Received: 2 May 2020 | Accepted: 6 August 2020 | Published online: 15 September 2020

DOI: 10.1002/jeq2.20138

Journal of Environmental Quality

SPECIAL SECTION:
GRA N_2O CHAMBER METHODOLOGY GUIDELINES

Global Research Alliance N_2O chamber methodology guidelines: Guidelines for gap-filling missing measurements

Christopher D. Dorich[1] | **Daniele De Rosa**[2] | **Louise Barton**[3] | **Peter Grace**[2] | **David Rowlings**[2] | **Massimiliano De Antoni Migliorati**[2] | **Claudia Wagner-Riddle**[4] | **Cameron Key**[5] | **Daqi Wang**[1] | **Benjamin Fehr**[6] | **Richard T Conant**[1]

[1] Natural Resource Ecology Lab., Colorado State Univ., Fort Collins, CO 80523, USA

[2] Institute for Future Environments, Queensland Univ. of Technology, Brisbane, QLD 4000, Australia

[3] SoilsWest, School of Agriculture and Environment (M079), Faculty of Science, The Univ. of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

[4] School of Environmental Sciences, Univ. of Guelph, Guelph, ON N1G 2W1, Canada

[5] Dep. of Electrical and Computer Engineering, Colorado State Univ., Fort Collins, CO 80523, USA

[6] Dep. of Agriculture, Nutrition and Food Systems, Univ. of New Hampshire, Durham, NH 03823, USA

Correspondence
Christopher D. Dorich, Natural Resource Ecology Lab., Colorado State Univ., Fort Collins, CO 80523, USA.
Email: chris.dorich@colostate.edu

Assigned to Associate Editor Mike Harvey.

Funding information
USDA AFRI, Grant/Award Numbers: 2018, -67019-27812; USDA Office of the Chief Economist, Grant/Award Number: 58-0111-19-019

Abstract

Nitrous oxide (N_2O) is a potent greenhouse gas that is primarily emitted from agriculture. Sampling limitations have generally resulted in discontinuous N_2O observations over the course of any given year. The status quo for interpolating between sampling points has been to use a simple linear interpolation. This can be problematic with N_2O emissions, since they are highly variable and sampling bias around these peak emission periods can have dramatic impacts on cumulative emissions. Here, we outline five gap-filling practices: linear interpolation, generalized additive models (GAMs), autoregressive integrated moving average (ARIMA), random forest (RF), and neural networks (NNs) that have been used for gap-filling soil N_2O emissions. To facilitate the use of improved gap-filling methods, we describe the five methods and then provide strengths and challenges or weaknesses of each method so that model selection can be improved. We then outline a protocol that details data organization and selection, splitting of data into training and testing datasets, building and testing models, and reporting results. Use of advanced gap-filling methods within a standardized protocol is likely to increase transparency, improve emission estimates, reduce uncertainty, and increase capacity to quantify the impact of mitigation practices.

1 | INTRODUCTION

Soil nitrous oxide (N_2O) emissions are notoriously variable—through time, across space, and with

Abbreviations: ARIMA, autoregressive integrated moving average; EF, emission factor; GAM, generalized additive model; MLP, multilayer perceptron; NN, neural network; RF, random forest.

management and environmental conditions (Galloway et al., 2003). Drivers of soil N_2O emissions can be largely broken down into three categories: availability of substrates (nitrogen [N], carbon, oxygen, etc.), site factors (soil properties, climate, management, etc.), and biological (microbial and fungi) composition (Davidson, Keller, Erickson, Verchot, & Veldkamp, 2000; McDaniel et al., 2017). These processes and their interactions with environmental variables lead to large spatial (hot spots) and temporal (hot moments) N_2O variability (McDaniel et al., 2017). Nitrogen application rates, and N surplus, play a leading role in these soil N_2O emissions, often showing a nonlinear annual response in emissions with increasing N surplus (Hoben, Gehl, Millar, Grace, & Robertson, 2011; Kim, Hernandez-Ramirez, & Giltrap, 2013; Shcherbak, Millar, & Robertson, 2014). Large pulses are also common after wetting events (Abdalla, Jones, Smith, & Williams, 2009; Rafique, Anex, Hennessy, & Kiely, 2012), as well as in response to soil disturbance events (e.g., freeze–thaw cycles; Congreves, Wagner-Riddle, Si, & Clough, 2018; Flesch et al., 2018; Flessa, Dorsch, & Beese, 1995; Ludwig, Wolf, & Teepe, 2004; Teepe, Brumme, & Beese, 2001; Wagner-Riddle et al., 2017). Emission peaks typically make up a considerable portion of annual emissions: up to 90% for freeze–thaw emissions (Wagner-Riddle et al., 2017), up to 97% postharvest and after summer rainfall in semiarid regions (Barton, Hoyle, Stefanova, & Murphy, 2016), up to 50% within the first weeks after N fertilizer events (Shcherbak et al., 2014), and up to 30% within a few weeks of manure application (Chadwick et al., 2011). Peak events have also been observed during tillage, residue management (Scheer et al., 2017), and irrigation, though these have been less studied and often coincide with N application events.

Emission factors (EFs) of N_2O (i.e., N_2O emitted as a fraction of applied N after correcting for background emissions) can only be well characterized if measurements are conducted over an extended period of time (de Klein et al., 2020). Long-term, multiyear field studies are recommended, as they can help determine mean annual and seasonal N_2O emissions, helping to differentiate the impact of management and climate on emissions. However, gap-free sampling is a rarity in greenhouse gas studies due to budgetary or labor constraints, equipment malfunction, and field access that prevent continuous time series data. Nitrous oxide meta-analyses (Albanito et al., 2017; Cayuela et al., 2017; Dorich et al., 2020; Stehfest & Bouwman, 2006; Shcherbak et al., 2014; Verhoeven et al., 2017) have shown that a majority of research has historically sampled for <50 d annually (<14% of the year), while fewer studies measured for >200 d in a year. Given the variable nature of N_2O emissions and the sporadic sampling, there is potentially large uncertainty in N_2O emissions due to the necessity of gap-filling between sampled data. Manual chamber studies generally have more gaps to fill and thus are more susceptible to uncertainty during the gap-filling process (Shang et al., 2020). Automatic chamber methods, which improve time series sampling, are being used more frequently and can be used to develop gap-filling methods for all N_2O flux data, but especially for manual chamber studies. Addressing gaps within sampling is required, as sampling limitations have been shown to affect N_2O emission estimates and thus EF estimates (Barton et al., 2015; Mishurov & Kiely, 2011; Parkin, 2008; Reeves & Wang, 2015; Shang et al., 2020; Smith & Dobbie, 2001). Ensuring that data from sampling campaigns are turned into unbiased and accurate representations of estimates is essential for accurately quantifying the impacts of management practices on emissions, as well as developing sound mitigation strategies (Parkin, 2008; Shang et al., 2020).

Core Ideas

- N_2O emissions are highly variable, and sampling strategy can affect cumulative emissions.
- Addressing gaps within sampling is required, as sampling limitations can affect emissions.
- There is a need to update methodologies within soil N_2O research to improve gap-filling.

Although there is known uncertainty in some areas of N_2O research (Kravchenko & Robertson, 2015), uncertainty associated with estimates reliant on gap-filling has not been well documented. As manual chamber sampling was the dominant methodology for soil N_2O measurement during the formative years of biogeochemical models (Del Grosso, Mosier, Parton, & Ojima, 2005; Giltrap, Li, & Saggar, 2010; Li, 1996) and in the formulation of EFs (IPCC, 2006), the uncertainty associated with large gaps in data is likely incorporated into these methods. A recent meta-analysis (Shang et al., 2020) of N_2O EFs showed that sampling method, strategy, frequency, and other methodological issues are often not reported with enough detail to be used within analyses. This incomplete reporting results in uncertainty in methodology that makes data comparison difficult (Albanito et al., 2017; Cayuela et al., 2017; Charles et al., 2017; Shang et al., 2020; Shcherbak et al., 2014; Verhoeven et al., 2017).

The standard practice for deriving an emission estimate over a specified time period (cumulative estimate) from sampled field points is to perform a simple linear interpolation between points and use the average value. Using linear interpolation with a sampling strategy that "chases

peaks" by sampling around expected high emission rates can invariably overestimate emissions. Similarly, studies that happen to miss these peak events are likely to underestimate emissions. Further, linear interpolation is generally more unreliable as the number and duration of gaps increases. With increasing gap numbers or gap duration it becomes more likely that gap-filling will be used over periods where soil conditions and chemistry, and thus emission rates, may change—adding to uncertainty in emission estimates. Beyond the already-mentioned challenge of underestimating emissions due to missing peak emission events, sampling strategies that do not sample outside of the growing season present issues, as they may be difficult to extrapolate to an annual value or accurate EF (Scheer et al., 2017; Wagner-Riddle et al., 2017). Although some studies have examined the influence of sampling frequency from chamber methods on reported annual emissions (Barton et al., 2015; Mishurov & Kiely, 2011; Savage, Phillips, & Davidson, 2014), research into determining and testing gap-filling methods has been limited (Bigaignon, Fieuzal, Delon, & Tallec, 2020; Cowan et al., 2019; De Rosa et al., 2018; Taki, Wagner-Riddle, Parkin, Gordon, & VanderZaag, 2018; Webb et al., 2019). Thus, there is a need for updating methodologies within soil N_2O research to improve gap-filling and develop estimation methods.

Although soil N_2O emissions are inherently more difficult to measure and predict than soil carbon dioxide (CO_2), the state of gap-filling methods in N_2O emissions trail efforts for other ecosystem exchange measures, like that used in eddy covariance CO_2 fluxes (Barba et al., 2018; Hoffmann et al., 2015; Moffat et al., 2007; Wutzler et al., 2018) or methane (CH_4) emissions (Dengel et al., 2013; Kim et al., 2019). Examination of some of these methods within soil N_2O emissions (Bayesian estimates of emissions [Cowan et al., 2020], autoregressive integrated moving averages [ARIMA, De Rosa et al., 2016; De Rosa et al., 2018], generalized additive model [GAMs; Cowan et al., 2019; De Rosa et al., 2020; Webb et al., 2019], and artificial neural networks [NNs; Bigaignon et al., 2020; Taki et al., 2018]) has begun, showing promise in their ability to improve estimates. However, there is little understanding about conditions under which these models can successfully be implemented, essential data requirements, and sampling strategies and limitations. Answering these questions is important for determining which gap-filling method to use and how to implement it, as well as potentially being useful for informing sampling strategies themselves. These advanced statistical methods provide an opportunity to better estimate or gap-fill N_2O emissions by using other associated variables (covariates) that are measured in the field and are known drivers of soil N_2O emissions (e.g., soil water content and temperature, soil inorganic N, climate information, soil physics and chemistry, vegetation information, biological indices). Although some of these covariates can be difficult or time consuming to measure (e.g., inorganic N, microbial populations, or biomass) and thus have sporadic measurements of their own, others (e.g., soil water content and temperature or weather data) are often readily available at a low cost and are thus measured continuously. Using these covariates within advanced statistical methods provides an opportunity to improve N_2O gap-filled estimates compared with linear interpolation by incorporating covariate drivers of N_2O emissions.

Improved gap-filling methodology is required for soil N_2O emissions in order to (a) estimate daily N_2O emissions when sampling was not conducted, (b) improve annual and seasonal estimates, (c) reduce uncertainty in emission estimates, and (d) ensure that mitigation practices are reducing emissions. This guidelines paper addresses the first topic by outlining the various gap-filling techniques that have been implemented for N_2O emissions. We evaluate strengths and weaknesses and present examples of how to select the best methods so that N_2O researchers unfamiliar with gap-filling can better utilize the methods. In the future, it is possible that other statistical techniques like generalized linear models (Davis et al., 2017), recurrent NNs, and other methods may be used for gap-filling. Here, however, we review several approaches selected because they are the ones that have been most commonly used for gap-filling within the trace gas flux research community. Further, and possibly more importantly, we outline a protocol for testing and reporting gap-filling methods. Consistent and appropriate use of these methods, as well as the adequate reporting of use, are needed so that comparisons and meta-analysis can be conducted. The gap-filling protocol we outline was developed using guidelines to help researchers utilize consistent, replicable, and comparable methods (Table 1, adapted from Wu, Dandy, & Maier, 2014).

2 | GAP-FILLING METHODS

Here, we review five potential methods for gap-filling soil N_2O emissions, ranging from simple to more advanced (Table 2). This does not attempt to serve as an exhaustive list or overview of methods, especially with NN models that are just starting to be explored within soil N_2O research. Rather, our intent is to conduct a brief review of methods and provide details for use within soil N_2O research. For each method, we outline (a) what each technique is and the basic logic involved in it, (b) how each technique is conducted, (c) strengths of the technique, and (d) weaknesses of the technique, providing

TABLE 1 Steps within the gap-filling model creation process for estimating soil N_2O emissions

Gap-filling step	Description	Best practices
Covariate selection	• Detail covariates measured onsite: ○ Continuous data (soil water content and temperature, and climate data collected at or near the site) ○ Noncontinuous data (crop yield data, mineral N, etc.)	• Test covariates for significance to N_2O emissions • Include useful covariates within models • Gap-fill covariate data where needed; report the gap-filling method used on covariates
Data splitting	• Split data into training and testing datasets for model creation and validation	• Data splitting regime should resemble the way the model will be used: ○ Consistent day sampling is recommended when static chamber measurements will be examined ○ Common neural network (NN) data splitting is random, with 70% of data for training and 30% for testing, and can be used in more complete data sets where appropriate
Model creation	• Architecture or structure: describe in detail the model architecture and structure so that it can be replicated • Calibration: approach used to train the methods	• Report detailed model methodology and architecture: ○ e.g., model structure for NN, how many nodes or layers? ○ e.g., for generalized additive models (GAMs), how many variables were allowed and how were they chosen? ○ What algorithm is being used to determine the best model? • Select the model and parameters based on your test statistics
Model validation	• Testing the method against the testing dataset to assess and minimize model bias • Test against an unaffiliated (not used in training) dataset	• Report statistics from the testing dataset.

examples of the technique when possible. Within this section, we will refer to average daily N_2O emissions as *points*.

There are disclaimers that apply to all gap-filling methods worth noting in advance, although highlighting these challenges may be important for specific methods. Regardless of the complexity of the gap-filling method used, more robust field sampling campaigns are preferred because fewer gaps in data inherently reduces uncertainty in estimates (Barton et al., 2015). The use of a gap-filling method outside of its designed training data range is generally considered an incorrect application of a method and inherently introduces uncertainty. Further, as models become more complex, they are more prone to overfitting and memorizing training data, especially in smaller datasets. Interpretable models allow for clearly assessing the influence of covariates, and this influence is an important component of the model validation procedure. Therefore we believe interpretability should still be a goal, because these gap-filling methods are still in early developmental phases, and influence of covariates are still being determined. The use of subdaily N_2O emissions is not covered within this analysis, as integrating to daily emissions entails a review of its own and gap-filling research has primarily focused on daily emissions thus far.

2.1 | Simple interpolation methods

Linear interpolation is the most simplistic, and commonly used, of the gap-filling methods. In linear interpolation, known points are connected by drawing a straight line between the points to create a complete time series. Another variation of the simple interpolation methods is that of moving (time) windows. Moving windows use a running average of N_2O points, which can reflect a variable number of days. For example, a 3-d moving window could use the current day along with the previous 2 d for

TABLE 2 List of gap-filling methods used within N_2O research

Method	Method complexity	Examples of use in N_2O research
Linear interpolation	Simple	Standard practice
Generalized additive models (GAMs)	Moderate	Webb et al. (2019), De Rosa et al. (2020)
Autoregressive integrated moving average (ARIMA)	Moderate	De Rosa et al. (2016), De Rosa et al. (2018), Mumford, Rowlings, Scheer, De Rosa, & Grace (2019)
Random forest (RF)	Complex	Philibert et al. (2013)
Neural network (NN)	Complex	Taki et al. (2018), Bigaignon et al. (2020)
Biogeochemical models[a]	Complex	Del Grosso et al. (2005), Giltrap et al. (2010), Giltrap et al. (2020)

[a]We did not review biogeochemical models in this article, but their capabilities for estimating emissions is worth mentioning, as they have been used across a range of sites and scenarios and are a valuable tool for scenario testing.

estimating the N_2O emission for that day. There are further alternatives that use a 3-d window but weight the days so that the last day carries a larger weight within the window. Moving windows are a precursor to the more advanced ARIMA method, which uses lag periods and adjustment factors in concert with the moving window approach, and thus are not covered in depth here.

2.1.1 | Strengths

The obvious strength of these techniques is their simplicity. Interpolating between points can be easily accomplished in any data sheet or statistical program. If there are a minimal number of gaps or short gaps (likely a few days or less), the linear method may be an accurate and easy to implement gap-filling method, whereas more advanced methods may be overkill and take longer to implement than is needed given the available field data.

2.1.2 | Weaknesses and challenges

Linear interpolation estimates do not reflect current soil conditions, which can be reflected in other methods via use of covariates. As gaps become more frequent or longer, emissions drivers (like freeze–thaw, wetting–drying, and substrate change, fertilizer or excretal inputs) become more likely to be missed, leading to increased uncertainty in emissions. In an examination of sampling frequency, Barton et al. (2015) showed that studies that sampled >7 d apart resulted in lower accuracy than those taken <4 d apart. In a study looking at manure and urine deposits, sampling every 3 d during a short-term study was found to provide no bias in estimates (van der Weerden, Clough, & Styles, 2013).

Sampling campaigns that do not cover an entire year period can be especially prone to uncertainty, as drawing linear interpolations out to unknown points is problematic. For example, a site in the United States that only samples during the growing season of a maize (*Zea mays* L.) crop (Julian days 125–300) still needs to have the interpolation drawn out to a 365-d period (e.g., Julian days 1–365) to get an accurate annual estimate. With no points for large portions of the year, the linear interpolation is a crude method. There are several approaches that have been used for gap-filling these off-sample periods: extending a straight line from first or last point(s), extending the first or last to the ends, taking an average of first or last few points, or using preceding or proceeding years with which you can use a linear interpolation. Sampling protocol generally recommends that sampling is started before major management events that are known to lead to emissions peaks (e.g., tillage, organic matter amendments, or fertilizer), but that does not exclude the possibility that the first point may have been sampled at a peak (e.g., freeze–thaw event), which will lead to inaccurate estimates. Currently, an average of several days within the beginning or ending period drawn to the ends may be the most acceptable of these limited methods, unless it is known that peak emissions may occur outside the growing season and not in response to N fertilizer additions.

2.2 | Generalized additive models

As environmental data are often nonlinear, GAMs are useful models, as they allow nonlinear responses of the dependent variable to covariates. The method at its base is a linear model that allows for potentially several regressions of smoothed, linear, and nonlinear nature to be determined based on covariates and N_2O emissions (De Rosa et al., 2020; Webb et al., 2019; Wood, Pya, & Safken, 2016). The linear regressions are then estimated with spline weights that allow functions to become nonlinear. Penalty terms are introduced to keep weights near zero

and reduce the impact of splines, preventing overfitting within the functions (Wood et al., 2016). A smoothness parameter is then implemented during cross-validation to control the flexibility of the curve. Model behavior and influence of covariates can be understood through the examination of partial dependence plots, making GAMs more understandable than more advanced methods. Further, GAMs allow for calculation of point-wise confidence intervals that are derived from the standard error associated with the posterior distribution of the model coefficients.

2.2.1 | Strengths

The GAM model has self-determining ability that allows for estimation and creation of regressions of potentially difficult and complex relationships (Webb et al., 2019). This can be especially useful when a nonlinear response to a covariate is expected or known, but the form of the equation is unknown (Webb et al., 2019). On the other hand, contrary to other gap-filling procedures presented, the shape constrained generalized additive model (SCAM) framework (Pya & Wood, 2015) offers the possibility to force particular covariates to conform to certain shape restrictions. For instance, pasture growth and fertilizer response curves are known to be monotonic, and thus covariate response curves can be adjusted to this understanding. This allows more control and constraints to be placed on the model than some of the more complex models.

2.2.2 | Weaknesses and challenges

Overreliance on GAM self-determination can result in overly complex models that become less interpretable and provide unrealistic results outside of the testing scenario. Due to this computational complexity and self-determination, GAMs can be prone to overfitting, and a priori knowledge should be used to properly parametrize the model. This a priori knowledge should be used to help guide in determination of the covariates, number and types of smooths, and penalties associated with splines (Webb et al., 2019). When fitting a time series with GAMs, concurvity between regressors, the nonparametric analogue of multicollinearity, can reduce the sensitivity of the model prediction by underestimating the variance of the fitted model parameters. The Pearson product moment correlation coefficient should be calculated to determine covariates that are highly correlated, with covariate(s) removed to address multicollinearity issues.

2.3 | Autoregressive integrated moving average

The ARIMA method is common for time series forecasting problems but has also been tested within gap-filling scenarios (De Rosa et al., 2016, 2018). Within the ARIMA method, the autocorrelation (relation between elements of a series at separate time intervals) between sequential N_2O points is examined and used for prediction of future emissions (Box & Pierce, 1970). Covariates can be used within the model to influence the N_2O autocorrelation estimates (De Rosa et al., 2016), though ARIMA can also be run without covariates. These interactions in the relationship between an observation, and its preceding observations and covariates, allow for gap-filling or future prediction of N_2O emissions (De Rosa et al., 2018).

2.3.1 | Strengths

The ARIMA method is a model with a moderate degree of difficulty that is best used for predicting future time series data but is also used for gap-filling. Although ARIMA can be run without covariates, it is also possible to include covariates and seasonal effects to improve model predictions, especially in longer term datasets. A main reason to use the ARIMA model is its use of previous observations, which allows the model to incorporate preceding conditions to inform current values. Within N_2O research, this is informative as emissions may lag current conditions, or preceding info can better inform current emissions (e.g., recognition that fertilizer was applied a few days ago). The smoothing component of the model is useful in incorporating long-term trends and seasonality and works to eliminate residual autocorrelation during the model development process. Uncertainty can be calculated with the ARIMA model using bootstrap simulations (or the normal distribution) to provide point-wise confidence intervals.

2.3.2 | Weaknesses and challenges

The ARIMA models require at least 50 points, so that the model has access to enough data for determining the length of lag observations, as well as relationships with covariates. If a seasonality effect is present, the data requirement increases. Like other autoregressive models, if no covariates are included in the model, it cannot forecast out of the boundaries of the training dataset. This impacts the model's ability to predict when gaps in the time series occur during or after events that

could trigger N_2O emissions, such as fertilization or large rainfall events. Moreover, if the gap in the time series is quite substantial, the model prediction capability tends to decrease. The ARIMA model is less accurate than other models because sharp changes in the dependent variable (N_2O) time series are not modeled well.

2.4 | Random forest

Random forest (RF) machine learning can be a powerful tool within classification-based problems. The basis of RF models is the decision tree, where parameters are sorted into different classes or groups (which can also be referred to as a nearest neighbor classification) and used within a relationship tree. Random forest models use many decision trees, in an ensemble approach, thus using many decision trees in the establishment of a forest. This ensemble approach allows for consideration of many parameters or relationships at once, which allows strong relationships to be highlighted while maintaining weak relationships by tying them to stronger relationships (Philibert, Loyce, & Makowski, 2013). The use of many decision trees within the forest ensures that the error from one decision tree cannot overly bias the performance of the model.

Covariate selection can be done using the importance test to determine which covariates to include within the model. In model development, trees are initially constructed on a subset of data (the default is generally one-third of available input data), where binary decisions trees are developed (Philibert et al., 2013). Random selection of covariates is determined by adjusting the "mtry" variable, or the number of covariates allowed in a tree, where it is generally set to a third of the total variables available (Breiman, Freidman, Olshen, & Stone, 1984; Liaw & Wiener, 2002). This is followed by a recursive selection process that works to create smaller selection nodes that minimize the misclassification rate, with the recursive selection process completed "ntree" number of times.

2.4.1 | Strengths

Relationships between covariates and the dependent variable are not required and are assessed within development of trees, allowing the model to be approached without knowing covariates relationships (Philibert et al., 2013). Random forest machine learning is also good at dealing with noisy data, as well as being able to handle missing input data points (by replacing missing data with the mean or mode value of similar proximity data; Philibert et al., 2013). The classification aspect of RF models is likely a strong fit for N_2O research given the episodicity of emissions linked to common conditions, and correctly classifying these peaks is important for estimating cumulative emissions (Bornø, Rønn, & Ekelund, 2020). This approach is a more advanced statistical method but has the benefits of being much less computationally intense and quicker than NN models, while also being more interpretable.

2.4.2 | Weaknesses and challenges

The main disadvantage of both RF and NN approaches is that they can be difficult to understand, and thus structural validity can be difficult to ascertain. Because the mtry parameter restricts the number of covariates within a tree, the RF model is best suited to datasets with a lot of covariate variables. Where covariates are limited, RF may be likely to overfit to the available data.

2.5 | Artificial neural networks

Neural networks are a fundamental building block of modern machine learning and data science and have become one of the strongest approaches for data-driven predictive modeling in many fields. Initial results using NN within N_2O flux data have been successful, with NN providing higher R^2 values than linear interpolations when used for gap-filling agricultural datasets (Bigaignon et al., 2020; Taki et al., 2018). A NN tries to predict an output variable (N_2O flux in this case) by determining the interconnected relationships between the covariate data (e.g., temperature, moisture, inorganic N, etc.) and that of the output (Rasmussen & Williams, 2006). An example of this model structure is seen in Figure 1 , which shows the network architecture of a basic multilayer perceptron (MLP) model. Further methods for building a NN can be found in the supplemental material, as well as in Bigaignon et al. (2020).

More complicated NN architectures include (a) long-short term memory (LSTM) and other recurrent NNs (RNNs), which excel at finding time-dependent relationships and using internal memory to recognize recurring patterns, and (b) convolutional NNs (CNNs), which have yet to be reviewed for their ability to estimate N_2O emissions and warrant further study (Schmidhuber, 2015). Within N_2O emission research where we expect fluxes at certain conditions (high moisture, after N application, etc.), these methods could potentially capture common flux patterns or be more useful in developing a universal model.

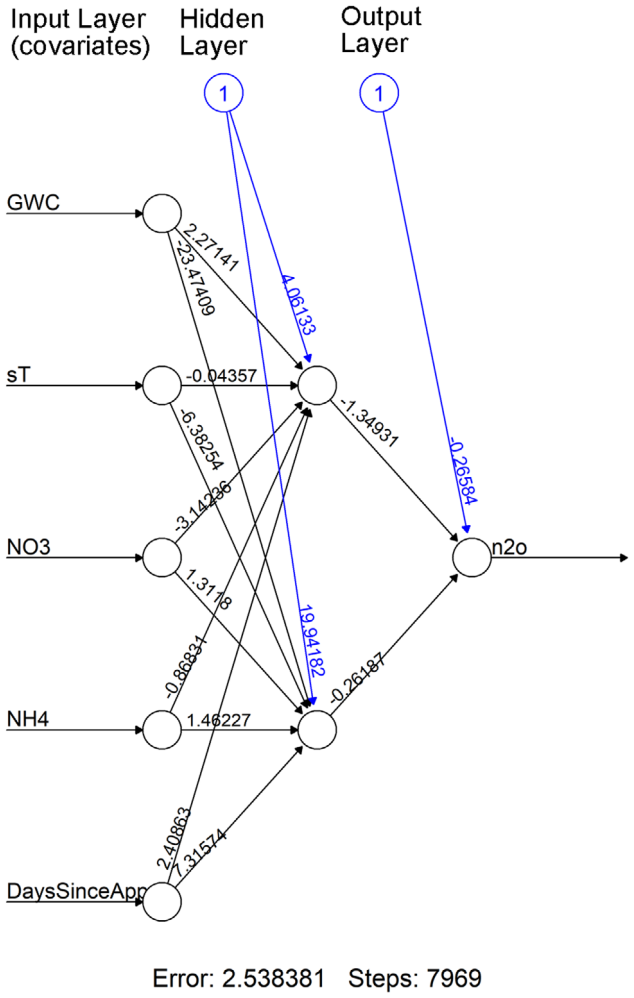

FIGURE 1 Example structure of a feed-forward multilayer perceptron (MLP) neural network (NN) for an N_2O site, where input layers are GWC (water-filled pore space), sT (soil temperature), NO3 (soil inorganic nitrate), NH4 (soil inorganic ammonium), and DaysSinceApp (a numerical count variable for the elapsed days since the last N application event). The values in the figure are the weights used within the NN. As NN algorithms are iterative, "Steps" indicates the number of parameter changes that were done in order to minimize the error within the NN model

2.5.1 | Strengths

Neural networks offer a highly general, flexible, and efficient approach to modeling unknown relationships between input and output variables given sufficient training data. They can achieve high accuracy with little user intervention beyond choosing a network architecture and training parameters like learning rate and batch size when given enough data. No relationship between variables must be specified by the user ahead of time, and networks can easily be expanded (including more layers, hidden neurons, etc.) to achieve higher accuracy.

2.5.2 | Weaknesses and challenges

The biggest challenge with NN is their dependence on large training datasets, likely requiring the most data of any method, which can be limiting within N_2O research. Because NNs assume no relationship (or a random relationship) between variables, all relationships they learn are only as good as the dataset they learn from. Multilayer perceptrons, like that described here, are also fully connected networks where one layer connects to the next. These sequentially connected layers are prone to overfitting of training data.

Beyond the heavy data requirements of training a NN, it is worth emphasizing some challenges of using them due to their *black box* nature. It is difficult to interpret the importance of each input in predicting the output (although we do have tools like Garson's algorithm which can help here; Garson, 1991), as well as to interpret the model itself. Because of difficulties in interpretability, network outputs should not be assumed to satisfy any physical constraints or relationships that actually dictate N_2O fluxes, and thus a structural validity check may be difficult to perform. Further, biases and errors in the training set become ingrained in the learned behavior of the network and can be manifest in two ways. First, if the training set is small, the network may have the capacity to simply memorize the data (this is called overfitting), rather than learning underlying relationships. Second, if the training data poorly represent the behavior we wish to model, the network will learn trends that are not useful to the problem of interest. Due to these limitations in interpretability, generalizability of the model for use on a dataset other than that on which it was trained should not be assumed, and limitations in the training data need to be considered.

3 | GAP-FILLING PROTOCOL

Utilization of the methods outlined above will be most successful and scientifically sound when implemented within a standardized gap-filling protocol. Here, we present a protocol that details the various steps (Table 1) that should be undertaken within a gap-filling exercise. The main goal of this gap-filling protocol is to create a systematic and consistent approach that allows for comparison across studies of soil N_2O emission. Further, establishment of a protocol will help researchers use statistically based methods for model selection and testing, rather than relying on past experience or trial-and-error methods that are too commonly used when no protocol is present. The development of this protocol also supports the recent push for FAIR (Findable, Accessible, Interoperable, and Reusable) data, analysis scripts, and model code that we believe can help

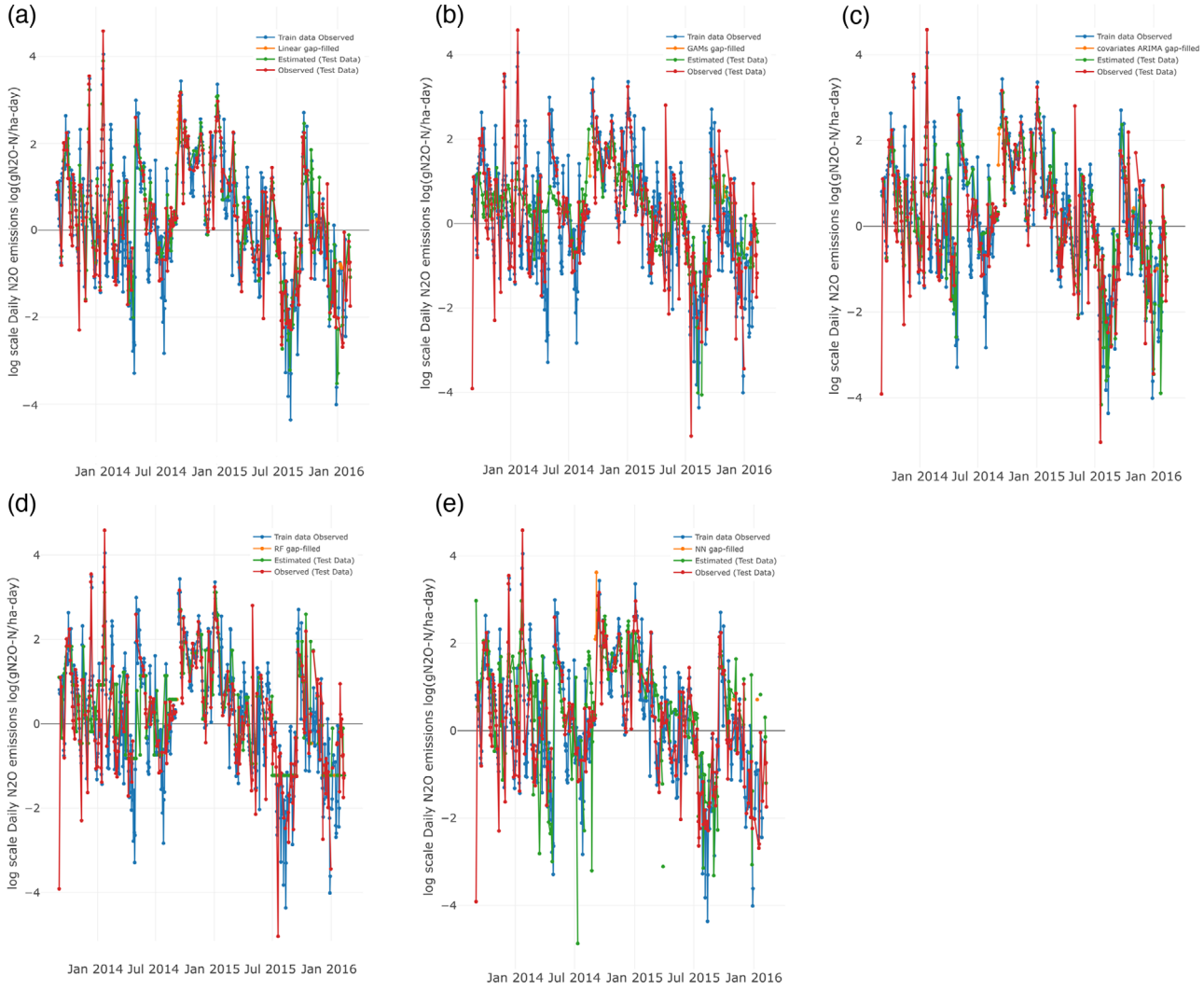

FIGURE 2 Nitrous oxide emissions across the time series from methods (a) linear, (b) generalized additive models (GAMs), (c) autoregressive integrated moving average (ARIMA), (d) random forest (RF), and (e) neural networks (NNs). Training and testing points were the same across models, with blue points and lines for the training data, red for the observed testing data (what models were trying to estimate), green for the model estimates of the testing points, and orange for the estimates for the actual gaps in the data

accelerate understanding within N_2O research by allowing more data to be examined and increase scenario testing (Wilkinson et al., 2016). Our analyses were conducted within the R statistical program (RStudio Team, 2015); however, there are many platforms that can conduct these analyses.

There are a number of limitations and methodology questions related to field sampling that need to be addressed and can potentially affect gap-filling efforts. As manual chamber measurements are a common sampling practice, this article focuses on daily mean N_2O emissions and associated covariate values. However, the *Journal of Environmental Quality* special section in which this article is published attempts to cover the entirety of methodology relevant to sampling soil N_2O with chamber systems. Future research is needed for a comprehensive treatment of gap-filling that can account for issues related to diurnality in fluxes (Grace et al., 2020) and the choice of flux calculation method (Venterea et al., 2020). The methods described in this paper have also been used for gap-filling non-chamber-based measurements (Taki et al., 2018).

Because covariates may not be continuously measured (e.g., soil inorganic N is generally sampled infrequently), a gap-filling approach is needed for these covariates as well. For our analysis, linear interpolation for covariates was used. This interpolation of covariates is inherently limiting by itself, but exploring these limits is beyond the scope of this work and requires further research. While these advanced methods are studied and developed, one approach that is of further interest is the use of biogeochemical process-based models for use in gap-filling of

covariates, as they have known mechanisms for estimating most of the covariates of interest (Giltrap et al., 2020).

The gap-filling protocol (as summarized in Table 1) consists of the following steps: covariate selection, data splitting, model creation, and model validation. In each section, we define terms, set standards for components of the protocol, and demonstrate the application of those standards. The application of the standards will apply the various steps of the protocol to an example N_2O flux measurement site and are found within boxes after each section. We have selected a site with near-continuous data (missing only 16 d out of a 892-d period), as this allows us to create artificial gaps within the dataset. With the approach of artificial gap creation, one is able to estimate how well the models estimate daily fluxes (e.g., by comparing the measured values removed from the dataset with the predicted values). As the goal of this paper is to highlight potential gap-filling methods, showing their ability to predict known values is a major goal. For statistical models like GAMs and ARIMA, daily uncertainty can be provided within the methods. For NN and RF models, bootstrapping can create a large number of time series with artificial gaps and can be used to provide uncertainty on emissions data.

3.1 | Covariate selection

Field-measured non-N_2O variables or covariates (e.g., soil temperature and moisture, inorganic N, climate parameters, crop parameters) can be used for informing the various methods. Creating event variables are useful in reporting management covariates (e.g., tillage events, day of fertilization) that have known impacts on emissions but are not measurable in a traditional sense like other covariates. The use of event variables is common practice for GAMs and ARIMA (De Rosa et al., 2016, 2018) up through RF and NN models.

Any data management that changes or interpolates field-measured covariates (e.g., gap-filling used within covariate measurements) needs to be reported, as this will affect results. Not all covariates are useful in estimating N_2O emissions and therefore an input selection process must be used to determine significant relationships between covariates and N_2O emissions. Input parameters should be tested for relationships and significance to N_2O emissions (using correlation matrices, regressions testing p values, importance test, stepwise methods), and any collinearity of covariates should be dealt with, likely by removing one of the covariates. Stepwise covariate selection was implemented for all methods, using a simple remove-one variable stepwise selection for GAMs and ARIMA, while using the importance test within RF and NN models.

Description of example site (Gatton)

In order to demonstrate use of this protocol, we tested these methods on an automated chamber dataset of 892 d from a vegetable N_2O trial at the Gatton Research Station in the Lockyer Valley, Queensland, Australia. The study was sampled from 3 Sept. 2013 until 11 Feb. 2016 (De Rosa et al., 2016, 2018). During this time period, there were only three gaps in sampled N_2O data, lasting 8, 3, and 5 d, respectively. This site was chosen for its completeness, which allowed us to make artificial gaps in the dataset so that we could test the accuracy of the models knowing the values removed when creating these artificial gaps. Nitrogen applications at the site consisted of raw chicken manure (Ma), composted chicken manure (Co), and two levels of inorganic fertilizer applications (Conv, the full application rate, and Red, the reduced application rate). The four treatments that were examined were Ma + Conv (raw chicken manure at 102 kg N ha^{-1} with mineral N at 310 kg N ha^{-1}), Ma + Red Conv (raw chicken manure at 102 kg N ha^{-1} with mineral N at 253 kg N ha^{-1}), Co + Conv (composted chicken manure at 240 kg N ha^{-1} with mineral N at 310 kg N ha^{-1}), and Co + Red (composted chicken manure at 240 kg N ha^{-1} with mineral N at 250 kg N ha^{-1}). More details on this study can be found in De Rosa et al. (2016, 2018), with the full methods examination report found in the supplemental material.

3.2 | Data splitting

Data splitting practices have become more common and better informed due to NN research and are needed for validating the accuracy of a gap-filling model. We recommend the standard practices from NN methods that focus on minimizing bias and variance, allowing for statistical analysis of methods, and improved generalizability of models (Wu, May, Dandy, & Maier, 2012). Data splitting methods suggests training data are 70% of the available information, with the remaining 30% used for testing (Wu et al., 2014). Although data splitting is generally done in a random format, we also advise another sampling scenario that does not split randomly but rather reflects the strategy of manual chamber sampling campaigns. This "consistent day" sampling method (Barton et al., 2015), uses the same days within a weekly dataset (e.g., Monday, Wednesday, Friday) for training and then uses remaining

Covariate selection at Gatton

Covariate analysis selected soil temperature, water-filled pore space, nitrate (NO_3), and ammonium (NH_4) as the covariates with the greatest relationship to N_2O. The covariate with the strongest relationship (CO_2 flux) had a correlation value of 0.37 but was not included in our analysis, as CO_2 flux is commonly sampled at the same time within a chamber sample as N_2O and thus data are often not available during days where N_2O is, limiting its use as a covariate for gap-filling those missing days. Mineral N concentrations were measured 48 times over the duration of the study, and thus a linear interpolation was used to gap-fill these values. The other parameters had small gaps (missing <60-d over the 874 d), and linear interpolation was used on these parameters when needed. As there was no collinearity detected and no strong correlations, no parameters were directly removed prior to model creation, allowing models to potentially use all covariates, removing them as suggested by the stepwise methods used. A covariate for days since fertilizer was applied (DaysSinceApp) was added, as this was found to be a useful variable for reflecting management practices.

data for testing. Although this sampling format risks poor model performance, due to limited data available for model training, it warrants examination as manual chamber studies are most in need of gap-filling methods. This consistent day approach allows for development of models in the manner in which they will be implemented for manual chamber studies. Concerns with limited data in this approach can be alleviated by using additional datasets or longer term studies to help in model development (e.g., using a similar site with near-continuous data). Additional considerations, like ensuring an adequate split in training–testing data points, from important emission periods (freeze–thaw in colder climates, Wagner-Riddle et al., 2017; precipitation in arid regions, Barton et al., 2016) should be examined within datasets. Similarly, when using multiple sites, the use of a site that experiences freeze–thaw emissions with a non-freeze–thaw site can lead to false peaks in a dataset, and testing should be done to determine if datasets can be examined appropriately together.

Data splitting at Gatton

We tested the models using six scenarios: consistent day sampling, as well as five variations of the random sampling (in increments of 10%, decreasing the percentage of data used in training data from 70 to 30%). The random sampling was examined across these training and testing data splits to examine the ability of models to use progressively less data in order to determine minimum data requirements, and better replicate issues that arise with chamber sampling. Splitting the data based on the preferred random sampling (70% training) resulted in ~260 sample d per year being used for training the model, whereas consistent day uses ~156 d for training, and the least training random sampling (30% training) used ~109 d for training. This level of training data (and our fake sampling) gives a wide variation and thus an ability to analyze the capabilities of the models against possible chamber sampling scenarios.

3.3 | Model creation

Once data splitting and covariates have been examined, the next step is creation of the model. This should start with a justification of why the selected method(s) were tested and used within the gap-filling. Why the method(s) were chosen, or if other methods were tested and not presented in results, should be explained, as this can help inform situations under which a method can be used. The model creation process consists of three phases: architecture selection, model structure selection, and model calibration (Wu et al., 2014). The models should be developed consistently (i.e., using the same data splitting for a NN and linear model) and documented with sufficient detail so that the model can be replicated and compared for future studies. This should go beyond reporting the final model details, with details also reported describing the input parameter selection process, what metrics were used in the selection process, and what tunable parameters were explored (e.g., hidden layers or weights within a NN). Model architecture and validation results should be reported (and preferably models and code should be made available) so that the method can be replicated by others (Wu et al., 2014).

Architecture selection details the covariates being used within the dataset and the flow of data within the model (e.g., do covariates interact or is it a feed-forward structure

in a NN). Parameters documented may be model dependent. For an ARIMA model, this will be detailing the number of lag observations used in the model (p; in our case, how many *preceding* days to use), number of times that raw observations are differenced (d; degree of differencing) and the size of the moving average window (q) variables within the model. For GAMs, it will be detailing the smooths used for covariates. For RF and NN, this will be detailing the forest and node structures and how these were determined.

The optimal model structure is determined by minimizing forecasting or prediction errors. The most common ways to achieve this are the use of backpropogation methods, global optimization methods (e.g., genetic algorithms), or stepwise models (e.g., pruning; Wu et al., 2014). Normalization is another feature that is typically needed within NN and RF approaches to prevent overfitting of models.

Model calibration, or model training, is often done concurrently with the model structure selection phase. Using the methods determined in the model structure selection the model will iterate through parameter options until it has reached a best fit based on the structural selection method.

3.4 | Model validation

Model validation, or proof of the accuracy of the model, needs to be provided. The protocol splits validation into three categories: replicative, predictive, and structural.

Replicative validity uses standard statistical methods to assesses the accuracy of the model on the testing dataset, ensuring confidence in the method. Common statistics for testing model performance include coefficient of determination (R^2), root mean squared error (RMSE), relative root mean squared error (RRMSE), mean bias error (MBE), and mean absolute percentage error (MAPE). This list is not exhaustive, and other statistics can be used with enough information provided.

Predictive validity is determined by testing the model (using the same statistics as replicative validity) with an independent dataset that was not used within model development. Using an affiliated but independent dataset tests the capabilities of the model and can be used to limit overfitting.

Structural validity is a sensitivity analysis that reflects a priori knowledge (Wu et al., 2014). This validity check is not steeped in statistical tests but rather ensures that the developed model is a realistic representation of current scientific understanding and natural processes. This presents an opportunity to check if your model is suggesting something that is unrealistic within N_2O research or is not physically possible, and thus correct it.

Model creation at Gatton

Stepwise selection was used within all models for determining which covariates should be included in the models. Based on this process, the ARIMA model used DaysSinceApp, soil moisture, and nitrate within the model, whereas the other methods used all available covariates.

For the GAMs model, a smooth term was applied to all of the parameters based on an enhanced stepwise selection looking at different curves for the covariates. More complex models with interactions and other smooths were not used here for interpretability reasons, though they could potentially lead to better results.

The ARIMA model was chosen based on maximum likelihood estimation (MLE) with values of 3, 0, and 0 for p, d, and q, where 0 values for parameters mean the component is not used in the model. Thus, only p, the number of lag observations (3 d), was used in the ARIMA model for which we show results.

Data were normalized for both random forest and neural network models. The random forest model used the ANOVA method in determination of the model, and stepwise selection to determine the ntree, mtry, and cp parameters. Ten-fold cross-validation (the default in randomForest package in R) was used within calibration of the model.

For the MLP NN, the model calibration was done with the sum of squared errors (SSE) approach for determination of weights in the NN with stepwise selection exploring different node sizes. The importance of covariates was tested with Garson's algorithm (Garson, 1991).

4 | CONCLUSION

Use of advanced gap-filling techniques is still new within soil N_2O emissions data but has shown promise in improving estimates. The choice of gap-filling method will likely depend on the number and duration of gaps, as well as the available covariate data. Although significant questions remain about method selection, we encourage the N_2O community to follow the gap-filling protocol outlined and to examine and report on multiple methods, even reporting negative results, as this will allow for comparability of studies and aid in determining which scenarios are best suited for a gap-filling model.

Model validation at Gatton

Model validation was done in the three respective steps. Models were checked for structural validity, examining if the model equations made sense for the N_2O emissions data. As there were only five covariates available (soil temperature, water-filled pore space, NH_4, NO_3, and days since fertilizer application) and all have known interactions with N_2O, the models selection of covariates seem reasonable. Part of this structural validity check was also minimizing complexity in models, sometimes to the detriment of model results. For example, the random forest model could potentially have used a lot more trees within the model, getting better results, but would have become extremely difficult to understand and likely would overfit the data. Similarly, interactions in GAMs or ARIMA could potentially improve results but would make comprehension of the model and reuse of it more difficult. Use of the other validation steps in concert can help in the decision making process for this. Replicative validity is where model metrics were generated for each of the models and sampling scenarios, showing statistics for both training and testing datasets, respectively, as well as the cumulative results. Replicative validity results are shown in Figure 2 where the time series is plotted with model results, as well as reporting the statistics of how models performed. The validation statistics (Table 3) are broken down to show both the results for the training (613 points) and testing (263 points) data splits, as well as the cumulative emissions data when gap-filled using respective methods. The sum of N_2O value therefore uses the 613 observed (training) data points and the 263 estimated testing days from the respective gap-filling methods in order to calculate an emission sum over the duration of the study. These are provided as an example and are not meant to propose the use of a particular model among the models proposed, as more work needs to be done testing across scenarios and within gap-filling scenarios. Predictive capability of the models would then be tested by running the model on another treatment from the De Rosa et al. (2018) study (i.e., running the model created from the Co+CONV treatment on the Ma+CONV treatment). The difference between these two treatments is the use of composted chicken manure and raw manure (at N application rates of 240 and 102 kg N ha^{-1}, respectively). Differences in the type of organic matter applied could be reflected in different inorganic N values between the treatments, though this may be hard to differentiate. Further, other impacts (water-holding capacity, infiltration, etc.) from the different organic matter additions are not reflected explicitly in the covariates that are used in the models but likely have an impact on emission rates (Barton & Schipper, 2001; van der Weerden et al., 2016). These are provided as examples of implementations of the methods, but further development of the methods across sites is warranted.

TABLE 3 Validation statistics for the Co+CONV (composted chicken manure at 240 kg N ha^{-1} with mineral N at 310 kg N ha^{-1}) treatment in Gatton. This used the standard 70:30 data splitting for training and testing (613 and 263 d, respectively, for the two datasets)

Statistic	Observed	Linear	GAMs	ARIMA	RF	NNs
Cumulative statistics						
Sum of N_2O, kg N_2O-N ha^{-1}	2.566	2.647	1.639	2.527	2.465	2.495
Test statistics						
No. of testing days	263					
Sum of N_2O, kg N_2O-N ha^{-1}	0.791	0.864	0.462	0.708	0.69	0.720
RMSE, g N_2O-N ha^{-1}		1.584	1.604	4.513	5.864	6.369
R^2		.672	.478	.718	.379	.240
MBE		−0.013	1.086	−0.181	−0.381	−0.267
Training statistics						
No. of training days	613					
Sum of N_2O, kg N_2O-N ha^{-1}	1.775	NA[a]	1.177	1.819	1.775	1.775
RMSE, g N_2O-N ha^{-1}		NA	1.669	2.776	3.052	3.374
R^2		NA	.374	.686	.614	.531
MBE		NA	1.186	0.095	0.000	0.003

Note. GAM, generalized additive model; ARIMA, autoregressive integrated moving average; RF, random forest; NN, neural network; MBE, mean bias error.
[a]NA, not applicable.

Although we have outlined a protocol for gap-filling, there are still many significant questions that remain. Gap-filling has primarily been done on daily averages; whether subdaily data for gap-filling are feasible, or would improve results, is uncertain. Covariate data can often contain gaps, and gap-filling of those data can present problems, as the N_2O estimate is now reliant on uncertainty beyond that of spatial variation and N_2O measurement uncertainty. Use of process-based biogeochemical models for gap-filling covariates may be an acceptable method in the near term and should be reviewed. Lastly, issues with methods around chamber sampling itself (integration used, filtering of minimum R^2, spatial and temporal coverage, and minimum detection limits) need to be examined within gap-filling techniques, which also requires reporting of full methods within publications.

ACKNOWLEDGMENTS

We are thankful for support for the publishing costs from the New Zealand Government, in support of the objectives of the Livestock Research Group of the Global Research Alliance on Agricultural Greenhouse Gases. Funding from the USDA Office of the Chief Economist through Cooperative Agreement no. 58-0111-19-019 and USDA AFRI Award 2018-67019-27812 was used for this research.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Christopher D. Dorich https://orcid.org/0000-0001-8005-5797
Daniele De Rosa https://orcid.org/0000-0002-0441-7722
Peter Grace https://orcid.org/0000-0003-4136-4129
David Rowlings https://orcid.org/0000-0002-1618-9309
Claudia Wagner-Riddle https://orcid.org/0000-0002-4802-6088

REFERENCES

Abdalla, M., Jones, M., Smith, P., & Williams, M. (2009). Nitrous oxide fluxes and denitrification sensitivity to temperature in Irish pasture soils. *Soil Use and Management*, *25*, 376–388. https://doi.org/10.1111/j.1475-2743.2009.00237.x

Albanito, F., Lebender, U., Cornulier, T., Sapkota, T. B., Brentrup, F., Stirling, C., & Hillier, J. (2017). Direct nitrous oxide emissions from tropical and sub-tropical agricultural systems: A review and modelling of emission factors. *Scientific Reports*, *7*. https://doi.org/10.1038/srep44235

Barba, J., Cueva, A., Bahn, M., Barron-Gafford, G. A., Bond-Lamberty, B., Hanson, P.J., … Vargas, R. (2018). Comparing ecosystem and soil respiration: Review and key challenges of tower-based and soil measurements. *Agricultural and Forest Meteorology*, *249*, 434–443. https://doi.org/10.1016/j.agrformet.2017.10.028

Barton, L., Hoyle, F. C., Stefanova, K. T., & Murphy, D. V. (2016). Incorporating organic matter alters soil greenhouse gas emissions and increases grain yield in a semi-arid climate. *Agriculture, Ecosystems & Environment*, *231*, 320–330. https://doi.org/10.1016/j.agee.2016.07.004

Barton, L., & Schipper, L. A. (2001). Regulation of nitrous oxide emissions from soils irrigated with dairy farm effluent. *Journal of Environmental Quality*, *30*, 1881–1887. https://doi.org/10.2134/jeq2001.1881

Barton, L., Wolf, B., Rowlings, D., Scheer, C., Kiese, R., Grace, P., … Butterbach-Bahl, K. (2015). Sampling frequency affects estimates of annual nitrous oxide fluxes. *Scientific Reports*, *5*. https://doi.org/10.1038/srep15912

Bigaignon, L., Fieuzal, R., Delon, C., & Tallec, T. (2020). Combination of two methodologies, artificial neural network and linear interpolation, to gap-fill daily nitrous oxide flux measurements. *Agricultural and Forest Meteorology*, *291*. https://doi.org/10.1016/j.agrformet.2020.108037

Bornø, M. L., Rønn, R., & Ekelund, F. (2020). Is wood ash amendment a suitable mitigation strategy for N_2O emissions from soil? *Science of the Total Environment*, *713*. https://doi.org/10.1016/j.scitotenv.2020.136581

Box, G. E. P., & Pierce, D.A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of American Statistical Association*, *65*, 1509–1526. https://doi.org/10.1080/01621459.1970.10481180

Breiman, L., Freidman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth.

Cayuela, M. L., Aguilera, E., Sanz-Cobena, A., Adams, D. C., Abalos, D., Barton, L., … Lassaletta, L. (2017). Direct nitrous oxide emissions in Mediterranean climate cropping systems: Emission factors based on a meta-analysis of available measurement data. *Agriculture, Ecosystems & Environment*, *238*, 25–35. https://doi.org/10.1016/j.agee.2016.10.006

Chadwick, D., Sommer, S., Thorman, R., Fangueiro, D., Cardenas, L., Amon, B., & Misselbrook, T. (2011). Manure management: Implications for greenhouse gas emissions. *Animal Feed Science and Technology*, *166–167*, 514–531. https://doi.org/10.1016/j.anifeedsci.2011.04.036

Charles, A., Rochette, P., Whalen, J. K., Angers, D. A., Chantigny, M. H., & Bertrand, N. (2017). Global nitrous oxide emission factors from agricultural soils after addition of organic amendments: A meta-analysis. *Agriculture, Ecosystems & Environment*, *236*, 88–98. https://doi.org/10.1016/j.agee.2016.11.021

Congreves, K. A., Wagner-Riddle, C., Si, B. C., & Clough, T. J. (2018). Nitrous oxide emissions and biogeochemical responses to soil freezing-thawing and drying-wetting. *Soil Biology & Biochemistry*, *117*, 5–15. https://doi.org/10.1016/j.soilbio.2017.10.040

Cowan, N., Carnell, E., Skiba, U., Dragosits, U., Drewer, J., & Levy, P. (2020). Nitrous oxide emission factors of mineral fertilisers in the UK and Ireland: A Bayesian analysis of 20 years of experimental data. *Environment International*, *135*. https://doi.org/10.1016/j.envint.2019.105366

Cowan, N., Levy, P., Drewer, J., Carswell, A., Shaw, R., Simmons, I., … Skiba, U. (2019). Application of Bayesian statistics to estimate nitrous oxide emission factors of three nitrogen fertilisers on UK grasslands. *Environment International*, *128*, 362–370. https://doi.org/10.1016/j.envint.2019.04.054

Davidson, E. A., Keller, M., Erickson, H. E., Verchot, L. V., & Veldkamp, E. (2000). Testing a conceptual model of soil emissions of nitrous and nitric oxides. *Bioscience*, *50*, 667–680. https://doi.org/10.1641/0006-3568(2000)050%5b0667:TACMOS%5d2.0.CO;2

Davis, B.W., Needelman, B.A., Cavigelli, M.A., Yarwood, S.A., Maul, J.E., Bagley, G.A., & Mirsky, S.B. (2017). A novel approach to estimating nitrous oxide emissions during wetting events from single-timepoint flux measurements. *Journal of Environmental Quality*, *46*, 247–254. https://doi.org/10.2134/jeq2016.02.0048

de Klein, C. A.M., Alfaro, M.A., Giltrap, D., Topp, C. F.E., Simon, P.L., Noble, A., & van der Weerden, T.J. (2020). Global Research Alliance N_2O chamber methodology guidelines: Statistical considerations, emission factor calculation, and data reporting. *Journal of Environmental Quality*. https://doi.org/10.1002/jeq2.20127

De Rosa, D., Rowlings, D.R., Biala, J., Scheer, C., Basso, B., & Grace, P.R. (2018). N_2O and CO_2 emissions following repeated applications of organic and mineral N fertilizer from a vegetable crop rotation. *Science of the Total Environment*, *637–638*, 813–824. https://doi.org/10.1016/j.scitotenv.2018.05.046

De Rosa, D., Rowlings, D.R., Biala, J., Scheer, C., Basso, B., McGree, J., & Grace, P.R. (2016). Effect of organic and mineral N fertilizers on N_2O emissions from an intensive vegetable rotation. *Biology and Fertility of Soils*, *52*, 895–908. https://doi.org/10.1007/s00374-016-1117-5

De Rosa, D., Rowlings, D.W., Fulkerson, B., Scheer, C., Friedl, J., Labadz, M., & Grace, P. R. (2020). Field-scale management and environmental drivers of N_2O emissions from pasture-based dairy systems. *Nutrient Cycling in Agroecosystems*, *117*, 299–315. https://doi.org/10.1007/s10705-020-10069-7

Del Grosso, S.J., Mosier, A.R., Parton, W.J., & Ojima, D.S. (2005). DAYCENT model analysis of past and contemporary soil N_2O and net greenhouse gas flux for major crops in the USA. *Soil and Tillage Research*, *83*, 9–24. https://doi.org/10.1016/j.still.2005.02.007

Dengel, S., Zona, D., Sachs, T., Aurela, M., Jammet, M., Parmentier, F. J. W., … Vesala, T. (2013). Testing the applicability of neural networks as a gap-filling method using CH_4 flux data from high latitude wetlands. *Biogeosciences*, *10*, 8185–8200. https://doi.org/10.5194/bg-10-8185-2013

Dorich, C. D., Conant, R. T., Albanito, F., Butterbach-Bahl, K., Grace, P., Scheer, C., … van der Weerden, T. J. (2020). Improving N_2O emission estimates with the global N_2O Database. *Current Opinion in Environmental Sustainability*, *47*, 13–20. https://doi.org/10.1016/j.cosust.2020.04.006

Flesch, T. K., Baron, V. S., Wilson, J. D., Basarab, J. A., Desjardins, R. L., Worth, D., & Lemke, R. L. (2018). Micrometeorological measurements reveal large nitrous oxide losses during spring thaw in Alberta. *Atmosphere*, *9*(4). https://doi.org/10.3390/atmos9040128

Flessa, H., Dorsch, P., & Beese, F. (1995). Seasonal variation of N_2O and CH_4 fluxes in differently managed arable soils in southern Germany. *Journal of Geophysical Research*, *100*, 23115–23124. https://doi.org/10.1029/95JD02270

Galloway, J. N., Aber, J. D., Erisman, J. W., Seitzinger, S. P., Howarth, R. W., Cowling, E. B., & Cosby, J. B. (2003). The nitrogen cascade. *Bioscience*, *53*, 341–356. https://doi.org/10.1641/0006-3568(2003)053%5b0341:TNC%5d2.0.CO;2

Garson, G. D. (1991). Interpreting neural network connection weights. *Artificial Intelligence Expert*, *6*, 46–51.

Giltrap, D., Yeluripati, J., Smith, P., Fitton, N., Smith, W., Grant, B., … Snow, V. (2020). Global Research Alliance N_2O chamber methodology guidelines: Summary of modelling approaches. *Journal of Environmental Quality*. https://doi.org/10.1002/jeq2.20119

Giltrap, D.L., Li, C., & Saggar, S. (2010). DNDC: A process-based model of greenhouse gas fluxes from agricultural soils. *Agriculture Ecosystems & Environment*, *136*, 292–300. https://doi.org/10.1016/j.agee.2009.06.014

Grace, P.R., van der Weerden, T.J., Rowlings, D.W., Scheer, C., Brunk, C., Kiese, R., … Skiba, U.M. (2020). Global Research Alliance N_2O chamber methodology guidelines: Considerations for automated flux measurement. *Journal of Environmental Quality*. https://doi.org/10.1002/jeq2.20124

Hoben, J. P., Gehl, R. J., Millar, N., Grace, P. R., & Robertson, G. P. (2011). Nonlinear nitrous oxide (N_2O) response to N fertilizer in on-farm corn crops of the US Midwest. *Global Change Biology*, *17*, 1140–1152. https://doi.org/10.1111/j.1365-2486.2010.02349.x

Hoffmann, M., Jurisch, B., Borraz, E.A., Hagemann, U., Drosler, M., Sommer, M., & Augustin, J. (2015). Automated modeling of ecosystem CO_2 fluxes based on periodic closed chamber measurements: A standardized conceptual and practical approach. *Agricultural and Forest Meteorology*, *200*, 30–45. https://doi.org/10.1016/j.agrformet.2014.09.005

IPCC. (2006). Emissions from livestock and manure management. In H.S. Eggleston, L. Buendia, K. Miwa, T. Ngara, & K. Tanabe (Eds.), *IPCC guidelines for national greenhouse gas inventories, Vol 4: Agriculture, forestry and other land use*. Kanagawa, Japan: Institute for Global Environmental Strategies.

Kim, D.G., Hernandez-Ramirez, G., & Giltrap, D. (2013). Linear and nonlinear dependency of direct nitrous oxide emissions on fertilizer nitrogen input: A meta-analysis. *Agriculture, Ecosystems & Environment*, *168*, 53–65.

Kim, Y., Johnson, M.S., Knox, S.H., Black, T.A., Dalmagro, H.J., Kang, M., … Baldocchi, D. (2019). Gap-filling approaches for eddy covariance methane fluxes: A comparison of three machine learning algorithms and a traditional method with principal component analysis. *Global Change Biology*, *26*, 1499–1518. https://doi.org/10.1111/gcb.14845

Kravchenko, A. N., & Robertson, G. P. (2015). Statistical challenges in Analyses of chamber-based soil CO_2 and N_2O emissions data. *Soil Science Society of America Journal*, *79*, 200–211. https://doi.org/10.2136/sssaj2014.08.0325

Li, C. (1996). The DNDC model. In D. S. Powlson, P. Smith, & J. U. Smith (Eds.), *Evaluation of soil organic matter models* (pp. 263–267). Berlin: Springer. https://doi.org/10.1007/978-3-642-61094-3_20

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*, 18–22.

Ludwig, B., Wolf, I., & Teepe, R. (2004). Contribution of nitrification and denitrification to the emission of N_2O in a freeze–thaw event in an agricultural soil. *Journal of Plant Nutrition and Soil Science*, *167*, 678–684. https://doi.org/10.1002/jpln.200421462

McDaniel, M. D., Simpson, R. R., Malone, B. P., McBratney, A. B., Minasny, B., & Adams, M. A. (2017). Quantifying and predicting spatio-temporal variability of soil CH_4 and N_2O fluxes from a seemingly homogenous Australian agricultural field. *Agriculture, Ecosystems & Environment*, *240*, 182–193. https://doi.org/10.1016/j.agee.2017.02.017

Mishurov, M., & Kiely, G. (2011). Gap-filling techniques for the annual sums of nitrous oxide fluxes. *Agricultural and Forest Meteorology*, *151*, 1763–1767. https://doi.org/10.1016/j.agrformet.2011.07.014

Moffat, A.M., Papale, D., Reichstein, M., Hollinger, D.Y., Richardson, A.D., Barr, A.G., … Stauch, V. J. (2007). Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agricultural and Forest Meteorology*, *147*, 209–232. https://doi.org/10.1016/j.agrformet.2007.08.011

Mumford, M.T., Rowlings, D.W., Scheer, C., De Rosa, D., & Grace, P.R. (2019). Effect of irrigation scheduling on nitrous oxide emissions in intensively managed pastures. *Agriculture, Ecosystems & Environment*, *272*, 126–134. https://doi.org/10.1016/j.agee.2018.11.011

Parkin, T. (2008). Effect of sampling frequency on estimates of cumulative nitrous oxide emissions. *Journal of Environmental Quality*, *37*, 1390–1395, https://doi.org/10.2134/jeq2007.0333

Philibert, A., Loyce, C., & Makowski, D. (2013). Prediction of N_2O emission from local information with random forest. *Environmental Pollution*, *177*, 156–163. https://doi.org/10.1016/j.envpol.2013.02.019

Pya, N., & Wood, S.N. (2015). Shape constrained additive models. *Statistics and Computing*, *25*, 543–559. https://doi.org/10.1007/s11222-013-9448-7

Rafique, R., Anex, R., Hennessy, D., & Kiely, G. (2012). What are the impacts of grazing and cutting events on the N_2O dynamics in humid temperate grassland? *Geoderma*, *181–182*, 36–44. https://doi.org/10.1016/j.geoderma.2012.03.006

Rasmussen, C., & Williams, C. (2006). *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.

Reeves, S., & Wang, W. (2015). Optimum sampling time and frequency for measuring N_2O emissions from a rain-fed cereal cropping system. *Science of the Total Environment*, *530*, 219–226. https://doi.org/10.1016/j.scitotenv.2015.05.117

RStudio Team (2015). *RStudio: Integrated development for R*. Boston, MA: RStudio. Retrieved from http://www.rstudio.com/

Savage, K., Phillips, R., & Davidson, E. (2014). High temporal frequency measurements of greenhouse gas emissions from soils. *Biogeoscience*, *11*, 2709–2720. https://doi.org/10.5194/bg-11-2709-2014

Scheer, C., Rowlings, D., Firrell, M., Deuter, P., Morris, S., Riches, D., … Grace, P. (2017). Nitrification inhibitors can increase post-harvest nitrous oxide emissions in an intensive vegetable production system. *Scientific Reports*, *7*. https://doi.org/10.1038/srep43677

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Shcherbak, I., Millar, N., & Robertson, G. P. (2014). A global meta-analysis of the nonlinear response of soil nitrous oxide (N_2O) emissions to fertilizer nitrogen. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 199–204. https://doi.org/10.1073/pnas.1322434111

Shang, Z., Abdalla, M., Kuhnert, M., Albanito, F., Zhou, F., Xia, L., & Smith, P. (2020). Measurement of N_2O emissions over the whole year is necessary for estimating reliable emission factors. *Environmental Pollution*, *259*, 113864. https://doi.org/10.1016/j.envpol.2019.113864

Smith, K. A., & Dobbie, K. E. (2001). The impact of sampling frequency and sampling times on chamber-based measurements of N_2O emissions from fertilized soils. *Global Change Biology*, *7*, 933–945. https://doi.org/10.1046/j.1354-1013.2001.00450.x

Stehfest, E., & Bouwman, L. (2006). N_2O and NO emission from agricultural fields and soils under natural vegetation: summarizing available measurement data and modeling of global annual emissions. *Nutrient Cycling in Agroecosystems*, *74*, 207–228.

Taki, R., Wagner-Riddle, C., Parkin, G., Gordon, R., & VanderZaag, A. (2018). Comparison of two gap-filling techniques for nitrous oxide fluxes from agricultural soil. *Canadian Journal of Soil Science*, *99*, 12–24. https://doi.org/10.1139/cjss-2018-0041

Teepe, R., Brumme, R., & Beese, F. (2001). Nitrous oxide emissions from soil during freezing and thawing periods. *Soil Biology and Biochemistry*, *33*, 1269–1275. https://doi.org/10.1016/S0038-0717(01)00084-0

van der Weerden, T.J., Clough, T.J., & Styles, T.M. (2013). Using near-continuous measurements of N_2O emission from urine-affected soil to guide manual gas sampling regimes. *New Zealand Journal of Agricultural Research*, *56*, 60–76. https://doi.org/10.1080/00288233.2012.747548

van der Weerden, T. J., Luo, J., Di, H. J., Podolyan, A., Phillips, R. L., Saggar, S., … Rys, G. (2016). Nitrous oxide emissions from urea fertilizer and effluent with and without inhibitors applied to pasture. *Agriculture, Ecosystems & Environment*, *219*, 58–70. https://doi.org/10.1016/j.agee.2015.12.006

Verhoeven, E., Pereira, E. I., Decock, C., Suddick, E. C., Angst, T. E., & Six, J. (2017). Toward a better assessment of biochar-nitrous oxide mitigation potential at the field scale. *Journal of Environmental Quality*, *46*, 237–246. https://doi.org/10.2134/jeq2016.10.0396

Venterea, R.T., Petersen, S.O., De Klein, C. A.M., Pedersen, A.R., Noble, A. D.L., Rees, R.M., … Parkin, T.B. (2020). Global Research Alliance N_2O chamber methodology guidelines: Flux calculations, *Journal of Environmental Quality*. https://doi.org/10.1002/jeq2.20118

Wagner-Riddle, C., Congreves, K. A., Abalos, D., Berg, A. A., Brown, S. E., Ambadan, J. T., … Tenuta, M. (2017). Globally important nitrous oxide emissions from croplands induced by freeze-thaw cycles. *Nature Geoscience*, *10*, 279–283. https://doi.org/10.1038/ngeo2907

Webb, J.R., Hayes, N.M., Simpson, G.L., Leavitt, P.R., Baulch, H.M., & Finlay, K. (2019). Widespread nitrous oxide undersaturation in farm waterbodies creates an unexpected greenhouse gas sink. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 9814–9819. https://doi.org/10.1073/pnas.1820389116

Wilkinson, M., Dumontier, M., Albersberg, I. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3. https://doi.org/10.1038/sdata.2016.18

Wood, S., Pya, N., & Safken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, *111*, 1548–1575. https://doi.org/10.1080/01621459.2016.1180986

Wu, W., Dandy, G.C., & Maier, H.R. (2014). Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality

modeling. *Environmental Modelling & Software*, *54*, 108–127. https://doi.org/10.1016/j.envsoft.2013.12.016

Wu, W., May, R., Dandy, G.C., & Maier, H.R. (2012). A method for comparing data splitting approaches for developing hydrological ANN models. Paper 394. Paper presented at the 6th International Congress on Environmental Modelling and Software, Leipzig, Germany.

Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Sigut, L., … Reichstein, M. (2018). Basic and extensible post-processing of eddy covariance flux data with REddyProc. *BiogeoScience*, *15*, 5015–5030. https://doi.org/10.5194/bg-15-5015-2018

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Dorich CD, De Rosa D, Barton L, et al. Global Research Alliance N_2O chamber methodology guidelines: Guidelines for gap-filling missing measurements. *J. Environ. Qual.* 2020;49:1186–1202.
https://doi.org/10.1002/jeq2.20138