Global Research Alliance N₂O chamber methodology guidelines: Summary of modeling approaches

Donna Giltrap¹ | Jagadeesh Yeluripati² | Pete Smith³ | Nuala Fitton³ | Ward Smith⁴ | Brian Grant⁴ | Christopher D. Dorich⁵ | Jia Deng⁶ | Cairistiona FE Topp⁷ | Mohamed Abdalla³ | Liyin L. Liáng¹ | Val Snow⁸

¹ Manaaki Whenua Landcare Research, Riddet Rd., Palmerston North 4442, New Zealand
² Information and Computational Sciences Group, The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, UK
³ Institute of Biological and Environmental Sciences, Univ. of Aberdeen, 23 St Machar Dr., Aberdeen AB24 3UU, UK
⁴ Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON K1A0C6, Canada
⁵ Natural Resource Ecology Laboratory, Colorado State Univ., Fort Collins, CO 80523, USA
⁶ Earth Systems Research Center, Institute for the Study of Earth, Oceans, and Space, Univ. of New Hampshire, Durham, NH 03824, USA
⁷ Scotland’s Rural College, Peter Wilson Building, King’s Buildings, West Mains Road, Edinburgh EH93JG, UK
⁸ AgResearch Lincoln Research Centre, PB 4749, Christchurch 8140, New Zealand

Correspondence
Email: Giltrapd@LandcareResearch.co.nz

Abstract
Measurements of nitrous oxide (N₂O) emissions from agriculture are essential for understanding the complex soil–crop–climate processes, but there are practical and economic limits to the spatial and temporal extent over which measurements can be made. Therefore, N₂O models have an important role to play. As models are comparatively cheap to run, they can be used to extrapolate field measurements to regional or national scales, to simulate emissions over long time periods, or to run scenarios to compare mitigation practices. Process-based models can also be used as an aid to understanding the underlying processes, as they can simulate feedbacks and interactions that can be difficult to distinguish in the field. However, when applying models, it is important to understand the conceptual process differences in models, how conceptual understanding changed over time in various models, and the model requirements and limitations to ensure that the model is well suited to the purpose of the investigation and the type of system being simulated. The aim of this paper is to give the reader a high-level overview of some of the important issues that should be considered when modeling. This includes conceptual understanding of widely used models,
common modeling techniques such as calibration and validation, assessing model fit, sensitivity analysis, and uncertainty assessment. We also review examples of N$_2$O modeling for different purposes and describe three commonly used process-based N$_2$O models (APSIM, DayCent, and DNDC).

## 1 INTRODUCTION

A model is a numerical representation of a real-world system. The aim of a model developer is to create a model that can approximate the behaviors of interest of this system. Models always involve some simplifying assumptions, but they enable the exploration of system behavior over a wide range of conditions much faster and cheaper than is possible by direct observation. The purpose of the model will determine which aspects of the real-world system the model developer chooses to represent and how. Therefore, when selecting a model, it is important to understand the underlying model assumptions to determine whether it is fit for a purpose.

Chambers are valuable for measuring the impacts of soils, climate, and management on nitrous oxide (N$_2$O) emissions from a range of sources. However, for pragmatic reasons, observational study periods usually cover only a few years under a narrow range of conditions and are often not continuous over that time. The ability of models to operate on longer timescales, and under more manipulations, than chamber studies allows for climate variation to be accounted for in simulations, ensuring that a single anomaly year captured in a chamber study is not too heavily weighted. Models can also be used to broaden the range of climate conditions, soil–environmental drivers, and management practices that can be assessed.

Models can either be empirical (derived from observed statistical or mathematical relationships) or process based (simulation) seeking to emulate the underlying processes. Empirical models tend to be simpler to run and have fewer data requirements, but they have less flexibility in the range of conditions they can model (and their limitations are often unclear), and are unable to simulate feedbacks between processes. However, they may perform better than process-based models for the specific situations under which they were developed. A typical schematic for a process-based model and an empirical model is included in the Supplemental Figure S1.

A large number of N$_2$O models are available. However, not all models are equally suitable for all purposes. The aim of this paper is to give the reader a basic understanding of modeling principles and current practices to help them determine which approaches would be most useful for their purposes. A number of textbooks on environmental modeling (such as Smith & Smith, 2007) are available for readers seeking further information on general modeling topics.

We begin by discussing different model applications and types of N$_2$O models with examples of how these have been implemented in published studies. Next, we discuss model uncertainty, sensitivity analysis, and the fundamental processes of calibration, validation, and assessing model fit. A list of some commonly used modeling terminology is available in Supplemental Material 1.

## 2 PURPOSE OF MODEL USE

When selecting a model for a particular use, the purpose of the modeling exercise needs to be considered. For example, if the purpose is to determine the effectiveness of a mitigation option, then it is important that this mitigation is adequately simulated by the selected model (Hillier et al., 2016). In the sections below, we describe a number of common modeling purposes and identify some of the key issues that modelers will need to consider.

### 2.1 Assessing effectiveness of mitigation

Chambers measurements may identify potential practices that could mitigate N$_2$O emissions. Validated models can be used to interpolate N$_2$O emissions into continuous assessments, verify the impacts of soil–environmental drivers on emissions, and extrapolate observational assessments in order to quantify the mitigation potential across a wider range of soil properties and climate variability.

### 2.2 Running scenarios

Well-developed, calibrated, and validated models can be used to facilitate new investigations of N$_2$O emissions under varying scenarios. These may include management interactions or may extend to different crops and management practices and could include the impacts of climate variability and climate change (Table 1). The use of several models (i.e., a model ensemble) to simulate the same scenarios can provide both a measure of uncertainty and
a degree of confidence if the simulated results are aligned (Tian et al., 2019). However, only a few models exist that are capable of simulating certain practices (e.g., urease and nitrification inhibitors, intercropping, tile drainage), so full comparisons between models are not always feasible.

2.3 | Estimating cumulative emissions

Methods that sample N₂O concentrations continuously are becoming more common, yet the majority of field studies still do not sample on a continuous basis. Even for continuous measurement systems, gaps in data due to malfunction, resource issues, or adverse meteorological conditions are still frequent. As a microbially produced gas, N₂O is highly variable, and gap-filling techniques (see Dorich, Conant, & Grace, 2020) are needed to approximate missed sampling time periods and estimate the cumulative annual emissions (Barton et al., 2015; Savage, Phillips, & Davidson, 2014). The most common practice currently used is area under the curve (AUC) using linear interpolation (i.e., drawing a straight line between consecutive N₂O samples) and integrating to obtain the cumulative emission. There are known limitations to this method. Nitrous oxide emissions largely occur during peak events and not on a steady basis (Barton et al., 2015; Davidson, Keller, Erickson, Verchot, & Veldkamp, 2000; Wagner-Riddle et al., 2017). Unsurprisingly, sites that are sampled at higher frequencies have shown better results in linear interpolation methods compared with less frequently sampled sites (Barton et al., 2015). Further methods, such as look-up tables or regional-level estimates, have been used to estimate annual emissions as well (Berdanier & Conant, 2012; Mishurov & Kiely, 2011). Recently, advanced methods, such as artificial neural networks (ANN), have started to be tested and show promise to improve estimates beyond that of simple linear interpolations (Taki, Wagner-Riddle, Parkin, Gordon, & VanderZaag, 2018).

Validated biogeochemical models represent another means for estimating annual cumulative emissions. Daily time step models can approximate the processes occurring in the soil with respect to water, carbon (C), nitrogen (N), and microbial dynamics to estimate daily N₂O emissions (del Grosso et al., 2006; Li et al., 2012; Thorburn, Biggs, Collins, & Probert, 2010). The sum of modeled emissions over a year is therefore an estimate of cumulative annual emissions. Although these models may not be ideal for representing daily N₂O fluxes, they are formulated from mechanistic understanding and field data and are composed of well-understood soil processes, allowing them to more reliably represent emissions over time than other empirical methods. The cumulative annual emissions reported by a validated model are often a reliable comparison with the field data (del Grosso et al., 2009; Ehrhardt et al., 2018). One of the strongest motives for using models to estimate annual emissions is their ability to be run for long periods of time, and across many treatments. Unlike empirical models, process-based models characterize the effect of soil C and N feedbacks over time on emissions.

2.4 | Using models for scaling up

Validated models are useful for scaling up N₂O emissions estimates to larger regions. Such estimates are needed for exploring mitigation options that may reduce greenhouse gas emissions from a region, or for estimating national greenhouse gas inventories. Scaling up assessments need to be managed with regards to the spatial scale (land unit), temporal scale, and management options under investigation. This will depend on the project objectives, the time and resource costs of running at a small spatial scale, data resolution and quality, and the feasibility and reliability of models being contemplated. If the quality of the model input data and resolution between soils, crops, climate, and management are not of sufficient detail, then a simpler empirical modeling approach may provide a good average estimate.

Models vary greatly in complexity, from very simple emissions factors (i.e., Tier I IPCC [Intergovernmental Panel on Climate Change]) to empirical algorithms based on statistical synthesis of regional observations (i.e., Tier II IPCC, Cool Farm Tool; Hillier et al., 2011), meta-models (Giltrap & Ausseil, 2016), and process-based models that attempt to simulate interacting soil–plant–climate processes (Table 2). As models become more mechanistic, in principle, they should be better suited for simulating the impacts of site-specific climate, soil drivers, and management on N₂O emissions. These improvements, however, come at the cost of requiring more detailed inputs, which may not be available for larger regions or may contain more error. Mechanistic models also require rigorous validation procedures.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Model(s) used</th>
<th>Management considered</th>
<th>Crops and locations</th>
<th>( \text{N}_2\text{O} ) outcome</th>
<th>Study limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molina-Herrera et al. (2016)</td>
<td>Landscape-DNDC</td>
<td>Optimized planting, harvest, and fertilizer. Fertilizer rate and residue management.</td>
<td>Eight European cropland and grassland sites</td>
<td>Optimized management (primarily 10% less fertilizer) reduced ( \text{N}_2\text{O} ) from croplands by 21% and from grasslands by 55%.</td>
<td>Mitigation options for alternative fertilizer types, split fertilizer, legumes, alternative crops were not considered.</td>
</tr>
<tr>
<td>Abalos et al. (2016)</td>
<td>DNDC</td>
<td>4R nutrient management for nine fertilizer practices</td>
<td>Maize at two sites in eastern Canada</td>
<td>( \text{N}_2\text{O} ) from side-dress 60% lower than fall applied. Nitrification inhibitors reduced ( \text{N}_2\text{O} ) by 10%, but urease inhibitors had no impact.</td>
<td>The DNDC model had not been validated for simulating urea and nitrification inhibitors.</td>
</tr>
<tr>
<td>Sándor et al. (2018)</td>
<td>Eight process-based models</td>
<td>Reduced N fertilizer and grazing intensity</td>
<td>Five grassland sites globally</td>
<td>Emissions linearly increased with fertilizer and livestock stocking rate.</td>
<td>No comparisons against observations. Limited grassland impacts assessed due to model limitations.</td>
</tr>
<tr>
<td>Cheng, Ogle, Parton, &amp; Pan (2014)</td>
<td>DayCent</td>
<td>N inhibitor, split N, 70% reduced N, no-till</td>
<td>Maize, wheat, soybean. Global study.</td>
<td>( \text{N}_2\text{O} ) reduction was greatest for reduced N, flowed by N inhibitor, no-till and split N.</td>
<td>The scale of weather and soils data was coarse. N amendment rates were based on national level averages.</td>
</tr>
<tr>
<td>Thorburn et al. (2010)</td>
<td>APSIM</td>
<td>Fertilizer application rate, irrigation rate, residue retained or removed</td>
<td>Sugarcane in Australia</td>
<td>Reduced fertilizer and residue removal decreased ( \text{N}_2\text{O} ) and maintained yields.</td>
<td>Limited management considered. Could investigate split, slow-release fertilizer, and inhibitors.</td>
</tr>
<tr>
<td>del Grosso et al. (2009)</td>
<td>DayCent</td>
<td>Reduced fertilizer, nitrification inhibitors, split fertilizer, no-till cultivation</td>
<td>Corn, wheat, soybean. Global study.</td>
<td>Fertilizer reduction lowered ( \text{N}_2\text{O} ), but also yields, proportionally. N inhibitors reduced ( \text{N}_2\text{O} ) by 10%, whereas split and no-till showed minor reductions. Net greenhouse gas was greatly reduced for no-till.</td>
<td>The rates and scheduling of fertilizer were assumed to be identical for crop types across regions (was not soil-climate specific).</td>
</tr>
<tr>
<td>He et al. (2018)</td>
<td>DNDC</td>
<td>Crop and cultivar type under climate change, conventional vs. no-till</td>
<td>Maize, soybeans, winter wheat in Ontario, Canada</td>
<td>Increased ( \text{N}_2\text{O} ) under climate change, more so for conventional than no-till. New cultivars reduced ( \text{N}_2\text{O} ) losses.</td>
<td>Changed planting date and fertilizer rate under climate change was not considered.</td>
</tr>
</tbody>
</table>

(Continues)
<table>
<thead>
<tr>
<th>Reference</th>
<th>Model(s) used</th>
<th>Management considered</th>
<th>Crops and locations</th>
<th>( \text{N}_2\text{O} ) outcome</th>
<th>Study limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith et al. (2019)</td>
<td>DNDC</td>
<td>18 fertilizer options including fertilizer placement, timing, type (organic vs mineral). Reactive N tradeoff simulated</td>
<td>Corn silage. Two sites in eastern Canada and one in the U.S. Midwest.</td>
<td>( \text{N}_2\text{O} ) emissions were higher for organic fertilizer and fall-applied fertilizer. Relative to preplant, split and side-dress were similar for ( \text{N}_2\text{O} ) but had less reactive N loss.</td>
<td>Impacts from urease and nitrification inhibitors was not included. Impacts were only determined for silage corn.</td>
</tr>
<tr>
<td>Chen et al. (2019)</td>
<td>DNDC</td>
<td>Different mulching (no mulching, plastic mulching, and straw mulching) under climate change</td>
<td>Winter wheat and summer maize rotation in the south Loess Plateau of China.</td>
<td>DNDC predicted increased ( \text{N}_2\text{O} ) emissions under future warmer and wetter climate scenarios for all mulching practices, with plastic mulching &gt; straw mulching &gt; no mulching. Straw mulching is the optimum mulching management under future climate scenarios by considering crop yield and ( \text{N}_2\text{O} ) emission.</td>
<td>Did not consider potential changes in plant physiological characteristics and varieties along with climate change, although these changes could affect both crop production and ( \text{N}_2\text{O} ) emissions.</td>
</tr>
<tr>
<td>Deng et al. (2018)</td>
<td>DNDC</td>
<td>Alternative practices in N fertilization, tillage, irrigation, and cover crop</td>
<td>Diverse cropping systems in California, including alfalfa, wheat, lettuce, vineyards, and almond orchards.</td>
<td>( \text{N}_2\text{O} ) emissions were reduced by reducing tillage, reducing the N application rate, using low-volume irrigation, and reducing the period of fallow through cultivation of nonleguminous cover crops.</td>
<td>Only simulated short-term (several years) impacts of the management practices on ( \text{N}_2\text{O} ) emission, and long-term impacts arising from persistently applying these practices were not evaluated.</td>
</tr>
<tr>
<td>Xin and Tao (2019)</td>
<td>APSIM</td>
<td>Management (irrigation, fertilization) and cultivar selection</td>
<td>Wheat–maize rotations in the North China Plain</td>
<td>Demonstrated that substantial improvement in emissions (( \sim )40%) was possible without yield losses but requires proper selection of cultivar, as well as agronomic, irrigation, and fertilization practices.</td>
<td>Assumed that farmers have good knowledge of the soil and cultivar and that all management actions are done as planned. Did not consider long-term changes in soil pools.</td>
</tr>
</tbody>
</table>
The quality and resolution of input data available for scaling up are of great importance. National databases are often based on aggregated or estimated data, scaled up from limited measurements taken decades earlier. Gaps in weather data are also common but can be addressed using estimates derived from international databases (e.g., POWER Project datasets, https://power.larc.nasa.gov/). Most studies require some degree of input estimation or gap filling to meet the modeling requirements. For instance, Fitton et al. (2017) found that several parameters were not available, and thus values from generic site-specific calibrations were sometimes used. Due to data limitations of running the NZ-DNDC (New Zealand Denitrification–Decomposition) model, Giltrap and Ausseil (2016) ran Monte Carlo simulations to develop a regression-based meta-model that produced simplified emissions factors for N$_2$O.

The advantages of using process-based models for scaling up are that they can avoid double counting of N losses that may occur if each source of N loss (e.g., gaseous emissions and nitrate leaching) is calculated independently. As the mass balances of nutrients are considered, process-based models can simulate emissions from integrated management and often simulate a larger range of outcomes in addition to N$_2$O emissions. It is warranted to use biogeochemical models when sufficient data quality, resolution, and modeler expertise are available.

2.5 Using models for understanding and expanding on drivers and mechanisms

Process-based models describe the main processes of N cycles in ecosystems and synthesize our current understandings from experimental results at different scales. In a complex system like soil, departures between model predictions and observations are rather common for N$_2$O emissions because of the complicated interactions between soil microbiological, chemical, and physical processes and plant processes (Butterbach-Bahl, Baggs, Dannenmann, Kiese, & Zechmeister-Boltenstern, 2013). Because of the complexity, models that simulate N$_2$O emissions are often limited to specific climates and ecosystems.

Models are particularly useful for integrating knowledge acquired from experiments at the laboratory, plot, or site scale to high spatial levels (Haas et al., 2013). Nitrogen transfers and transformations across different landscapes, such as agricultural land, natural grasslands, forests, and urban regions through atmospheric and hydrological processes could potentially affect the N inputs and outputs of a given system, further changing the N$_2$O emissions (Sutton et al., 2007). The Nitroscap model (Duretz et al., 2011) showed that 10% of the total N$_2$O emissions in a
landscape composed of pig–crop farms interspersed with unmanaged ecosystems were indirect emissions from N transferred across landscapes. This indirect component is not usually captured by site-level measurements. In this case, Nitroscape provided new insights for the estimation of greenhouse gas emissions at the landscape scale. Developments in measurement techniques with instruments that provide continuous measurements of N\textsubscript{2}O fluxes at high temporal scale (minute to hour interval) have demonstrated a clear diurnal pattern of N\textsubscript{2}O fluxes (Liang, Campbell, Wall, & Schipper, 2018; Shurpali et al., 2016), suggesting that the temperature variation and interactions between plants and soil microbial activities likely play important roles in regulating the N\textsubscript{2}O fluxes at a daily scale. Therefore, neglecting these diurnal variations could introduce errors in the estimation of N\textsubscript{2}O emissions.

Pulse emissions of N\textsubscript{2}O after rain, frost–thaw, and rewetting events have also been found to make significant contributions to the annual N\textsubscript{2}O emissions (de Bruijn, Butterbach-Bahl, Blagodatsky, & Grote, 2009; Grant & Pattle, 1999; Liang et al., 2018). Several studies have investigated methods of incorporating frost–thaw effects on N\textsubscript{2}O emissions in DNDC (Denitrification–Decomposition; Dutta et al., 2018; Kariyapperuma, Wagner-Riddle, Furona, & Li, 2011, Wolf et al., 2012). The high temporal and spatial variations of N\textsubscript{2}O fluxes present a great challenge for model developments to account for reactions and processes happening at different temporal and spatial scales. Adequately describing the complexity of N\textsubscript{2}O emissions requires collaboration among modelers, plant physiologists, soil scientists, hydrologists, GIS experts, and scientists from measuring communities. Such collaborations are beneficial for model development and expanding understanding of the mechanisms for N\textsubscript{2}O emissions.

3 | EMPIRICAL MODELING

In this section, we describe the development of an empirical model to calculate country- and region-specific emissions factors for direct emissions from managed soils. Such models are often developed using chamber measurements (Rochette et al., 2018), since they are the most common set of N\textsubscript{2}O observations readily available. These models can also be used to determine the effectiveness of mitigation strategies such as fertilizer rate but are often not robust enough to empirically describe the effects of the full range of management practices on N\textsubscript{2}O emissions.

The variables included in the model formulation must be available at the national or regional scale. Data must be collected from a representative range of soils and climates within the country or region. The data required include the site information outlined in de Klein et al. (2020) and appropriate rainfall and temperature data. A categorical variable is sometimes used to represent land use (e.g., grassland or arable land). The model is developed from a set of cumulative N\textsubscript{2}O emission values measured over a consistent length of time (ideally a 12-mo period to ensure direct compatibility with the IPCC default Tier 1 emissions factors). Developing the model from the plot data as opposed to the treatment-level data allows a better understanding of the variation. Data exploration should be carried out to assess if there are discontinuities in the relationships between the emissions and the key drivers. A mixed-effect model (e.g., generalized linear mixed model (GLMM); Harrison et al., 2018) permits the definition of random terms (nuisance terms such as site and year), as well as fixed effects. The GLMM can be fitted using either restricted maximum likelihood (REML) or maximum likelihood (ML) algorithms. The REML algorithm produces unbiased estimates of variances, but biased estimates of the fixed effects, whereas the reverse is true for ML. For a much more detailed explanation, see Gelman and Hill (2007). The data should be checked to ensure that they meet the assumption of the statistical technique chosen and transformed if necessary (N\textsubscript{2}O data typically require a log transformation). The residual plots should be checked for outliers, which, if they are biologically implausibale, can be removed from the analysis.

Separate models should be fitted individually for each N source (e.g. dung, urine, farmyard manure, slurry, manufactured fertilizer, etc.). This can be further disaggregated (e.g., animal type) if sufficient data are available. It is important to note that with GLMM approaches, the order of the variables in the model fitting can affect their significance level. The model selection can be based on the deviance (Madden, Turechek, & Nita, 2002), Akaike information criterion (AIC) or Bayesian Information criterion (BIC) coefficients (Spiegelhalter, Best, Carlin, & Van der Linde, 2014).

In order to predict the emissions, in addition to N application rates, regional data for the model parameters and long-term climate data are required. The emissions should be predicted over several years for the typical N fertilizer rates and a 0 N control. Thus, the average emission factor for that region can be determined.

4 | PROCESS-BASED MODELS

Process-based models are designed to simulate the multiple impacts of biophysical processes and management practices. They have an advantage over empirical models for assessing integrated management impacts (which can be nonadditive in behavior) and implicitly account for the mass balance of N. However, they are significantly
more data intensive, less transparent, and require expertise to apply. Due to their complexity, and because some site-specific parameters are not usually measured (e.g., microbial activity), there is always some uncertainty in the outputs from process-based models. This uncertainty can be reduced, by calibration and validation, not only for N2O emission but also for important drivers of emissions such as soil temperature, water content, soil N, and crop biomass. When available, these additional measurements are valuable for understanding the drivers and processes that affect N2O emissions.

The impacts of nutrient management on N2O emission mitigation have been successfully assessed in several large-scale studies, and in each of these cases, extensive calibration was performed (Abalos et al., 2016; Molina-Herrera et al., 2016; Sándor et al., 2018). In addition to simulating N2O emissions, it is important that a model effectively simulates crop N uptake, management actions, and other N loss pathways that can strongly affect N2O emissions. When a model is used to assess mitigation strategies across contrasting crops, soils, and climate, it is not uncommon to find weaknesses in model structure and processes that are then targeted for future development.

In addition to simulating N2O emissions, process-based models are valuable for estimating other outcomes, such as crop biomass, soil organic C change, nitrate leaching and runoff, ammonia volatilization, and methane emissions, in order to examine the tradeoffs in mitigation potential of management practices. This is particularly important when examining the overall climatic impact of the emissions. Model simulations of soil processes can be used as inputs for life cycle assessments at the farm gate to include a more holistic assessment of mitigation potential (Goglio et al., 2018).

### 4.1 Overview of current process-based models

There are many different process-based models of N2O, each with different emphases and approaches to these processes. However, there are a few major models that are widely used. A search on Web of Science for “process-based model” (no quotes) and “nitrous oxide” resulted in 47 relevant papers published since 2014 (date of search: 28 Jan. 2019). A total of 28 different models were referenced across these papers. The two most commonly referenced papers were DNDC (Li, Frolking, & Frolking, 1992a, 1992b; 21 references) and DayCent (Parton, Ojima, Cole, & Schimel, 1994; 11 references). Almost 60% of publications referenced at least one of these two models. Three models shared third place with three references each: APSIM (Agricultural Production System sIMulator; Holzworth et al., 2014), SPACSYS (Wu, McGechan, McRoberts, Baddeley, & Watson, 2007; Wu et al., 2015), and IFSM (Integrated Farm System Model; Rotz et al., 2018). In this section, we describe three of these models in more detail: DNDC, DayCent, and APSIM. All three models have a structure similar to that in Supplemental Figure S1a.

### 4.1.1 DNDC

The DNDC model (Li et al., 1992a, 1992b; Li, Frolking, & Harriss, 1994) is a biogeochemical model originally developed for quantifying C sequestration and N2O emissions from croplands. It has since been expanded to simulate C and N dynamics in different ecosystems (Gilhespy et al., 2014; Giltrap, Li, & Saggar, 2010; Li, Aber, Stange, Butterbach-Bahl, & Papen, 2000; Li et al., 2012; Zhang, Li, Trettin, Li, & Sun, 2002). The model has incorporated a relatively complete suite of biogeochemical processes governing C and N cycling, including decomposition, fermentation, ammonia volatilization, nitrification, and denitrification. The DNDC model is composed of two components. The first component consists of the soil climate, crop growth, and decomposition submodels and converts primary drivers, such as climate, soil properties, vegetation, and anthropogenic activity, into soil environmental factors. The second component consists of the nitrification, denitrification, and fermentation submodels and simulates C and N transformations that are mediated by soil microbes. In DNDC, soil N primarily exists in several pools—organic N, ammonium, ammonia, and nitrate. Dynamics of soil N in each pool are simulated at an hourly or daily time step through a series of biogeochemical reactions: decomposition, microbial assimilation, plant uptake, ammonia volatilization, ammonium adsorption, nitrification, denitrification, and nitrate leaching. Fluxes of N gases (i.e., nitric oxide [NO], N2O, and dinitrogen [N2]) are predicted as either products or intermediate products by simulating the relevant N transformation processes, primarily nitrification and denitrification.

The DNDC model has been successfully applied in different ecosystem types, including cropland, forest, grassland, wetland, peatland, and livestock farms (Deng et al., 2014; Gilhespy et al., 2014; Giltrap et al., 2010; Li et al., 2012). The model simulations have been extensively evaluated against datasets of N2O as well as methane fluxes that were measured worldwide (Gilhespy et al., 2014; Giltrap et al., 2010). The DNDC model is freely available at http://www.dndc.sr.unh.edu/.
4.1.2 | DayCent

DayCent is a biogeochemical model that simulates the C and N fluxes between the soil, atmosphere, and vegetation for forest, grassland, and cropland ecosystems. Based on the CENTURY model (Parton, Hartman, Ojima, & Schimel, 1998), DayCent (del Grosso et al., 2001, 2006; Parton et al., 2001) operates on a daily times step and includes subroutines that control soil organic matter pools, soil water content, soil temperature, methane oxidation, and N emissions via nitrification and denitrification. DayCent also includes a plant growth submodel with dynamic C allocation among the above- and belowground biomass pools, where plant growth is dependent on temperature and water or nutrient limitations. Soil C is distributed into a microbial pool and three soil organic matter pools with distinct decomposition rates.

To date, the model has been successfully used in a wide range of ecosystem types (cropland, grassland, and forest) in a number of different countries (Abdalla, Jones, Ambus, & Williams, 2010; Cameron et al., 2013; Fitton et al., 2014; Sansoulet et al., 2014). However, there are a number of limitations to the model structure, including a relatively simplistic grazing submodel, as well as an inability to simulate multiple plant species, meaning that for grassland ecosystems, in particular, there is an underrepresentation of the more complex species mixtures or management types associated with them.

The DayCent model, example simulations, and a publication list are freely available via the Natural Resource Ecology Laboratory (NREL) website (www2.nrel.colostate.edu/projects/daycent-home.html)

4.1.3 | APSIM

The APSIM model is a process-based model of the soil–plant–atmosphere system with a strong emphasis on the effect of management on the production system. It can run in single- or multi-point mode (e.g., multiple paddocks on a farm) and is designed to run over many years with changing management. The APSIM model’s origins were in cropping rotations, but it has been continually developed and successfully used in pasture (Snow, Smale, & Dodd, 2013), pasture mixtures (Fitton et al., 2019), agroforestry (Dilla, Smethurst, Barry, Parsons, & Denboba, 2018; Huth, Carberry, Poulton, Brennan, & Keating, 2002), and tree crops such as oil palm (Elaeis guineensis Jacq.; Huth, Banabasb, Nelsonc, & Webb, 2014). Holzworth et al. (2014) detailed the complex lineage of APSIM’s development. The APSIM model primarily uses a daily time step, but some internal models use shorter steps. The model includes many dynamic crop (including trees and pasture) models with dynamic partitioning of C and N to above- and belowground pools. A particular strength of APSIM is its inclusion of responsive rule-based management (Moore et al., 2014) where management practices can be dynamically determined by crop or soil conditions. The soil C and N processes operate on a daily time step (Probert, Dimes, Keating, Dalal, & Strong, 1998) with denitrification and N$_2$O emissions (from both nitrification and denitrification) based on DayCent methods (Thorburn et al., 2010). The APSIM model’s origins are in low-input agriculture, so it emphasizes the role of decomposition of plant litter in replenishing soil pools (Probert et al., 1998).

Recent advances include better representation of nutrient returns from grazing animals (Snow, Cichota, McAuliffe, Hutchings, & Vejlin, 2017). Limitations of the model include lack of freeze–thaw processes, transport processes through the soil for generated N$_2$O, soil CH$_4$ emissions or uptake, and the relatively (e.g., compared with DayCent) simple pool structure for humic soil organic matter.

The APSIM model is free for noncommercial purposes from https://www.apsim.info/, and that website includes links to documentation and training material. The source code for APSIM 7.x (development of this code base ceased in 2017) is available from https://github.com/APSIMInitiative/APSIMClassic, and the code for APSIM Next Generation (Holzworth et al., 2018) is available from https://github.com/APSIMInitiative/ApsimX.

4.2 | Modeling networks and online resources

Online resources and networks have been developed to facilitate collaboration between international researchers using some of the more common N$_2$O models. These include the Global Research Alliance Modeling Platform (GRAMP) and the Global DNDC Network, which are described below.

4.2.1 | Global Research Alliance Modeling Platform (GRAMP)

The GRAMP platform (http://www.gramp.org.uk/) is a platform for modelers to share information and connect with each other (Yeluripati et al., 2015). The website currently features two biogeochemical models: DNDC and ECOSSE (Estimation of Carbon in Organic Soils—Sequestration and Emissions; Smith et al., 2010). The website includes forums, publication lists, and field data.
4.2.2 Global DNDC Network

The Global DNDC Network (http://www.globaldnnc.net/) is an international network of researchers and model developers using the process-based DNDC mode. The website contains a list of relevant publications and news of upcoming events. There is also an email discussion group for model users.

4.3 Comparisons of models

Worldwide, various simulation models have been developed and applied to predict N₂O emissions. However, these models vary in the level of detail and how the processes are described. Brilli et al. (2017) identified almost 200 different processes and approaches across nine different agroecosystem C and N flux models. On the basis of published modeling studies, the most common weaknesses in the models tended to be in the simulation of pedoclimatic conditions, rather than management practices.

Table 3 compares three models (DNDC, DayCent, and APSIM) and presents the key differences and similarities with respect to their simulation of nitrification, denitrification, and N₂O emissions. Important considerations when selecting a process-based model include the input data requirements, whether the model produces all the output data needed, and whether it is capable of simulating the climate, management practices, and crop types of interest. Depending on the research question, it may be necessary to examine the modeled processes in more detail (e.g., to investigate the effect of increased atmospheric carbon dioxide [CO₂] on N₂O emissions, a model that accounts for the effect of atmospheric CO₂ on plant growth should be selected). Reviewing recent literature will also help identify models that have been used in similar studies and how well they performed.

5 UNCERTAINTY IN MODELS

5.1 Sources of uncertainty

There are two main types of errors that can affect a model: model errors and propagation of input uncertainties. Although random errors and sampling errors are common in measured data, most models are deterministic and have no variability in their output for a given input.

5.1.1 Model errors

Model errors are the result of differences between reality and the way it is represented in the model. These differences could be due to simplifying assumptions, missing processes, or misconceptions about how the system works. In many cases, model errors cannot be quantified. However, there are a few instances where it may be possible to estimate the effects of known simplifications (e.g., when a simplified version of a more complex equation is used).

5.1.2 Propagation of input uncertainties

The input data used to run a model simulation will always have some uncertainty attached to it. Likewise, there can be uncertainty in the model parameters. These uncertainties propagate through the model, leading to uncertainty in the model output. The impact of uncertainties in the model input on the model output can be investigated using the methods described in Section 5.1.3.

5.2 Uncertainty and sensitivity analysis

Sensitivity analysis and uncertainty analysis are related concepts. However, whereas uncertainty analysis is concerned with the propagation of uncertainty from model inputs and parameters up to the model output, sensitivity analysis is the assessment of the relative importance of the sources of uncertainty in the output uncertainty (Crosetto & Tarantola, 2001).

5.2.1 One-at-a-time

Most models have multiple inputs, and frequently multiple outputs. The simplest form of uncertainty or sensitivity analysis is to look at the effect of varying a single input variable on a single output variable assuming that all other inputs are held constant. For an uncertainty assessment, one might use the extreme values of the input variable or a probability distribution function (PDF).

A dimensionless “sensitivity” value can be calculated to compare the relative effects of changes in different inputs on the model output. Some methods for calculating model sensitivity are described in Supplemental Material 2. As many process-based models have a large number of input variables, sensitivity analysis can be a useful tool to determine which variables will have the most effect on model uncertainty and which can be safely neglected. Morris
<table>
<thead>
<tr>
<th>Parameter</th>
<th>DNDC</th>
<th>DayCent</th>
<th>APSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documentation</td>
<td><a href="http://www.dndc.sr.unh.edu">http://www.dndc.sr.unh.edu</a></td>
<td><a href="http://www.nrel.colostate.edu/projects/daycent/">http://www.nrel.colostate.edu/projects/daycent/</a></td>
<td></td>
</tr>
<tr>
<td>Accessibility</td>
<td>Executable files available to download, source code on request</td>
<td>On request</td>
<td>Free for noncommercial use, download from GitHub, registration required</td>
</tr>
<tr>
<td>Key input data</td>
<td>Daily meteorological data (max. and min. air temperatures, precipitation); soil properties (soil texture, bulk density, clay content, organic matter content, and pH); plant growth parameters; management practices (e.g., planting and harvest, tillage, fertilization, manure amendment, irrigation, flooding, and residue management)</td>
<td>Daily meteorological data (max. and min. temperature, precipitation; soil texture; plant growth parameters; management practices (timing and intensity of events such as harvest or fertilization)</td>
<td>Daily meteorological data (max. and min. temperature, precipitation, solar radiation); soil properties (bulk density, cardinal water contents, pH); crop and animal parameters; management practices (date-based or dynamic)</td>
</tr>
<tr>
<td>Quality of N\textsubscript{2}O simulation</td>
<td>Good\textsuperscript{a}</td>
<td>Good\textsuperscript{a}</td>
<td>Good\textsuperscript{a}</td>
</tr>
<tr>
<td>Initialization</td>
<td>Spin-up for 2–3 yr</td>
<td>Spin-up for &gt;1,000 yr</td>
<td>Spin-up for 2–20 yr</td>
</tr>
<tr>
<td>Spatial scale</td>
<td>Field/regional</td>
<td>Field/regional</td>
<td>Field/regional</td>
</tr>
<tr>
<td>Time step</td>
<td>Hourly for nitrification–denitrification; daily for other processes</td>
<td>Daily</td>
<td>Mostly daily, some submodels as shorter dynamic time steps</td>
</tr>
<tr>
<td>Output time steps</td>
<td>Daily/annual</td>
<td>Daily/monthly/annual</td>
<td>User defined</td>
</tr>
<tr>
<td>Ecosystem</td>
<td>Arable land/grassland/forest/wetland</td>
<td>Arable land/grassland/forest</td>
<td>Arable land/grassland/grazed pasture/managed forestry</td>
</tr>
<tr>
<td>Max. simulation depth, cm</td>
<td>50</td>
<td>20</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Range of management options</td>
<td>Very broad</td>
<td>Limited</td>
<td>Very broad</td>
</tr>
<tr>
<td>Multiple simultaneous crops</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Phenological response to N stress</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes (but crop dependent)</td>
</tr>
</tbody>
</table>

(Continues)
<table>
<thead>
<tr>
<th>Parameter</th>
<th>DNDC</th>
<th>DayCent</th>
<th>APSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>C/N ratio</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Temperature</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Moisture</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Crop module</td>
<td>Simple</td>
<td>Simple</td>
<td>Advanced</td>
</tr>
<tr>
<td>Over-winter processes</td>
<td>Snow dynamics, soil freeze–thaw</td>
<td>Snow dynamics, soil freeze–thaw</td>
<td>No</td>
</tr>
<tr>
<td>N transformations expressed by</td>
<td>Microbial growth</td>
<td>Microbial growth</td>
<td>Microbial growth and abiotic factors</td>
</tr>
<tr>
<td>Release of produced gases affected by soil physical properties</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Denitrification kinetics</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Nitrification kinetics</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>$N_2$ from denitrification calculated</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>$N_2O$ losses from nitrification</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Dissolved organic N</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

*Previous studies have shown good correlations between the field measured and simulated N$_2$O.*
presented a scheme for determining which inputs have important effects by performing a series of individually randomized one-at-a-time simulations.

5.2.2  Most sensitive factor

Li, Narayanan, and Harriss (1996) investigated the sensitivity of the DNDC model to variations in soil parameters by running simulations with each parameter at its maximum or minimum (with the others held at the mean values) and comparing model results with the extreme scenario when all parameters were at an extreme. The parameter that resulted in the greatest output variability was termed the most sensitive factor (MSF). Different outputs were sensitive to different factors. Therefore, in regional mode, DNDC estimates the model uncertainty by running the model with the maximum and minimum values of the MSF for the output of interest.

5.2.3  Monte Carlo analysis

The disadvantage of one-at-a-time sensitivity analysis is that it does not account for interactions between input variables. To account for these, it is necessary to investigate varying multiple parameter values simultaneously. In a Monte Carlo analysis, a PDF is estimated for each input variable of interest. This can include correlations between variables if known. These PDFs are then used to randomly generate a large number of input datasets. The model is then run for each set of input data values and the results are analyzed statistically. One easy statistic to estimate from Monte Carlo output is the 95% confidence, which is simply the 2.5th and 97.5th percentiles of the results. Sobol (2001) outlines more advanced techniques for global sensitivity analysis.

6  CALIBRATION AND VALIDATION

Values of model parameters vary, depending on the system simulated so model parameter values need to be established in each new system. The most common procedure for finding model parameter values (calibration) is by using an optimization method. This involves assigning initial values to parameters (based on some prior knowledge about the system). Simulated model results are then compared with the corresponding observations and parameters adjusted until a satisfactory solution is obtained. This comparison could be a simple visual comparison or could use one of the goodness-of-fit measures described in the section below.

This method works well for simple models with few parameters. However, as many biogeochemical models are very complex with tens to hundreds of parameters, leading to a vast number of possible combinations, it is common for calibration to be carried out using only a subset of parameters. Sensitive parameters can be identified by model sensitivity analysis to reduce the number of parameters that need to be calibrated (Babu, Li, Froeling, Nayak, & Adhya, 2006). This can lead to biases due to assumptions made by individual modelers, and this effect is evident in Ehrhardt et al. (2018), where multiple groups used the same models and produced different results.

With increasing availability of computational power automation of the calibration process is possible (Necpálová et al., 2015). Automation requires the criteria to evaluate model fit to be predefined. In the literature, several statistical functions have been used for this purpose (Babu et al., 2006; Beheydt, Boeckx, Sleutel, Li, & Van Cleemput, 2007). Once the model parameter values have been fitted, the model can be validated by testing the model fit using a data set independent of that used for calibration.

7  GOODNESS-OF-FIT MEASURES

Quantitative analysis of a model simulation tells us how well the simulated values match measured data. Two types of analyses are most frequently used—analysis of coincidence and analysis of association—and a thorough analysis will include both. A model with a small difference (i.e., high coincidence) and high association between simulated and measured values simulates the measured data accurately. It is possible, however, for the difference between simulated and measured values to be small, but the association to be low, suggesting that the fit between simulated and measured data may just be happenstance. Equally, it is possible for the measured and simulated values to be closely associated, but the difference between them to be high, suggesting the model is good, but some systematic error is causing a shift in the simulations and preventing a good fit.

For all analyses, the different statistical tests tell us different things about the goodness-of-fit of the model (Smith & Smith, 2007). For example, the correlation coefficient $r$ is a measure of how well trends in measured values are simulated; $RMSD$ (Loague & Green, 1991; Smith & Smith, 2007) is a measure of the total error; mean difference ($M$) or the relative error ($E$) (Addiscott & Whitmore, 1987) indicate bias or systematic error; and lack of fit (LOFIT; Whitmore, 1991) separates model error from variation in the measurements. Supplemental Material 3 contains the details of these measures of model fit and how they are
TABLE 4  Summary of modeling approaches

<table>
<thead>
<tr>
<th>Issue</th>
<th>Types of approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose of model</td>
<td>Assessing effectiveness of mitigation</td>
</tr>
<tr>
<td></td>
<td>Running scenarios</td>
</tr>
<tr>
<td></td>
<td>Estimating cumulative emissions</td>
</tr>
<tr>
<td></td>
<td>Scaling up</td>
</tr>
<tr>
<td></td>
<td>Understanding drivers and mechanisms</td>
</tr>
<tr>
<td>Types of model</td>
<td>Empirical</td>
</tr>
<tr>
<td></td>
<td>Process-based</td>
</tr>
<tr>
<td>Sensitivity analysis</td>
<td>One parameter at a time</td>
</tr>
<tr>
<td></td>
<td>Monte Carlo</td>
</tr>
<tr>
<td></td>
<td>Global/local</td>
</tr>
<tr>
<td>Goodness of fit</td>
<td>Correlation coefficient</td>
</tr>
<tr>
<td>Measures of association</td>
<td>RMSD</td>
</tr>
<tr>
<td>Measures of coincidence</td>
<td>Mean error/relative mean error</td>
</tr>
<tr>
<td>Measures of bias</td>
<td>LOFIT*, lack of fit</td>
</tr>
<tr>
<td>Separating model error from measurement</td>
<td></td>
</tr>
</tbody>
</table>

*LOFIT, lack of fit.

calculated. Other helpful references include Legates and McCabe (1999), Moriasi et al. (2007), and Bennett et al. (2013).

The significance tests described in the supplemental material assume a normal distribution. However, N₂O emissions are frequently right skewed rather than normally distributed (Giltrap, Berben, Palmada, & Saggar, 2014). For this reason, the normality of the distribution of difference between measured and simulated values should be checked using standard statistical tests (e.g., the Shapiro–Wilk test), and if the distribution is not normal, an appropriate transformation applied. Examples of model assessment using these goodness-of-fit measures can be seen in Abdalla, Jones, Yeluripati, et al. (2010), Abdalla et al. (2014), and Bell et al. (2012).

8 | CONCLUSIONS

Models have the potential to significantly increase our understanding of N₂O emissions from soils compared with what would be possible from measurements alone. Potential applications of N₂O modeling include gap filling of measured data; upscaling emissions estimates to a regional, national, or global scale; determining the long-term variability of N₂O emissions; and running scenarios to assess the potential benefits of mitigation strategies (Table 4). However, the processes controlling N₂O emissions are complex, and it is important to ensure that the model selected is fit for purpose and well calibrated.

ACKNOWLEDGMENTS

Funding for this publication was provided by the New Zealand Government to support the objectives of the Livestock Research Group of the Global Research Alliance on Agricultural Greenhouse Gases. Individual authors’ work contribute to the following projects for which support has been received: Climate Smart Use of Norwegian Organic Soils (MYR, 2017–2022) project funded by the Research Council of Norway (Decision no. 281109); Scottish Government’s Strategic Research Programme, SuperG (under EU Horizon 2020 Programme); DEVIL (NE/M021327/1), Soils-R-GREAT (NE/P019455/1), and the EU H2020 Project under Grant Agreement 774378—Coordination of International Research Cooperation on Soil Carbon Sequestration in Agriculture (CIRCASA); Project J-001793, Science and Technology Branch, Agriculture and Agri-Food Canada; and New Zealand Ministry of Business, Innovation and Employment (MBIE) core funding. Thanks to Alasdair Noble and the anonymous reviewers for helpful comments on a draft of this paper, and to Anne Austin for editing services.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Donna Giltrap  https://orcid.org/0000-0001-7919-0414
Ward Smith  https://orcid.org/0000-0001-7462-5247
Christopher D. Dorich  https://orcid.org/0000-0001-8005-5797
Val Snow  https://orcid.org/0000-0002-6911-8184
REFERENCES


