites. For RMSE, Zimmermann et al. (2018) found comparable and higher relative values for various sites between 140–

234% in DayCent, 261–503% in DNDC 9.4, 352–652% in DNDC 9.5, and 160–2,079% in ECOSSE based on daily data. RMSE in Fitton et al. (2014) ranged between 144% and 213% across sites for DayCent. The relative RMSE at our site was similar to lower with 168–173% for DayCent, 221% for PaSim, and 158−161% for APSIM. Thus, our presented validation results are within the ranges of previous findings but add reliability to the long measurement period and at high temporal (daily) resolution.

4.2. Best Fits, Blind Spots, and Phantoms Per Model

4.2.1. DayCent

Consistent with our study, for DayCent, Zimmermann et al. (2018) observed large underestimations of N_2O production by denitrification, particularly after fertilizer applications. Zimmermann et al. (2018) further showed that DayCent performed most accurately at simulating background N_2O emissions, while directly after management events DayCent did not reflect observed emission peaks, which was consistent with our findings. We frequently observed DayCent to underestimate SWC. Immediate drainage above field capacity is assumed by DayCent (and DNDC), which prevented these models simulating higher water contents than field capacity in poorly drained soils (Brilli et al., 2017) such as at our site Chamau during springtime. This can lead to inaccurate N_2O emission estimates due to underestimated soil water content. Alternatively, too low denitrification rates or too low N_2O:N_2 ratios can cause mismatches even if mineral N is simulated accurately. Stehfest and Müller (2004) found that DayCent generally simulated the denitrification-related fluxes from a urine-affected grassland in NZ well, while it overestimated total N_2O emissions by 318% due to overestimated nitrification-related N_2O. This was attributed to DayCent's fixed nitrification factor, which does not allow nitrification to happen without significant N_2O emissions. In contrast to Stehfest and Müller (2004), a clear separation between nitrification and denitrification-related N_2O emissions was not possible in our validation exercise. Soil NO_3^- concentrations were underestimated by a factor of 2–4 by DayCent in Stehfest and Müller (2004). Similarly, Senapati et al. (2016) found underestimations of NO_3^- concentrations and quite accurate NH_4^+ concentrations. We observed this in DC1 but not DC2. In DayCent, an underestimation in soil water content coincided with overestimation in NH_4^+ concentrations, potentially shifting the N_2O emissions that were related to denitrification to nitrification, and resulting in relatively accurate annual N_2O emissions, but due to inadequate reasons.

4.2.2. PaSim

Validating PaSim at three intensively and two extensively managed European sites, Calanca et al. (2007) found simulated annual N_2O emissions to be 2–10 times higher than observations, compared to 0.4 times higher at our site. The pattern of larger simulated fluxes during weeks after fertilization as shown in our study was reported by Calanca et al. (2007) and also by Schmid et al. (2001), where the N_2O emission peaks seem prolonged and as a consequence background fluxes overestimated by PaSim. In our study, we found the best PaSim performance in summer 2016 in the clover parcel. At this date, N_2O emissions occurred without direct N input by a management event. Interestingly, these fluxes were quite high, in the order of magnitude of fertilizer-induced peaks, but not depicted by most other models (AP1, AP2, and DC1). The pattern that in the nonfertilized treatment PaSim performed particularly well may be attributed to the fact that PaSim was developed to simulate N cycling in low input systems (i.e., clover-ryegrass vegetation).

Generally, PaSim simulated the mineral N concentrations relatively well and also the onset of fertilizer-related peaks. For instance, in some months all drivers were simulated accurately; however, simultaneous N_2O emissions were too high (Figure 6). As a consequence of ongoing emissions for weeks after fertilization and several phantom peaks, cumulative N_2O emissions were largely overestimated by PaSim, even if the model performed well in the simulation of mineral N. This systematic overestimation could be caused by too high nitrification and/or denitrification rates as defaults.

4.2.3. APSIM

In our study, APSIM represented the onset, magnitude, and duration of postfertilizer peaks best (Figure 4). However, APSIM omitted several observed peaks and had a general tendency to underestimate N_2O fluxes (Figure 6). APSIM's N_2O fluxes were validated by Thorburn et al. (2010), who found that the previously used default denitrification coefficient of 0.0006 mg kg^{-1} caused underestimated N_2O emissions. They suggested an optimized parameter ($k_{denit} = 0.001379$) at their location. This was not changed in the default parameterization as used here, but an informal investigation showed that the value proposed by Thorburn and colleagues would improve the N_2O predictions. Similarly, denitrification in APSIM was shown to underestimate

N_2O emissions as reported by Xing et al. (2011). Xing et al. (2011) validated denitrification N_2O emissions using soil incubation measurements. In their study, the underestimation of N_2O emissions was attributed to a too weak response of denitrification to temperature and soil moisture in the model. Xing et al. (2011) therefore suggested modifying the parameters to obtain a stronger temperature response for denitrification. Up to date, there has not been sufficient work done to propose a widely applicable response function. We here showed that an overprediction of exported biomass N was associated with underpredicted N_2O emissions in APSIM. This is not necessarily a direct causative relationship, but it does highlight the importance of accurately predicted biomass for robust N_2O flux predictions.

APSIM performed best in summer 2015, which was warm and dry (favorable for nitrification) and preceded by very wet spring conditions (very high soil water content leading to complete denitrification). Thus, we might conclude that nitrification was reflected quite well in APSIM. In contrast, denitrification was probably underestimated during the observation period, as for instance shown in summer 2016. APSIM did not predict the N_2O emission peaks during summer 2016 in Clo and underpredicted them in Ctr. The underestimated N_2O emissions coincided with underestimated soil NO_3^- (and sometimes NH_4^+) concentrations, while temperature and water conditions were simulated accurately or moderately overestimated. This suggests that low mineralization rates could be a reason for the low N_2O emissions.

5. Conclusion

We challenged three biogeochemical process models to simulate N_2O emissions and their driver variables and provided new insights into strengths and limitations of each model for facilitating decision-making. Using eddy covariance N_2O data, we overcame the limitation in temporal coverage that usually leads to large uncertainties in annual emissions estimated from sporadic chamber measurements. We recommend that PaSim's parameterization of nitrification and denitrification should be revised and potentially nitrification/denitrification rates reduced. APSIM should be used with the optimized, higher denitrification factor as suggested by Thorburn et al. (2010). DayCent predictions could likely be improved by an improved soil water module. In choosing an appropriate model, it appears that DayCent would be a good choice if the goal was to provide annual estimates. However, APSIM better reflects daily variability and therefore might be chosen if the temporal dynamics are important for the question of interest. Even though single model performance showed significant deficits, the model ensemble improved the assessment of the mitigation potential of the clover-based treatment in comparison to the IPCC-Swiss calculations. This study thus highlights some of the challenges that remain in modeling the complex biogeochemistry associated with N_2O emissions from agricultural systems.

Acknowledgments
We thank Charlotte Decock for kindly providing additional soil nitrate and ammonium data for the year 2013. This study was conducted under the Models4Pastures project within the framework of FACCE-JPI. Lutz Merbold and Kathrin Fuchs acknowledge funding for the Swiss contribution to Models4Pastures (FACCE-JPI project, SNSF funded contract: 40FA40_154245/1) and Kathrin Fuchs for the Doc.Mobility fellowship (SNSF funded project: P1EZP2_172121). Nina Buchmann acknowledges funding from GHG-Europe (FP7, EU contract 244122). The NZ coauthors acknowledge funding from the New Zealand Government Ministry of Primary Industries for supporting the Livestock Research Group of the Global Research Alliance on Agricultural Greenhouse Gases and from AgResearch's Strategic Science Investment Fund (the Forages for Reduced Nitrate Leaching (FRNL) Research Programme). The UK contributors acknowledge funding by DEFRA and the RCUK projects: N-Circle (BB/N013484/1), UGRASS (NE/M016900/1), and GREENHOUSE (NE/K002589/1) and further funding from the Scottish Government Strategic Research Programme. Lorenzo Brilli and Marco Bindi received funding from the Italian Ministry of Agricultural Food and Forestry Policies (MiPAAF). The FR partners acknowledge funding from the French National Research Agency for the CN-MIP project (ANR-13-JFAC-0001) and ADEME (12-60-C0023). Susanne Rolinski acknowledges financial support from the project Climasteppe (BMBF under grant 01DJ18012). Acknowledgment is made to the APSIM Initiative, which takes responsibility for quality assurance and a structured innovation program for APSIM. APSIM is provided free for research and development use (see www.apsim.info for details). The data used in this study are available in the ETH Research collection under http://hdl.handle.net/20.500.11850/342267 as "Dataset: Multi–model evaluation of nitrous oxide emissions from an intensively managed grassland".

References

Abdalla, M., Jones, M., Yeluripati, J., Smith, P., Burke, J., & Williams, M. (2010). Testing DayCent and DNDC model simulations of N_2O fluxes and assessing the impacts of climate change on the gas flux and biomass production from a humid pasture. *Atmospheric Environment*, *44*(25), 2961–2970. https://doi.org/10.1016/j.atmosenv.2010.05.018

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., et al. (2013). Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, *3*(9), 827–832. https://doi.org/10.1038/nclimate1916

Aubinet, M., Vesala, T., & Papale, D. (2012). *Eddy covariance: A practical guide to measurement and data analysis*. Heidelberg: Springer. Retrieved. www.springer.com/de/book/9789400723504

Baggs, E. M. (2008). A review of stable isotope techniques for N_2O source partitioning in soils: Recent progress, remaining challenges and future considerations. *Rapid Communications in Mass Spectrometry*, *22*(11), 1664–1672. https://doi.org/10.1002/rcm.3456

Ball, B. C. (2013). Soil structure and greenhouse gas emissions: A synthesis of 20 years of experimentation. *European Journal of Soil Science*, *64*(3), 357–373. https://doi.org/10.1111/ejss.12013

Barton, L.,Wolf, B., Rowlings, D., Scheer, C., Kiese, R., Grace, P., et al. (2015). Sampling frequency affects estimates of annualnitrous oxide fluxes. *Scientific Reports*, *5*(1). https://doi.org/10.1038/srep15912

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., et al. (2013). Characterising performance of environmental models. *Environmental Modelling & Software*, *40*, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011

Blagodatsky, S., & Smith, P. (2012). Soil physics meets soil biology: Towards better mechanistic prediction of greenhouse gas emissions from soil. *Soil Biology & Biochemistry*, *47*, 78–92. https://doi.org/10.1016/j.soilbio.2011.12.015

Bowatte, S., Hoogendoorn, C. J., Newton, P. C. D., Liu, Y., Brock, S. C., & Theobald, P. W. (2018). Grassland plant species and cultivar effects on nitrous oxide emissions after urine application. *Geoderma*, *323*, 74–82. https://doi.org/10.1016/j.geoderma.2018.03.001

Brilli, L., Bechini, L., Bindi, M., Carozzi, M., Cavalli, D., Conant, R., et al. (2017). Review and analysis of strengths and weaknesses of agro-ecosystem models for simulating C and N fluxes. *Science of The Total Environment*, *598*, 445–470. https://doi.org/10.1016/j.scitotenv.2017.03.208

Butterbach-Bahl, K., Baggs, E. M., Dannenmann, M., Kiese, R., & Zechmeister-Boltenstern, S. (2013). Nitrous oxide emissions from soils: How well do we understand the processes and their controls? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1621). https://doi.org/10.1098/rstb.2013.0122

Calanca, P., Vuichard, N., Campbell, C., Viovy, N., Cozic, A., Fuhrer, J., & Soussana, J.-F. (2007). Simulating the fluxes of CO_2 and N_2O in European grasslands with the Pasture Simulation Model (PaSim). *Agriculture, Ecosystems & Environment*, *121*(1–2), 164–174. https://doi.org/10.1016/j.agee.2006.12.010

Conant, R. T., Cerri, C. E. P., Osborne, B. B., & Paustian, K. (2017). Grassland management impacts on soil carbon stocks: A new synthesis. *Ecological Applications*, *27*(2), 662–668. https://doi.org/10.1002/eap.1473

Conant, R. T., Paustian, K., Grosso, S. J. D., & Parton, W. J. (2005). Nitrogen pools and fluxes in grassland soils sequestering carbon. *Nutrient Cycling in Agroecosystems*, *71*(3), 239–248. https://doi.org/10.1007/s10705-004-5085-z

Cowan, N. J., Norman, P., Famulari, D., Levy, P. E., Reay, D. S., & Skiba, U. M. (2015). Spatial variability and hotspots of soil N_2O fluxes from intensively grazed grassland. *Biogeosciences*, *12*(5), 1585–1596. https://doi.org/10.5194/bg-12-1585-2015

Davidson, E. A. (2009). The contribution of manure and fertilizer nitrogen to atmospheric nitrous oxide since 1860. *Nature Geoscience*, *2*(9), 659–662. https://doi.org/10.1038/ngeo608

Del Grosso, S., Hartman, M., Brenner, J., Mosier, A., & Ojima, D. (2001). Simulated interaction of carbon dynamics and nitrogen trace gas fluxes using the DayCent model. In S. Hansen, M. Shaffer, L. Ma, & D. Schimel (Eds.), *Modeling carbon and nitrogen dynamics for soil management* (Chap. 8, pp. 303–332). Boca Raton, Florida: CRC Press. https://doi.org/10.1201/9781420032635.ch8

Ehrhardt, F., Soussana, J.-F., Bellocchi, G., Grace, P., McAuliffe, R., Recous, S., et al. (2018). Assessing uncertainties in crop and pasture ensemble model simulations of productivity and N_2O emissions. *Global Change Biology*, *24*(2), e603–e616. https://doi.org/10.1111/gcb.13965

Environmental Protection Agency. (2019). Inventory of U.S. greenhouse gas emissions and sinks: 1990–2017. (No. Chapter 5 Agriculture.). Retrieved from https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2017

Eugster, W., & Merbold, L. (2014). Eddy covariance for quantifying trace gas fluxes from soils. *SOIL Discussions*, *1*(1), 541–583. https://doi.org/10.5194/soild-1-541-2014

Finkelstein, P. L., & Sims, P. F. (2001). Sampling error in eddy correlation flux measurements. *Journal of Geophysical Research*, *106*(D4), 3503–3509. https://doi.org/10.1029/2000JD900731

Fitton, N., Bindi, M., Brilli, L., Cichota, R., Dibari, C., Fuchs, K., et al. (2019). Modelling biological N fixation and grass-legume dynamics with process-based biogeochemical models of varying complexity. *European Journal of Agronomy*, *106*, 58–66. https://doi.org/10.1016/j.eja.2019.03.008

Fitton, N., Datta, A., Hastings, A., Kuhnert, M., Topp, C. F. E., Cloy, J. M., et al. (2014). Thechallenge of modelling nitrogen management at the field scale: simulation andsensitivity analysis of N 2 O fluxes across nine experimental sitesusing DailyDayCent. *Environmental Research Letters*, *9*(9), 095003. https://doi.org/10.1088/1748-9326/9/9/095003

Federal Office for the Environment (2018). *Switzerland's greenhouse gas inventory 1990–2016: National inventory report*. 3003 Bern, Switzerland: Federal Office for the Environment FOEN, Climate Division. Retrieved from www.climatereporting.ch

Food and Agriculture Organization. (2013). FAO statistical yearbook: World food and agriculture. FAO.

Fuchs, K., Hörtnagl, L., Buchmann, N., Eugster, W., Snow, V., & Merbold, L. (2018). Management matters: Testing a mitigation strategy for nitrous oxide emissions using legumes on intensively managed grassland. *Biogeosciences*, *15*(18), 5519–5543. https://doi.org/10.5194/bg-15-5519-2018

Gilgen, A. K., & Buchmann, N. (2009). Response of temperate grasslands at different altitudes to simulated summer drought differed but scaled with annual precipitation. *Biogeosciences*, *6*(11), 2525–2539.

Giltrap, D. L., Vogeler, I., Cichota, R., Luo, J., Weerden, T. vander, & Klein, C. de. (2015). Comparison between APSIM and NZ-DNDC models when describing N-dynamics underurine patches. *New Zealand Journal of Agricultural Research*, *58*(2), 131–155. https://doi.org/10.1080/00288233.2014.987876

Groffman, P. M., Butterbach-Bahl, K., Fulweiler, R. W., Gold, A. J., Morse, J. L., Stander, E. K., et al. (2009). Challenges to incorporating spatially and temporally explicit phenomena (hotspots and hot moments) in denitrification models. *Biogeochemistry*, *93*(1–2), 49–77. https://doi.org/10.1007/s10533-008-9277-5

Hagedorn, R., Doblas-Reyes, F. J., & Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. *Tellus A*, *57*(3), 219–233. https://doi.org/10.1111/j.1600-0870.2005.00103.x

Holzworth, D. P., Huth, N. I., de Voil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., et al. (2014). APSIM—Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, *62*, 327–350. https://doi.org/10.1016/j.envsoft.2014.07.009

Hörtnagl, L., Barthel, M., Buchmann, N., Eugster, W., Butterbach-Bahl, K., Díaz-Pinés, E., et al. (2018). Greenhouse gas fluxes over managed grasslands in Central Europe. *Global Change Biology*, *24*(5), 1843–1872. https://doi.org/10.1111/gcb.14079

Huth, N. I., Bristow, K. L., & Verburg, K. (2012). SWIM3: Model use, calibration, and validation. *Transactions of the ASABE*, *55*(4), 1303–1313.

Intergovernmental Panel on Climate Change (2008). Chapter 11: N_2O emissions from managed soils, and CO_2 emissions from lime and urea application. In H. S. Eggleston, K. Miwa, N. Srivastava, & K. Tanabe (Eds.), *2006 IPCC Guidelines for National Greenhouse Gas Inventories—A primer, Prepared by the National Greenhouse Gas Inventories Programme* (Chap. 11, pp. 11.1–11.54). Japan: Published: IGES.

Intergovernmental Panel on Climate Change (2013). Myhre, G., D. Shindell, F.-M. Bréon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, J.-F. Lamarque, D. Lee, B. Mendoza, T. Nakajima, A. Robock, G. Stephens, T. Takemura and H. Zhang, 2013: Anthropogenic and natural radiative forcing. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgley (Eds.), *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Chap. 8, pp. 659–740). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Khalil, M. I., Abdalla, M., Lanigan, G., Osborne, B., & Müller, C. (2016). Evaluation of parametric limitations in simulating greenhouse gas fluxes from Irish arable soils using three process-based models. *Agricultural Sciences*, *07*(08), 503–520. https://doi.org/10.4236/as.2016.78051

Kuypers, M. M. M., Marchant, H. K., & Kartal, B. (2018). The microbial nitrogen-cycling network. *Nature Reviews Microbiology*, *16*(5), 263–276. https://doi.org/10.1038/nrmicro.2018.9

Leifeld, J., Zimmermann, M., & Fuhrer, J. (2007). Characterization of soil carbon stocks and site-specific sequestration potentials of agricultural soils (Final Report). Zurich, Swithzerland: Agroscope Reckenholz-Tänikon Research Station ART.

Li, F., Snow, V., & Holzworth, D. (2011). Modelling the seasonal and geographical pattern of pasture production in New Zealand. *New Zealand Journal of Agricultural Research*, *54*(4), 331–352. https://doi.org/10.1080/00288233.2011.613403

Li, Y., Chen, D., Zhang, Y., Edis, R., & Ding, H. (2005). Comparison of three modeling approaches for simulating denitrification and nitrous oxide emissions from loam-textured arable soils: Modeling denitrification and nitrous oxide. *Global Biogeochemical Cycles*, *19*, GB3002. https://doi.org/10.1029/2004GB002392

McCuen, R. H., Knight, Z., & Cutter, A. G. (2006). Evaluation of the Nash–Sutcliffe efficiency index. *Journal of Hydrologic Engineering*, *11*(6), 597–602. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:6(597)

Merbold, L., Eugster, W., Stieger, J., Zahniser, M., Nelson, D., & Buchmann, N. (2014). Greenhouse gas budget (CO_2, CH_4 and N_2O) of intensively managed grassland following restoration. *Global Change Biology*, *20*(6), 1913–1928. https://doi.org/10.1111/gcb.12518

Mosier, A., Kroeze, C., Nevison, C., Oenema, O., Seitzinger, S., & van Cleemput, O. (1998). Closing the global N_2O budget: Nitrous oxide emissions through the agricultural nitrogen cycle. *Nutrient Cycling in Agroecosystems*, *52*(2–3), 225–248. https://doi.org/10.1023/A:1009740530221

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Neftel, A., Flechard, C., Ammann, C., Conen, F., Emmenegger, L., & Zeyer, K. (2007). Experimental assessment of N_2O background fluxes in grassland systems. *Tellus B*, *59*(3), 470–482. https://doi.org/10.1111/j.1600-0889.2007.00273.x

Parton, W. J., Hartman, M., Ojima, D., & Schimel, D. (1998). DAYCENT and its land surface submodel: Description and testing. *Global and Planetary Change*, *19*(1–4), 35–48. https://doi.org/10.1016/S0921-8181(98)00040-X

Parton, W. J., Mosier, A. R., Ojima, D. S., Valentine, D. W., Schimel, D. S., Weier, K., & Kulmala, A. E. (2001). Generalized model for NO_x and N_2O emissions from soils. *Global Biogeochemical Cycles*, *10*(3), 401–412. https://doi.org/10.1029/96GB01455

Probert, M. E., Dimes, J. P., Keating, B. A., Dalal, R. C., & Strong, W. M. (1998). APSIM's water and nitrogen modules and simulation of the dynamics of water and nitrogen in fallow systems. *Agricultural Systems*, *56*(1), 1–28.

Rannik, Ü., Peltola, O., & Mammarella, I. (2016). Random uncertainties of flux measurements by the eddy covariance technique. *Atmospheric Measurement Techniques*, *9*(10), 5163–5181. https://doi.org/10.5194/amt-9-5163-2016

Ravishankara, A. R., Daniel, J. S., & Portmann, R. W. (2009). Nitrous oxide (N_2O): The dominant ozone-depleting substance emitted in the 21st Century. *Science*, *326*(5949), 123–125. https://doi.org/10.1126/science.1176985

Reay, D. S., Davidson, E. A., Smith, K. A., Smith, P., Melillo, J. M., Dentener, F., & Crutzen, P. J. (2012). Global agriculture and nitrous oxide emissions. *Nature Climate Change*, *2*(6), 410–416. https://doi.org/10.1038/nclimate1458

Riedo, M., Gyalistras, D., & Fuhrer, J. (2000). Net primary production and carbon stocks in differently managed grasslands: Simulation of site-specific sensitivity to an increase in atmospheric CO_2 and to climate change. *Ecological Modelling*, *134*(2–3), 207–227. https://doi.org/10.1016/S0304-3800(00)00356-2

Rihm, B., & Achermann, B. (2016). Critical loads of nitrogen and their exceedances. Swiss contribution to the effects-oriented work under the Convention on Long-range Transboundary Air Pollution (UNECE) (Environmental studies No. 1642) (p. 78). Bern, Switzerland: Federal Office for the Environment.

Roth, K. (2006). Bodenkartierung und GIS-basierte Kohlenstoffinventur von Graslandböden (diploma thesis). University of Zurich, Zurich, Switzerland.

Ruane, A. C., Goldberg, R., & Chryssanthacopoulos, J. (2015). Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. *Agricultural and Forest Meteorology*, *200*, 233–248. https://doi.org/10.1016/j.agrformet.2014.09.016

Schmid, M., Neftel, A., Riedo, M., & Fuhrer, J. (2001). Process-based modelling of nitrous oxide emissions from different nitrogen sources in mown grassland. *Nutrient Cycling in Agroecosystems*, *60*(1-3), 177–187.

Schulze, E. D., Ciais, P., Luyssaert, S., Freibauer, A., Janssens, I. A., & Soussana, J. F. (2009). Importance of methane and nitrous oxide for Europe's terrestrial greenhouse-gas balance. *Nature Geoscience*, *2*(12), 842–850. https://doi.org/10.1038/ngeo686

Senapati, N., Chabbi, A., Giostri, A. F., Yeluripati, J. B., & Smith, P. (2016). Modelling nitrousoxide emissions from mown-grass and grain-cropping systems: Testing andsensitivity analysis of DailyDayCent using high frequency measurements. *Scienceof The Total Environment*, *572*, 955–977. https://doi.org/10.1016/j.scitotenv.2016.07.226

Smith, J., & Smith, P. (2007). *Environmental modelling: An introduction*. New York: OUP Oxford.

Snow, V., & Huth, N. (2004). The APSIM-MICROMET module. *The APSIM–Micromet Module*, *2004*(12848), 21.

Soussana, J. F., Tallec, T., & Blanfort, V. (2010). Mitigating the greenhouse gas balance of ruminant production systems through carbon sequestration in grasslands. *Animal*, *4*(3), 334–350. https://doi.org/10.1017/S1751731109990784

Stehfest, E., & Müller, C. (2004). Simulation of N_2O emissions from a urine-affected pasture in New Zealand with the ecosystem model DayCent: Simulation of N_2O from grassland. *Journal of Geophysical Research*, *109*, D03109. https://doi.org/10.1029/2003JD004261

Thorburn, P. J., Biggs, J. S., Collins, K., & Probert, M. E. (2010). Using the APSIM model to estimate nitrous oxide emissions from diverse Australian sugarcane production systems. *Agriculture, Ecosystems & Environment*, *136*(3), 343–350. https://doi.org/10.1016/j.agee.2009.12.014

United Nations Environment Programme. (1987). Montreal protocol on substances that deplete the ozone layer (No. 1522 UNTS 3; 26 ILM 1550 (1987)). Montreal. Retrieved from https://treaties.un.org/doc/Treaties/1989/01/19890101%2003-25%20AM/Ch_XXVII_02_ap.pdf

United Nations Environment Programme (UNEP). (2013). Drawing down N2O to protect climate and the ozone layer: A UNEP synthesis report. Nairobi, Kenya: United Nations Environment Programme. Retrieved from http://www.unep.org/publications/ebooks/UNEPN2Oreport/

United Nations Framework Convention on Climate Change. (2015). The Paris Agreement. Retrieved March 11, 2019, from https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

van Groenigen, J. W., Huygens, D., Boeckx, P., Kuyper, T. W., Lubbers, I. M., Rütting, T., & Groffman, P. M. (2015). The soil N cycle: New insights and key challenges. *Soil*, *1*(1), 235–256. https://doi.org/10.5194/soil-1-235-2015

Vogeler, I., & Cichota, R. (2016). Deriving seasonally optimal nitrogen fertilization rates for a ryegrass pasture based on agricultural production systems simulator modelling with a refined AgPasture model. *Grass and Forage Science*, *71*(3), 353–365. https://doi.org/10.1111/gfs.12181

Wallach, D., Martre, P., Liu, B., Asseng, S., Ewert, F., Thorburn, P. J., et al. (2018). Multimodel ensembles improve predictions of crop-environment-management interactions. *Global Change Biology*, *24*(11), 5072–5083. https://doi.org/10.1111/gcb.14411

White, R., Murray, S., & Rohweder, M. (2000). *Grassland ecosystems*. Washington, DC, USA: World Resources Institute. Retrieved from. http://www.wri.org/wr2000

Xing, H., Wang, E., Smith, C. J., Rolston, D., & Yu, Q. (2011). Modelling nitrous oxide and carbon dioxide emission from soil in an incubation experiment. *Geoderma*, *167–168*, 328–339. https://doi.org/10.1016/j.geoderma.2011.07.003

Zimmermann, J., Carolan, R., Forrestal, P., Harty, M., Lanigan, G., Richards, K. G., et al. (2018). Assessing the performance of three frequently used biogeochemical models when simulating N_2O emissions from a range of soil types and fertiliser treatments. *Geoderma*, *331*, 53–69. https://doi.org/10.1016/j.geoderma.2018.06.004