Global research alliance N2O chamber methodology guidelines: statistical considerations, emission factor calculation, and data reporting

de Klein, CAM; Alfaro, Marta A; Giltrap, Donna; Topp, CFE; Simon, Priscila L; Noble, Alasdair; van der Weerden, TJ

Published in:
Journal of Environmental Quality

DOI:
10.1002/jeq2.20127

Print publication: 01/09/2020

Received: 28 February 2020 | Accepted: 26 June 2020

DOI: 10.1002/jeq2.20127

Journal of Environmental Quality

SPECIAL SECTION:
GRA N_2O CHAMBER METHODOLOGY GUIDELINES

Global Research Alliance N_2O chamber methodology guidelines: Statistical considerations, emission factor calculation, and data reporting

Cecile A. M. de Klein[1] | Marta A. Alfaro[2] | Donna Giltrap[3] | Cairistiona F. E. Topp[4] | Priscila L. Simon[1] | Alasdair D. L. Noble[5] | Tony J. van der Weerden[1]

[1] AgResearch, Invermay Agricultural Centre, Mosgiel 9053, New Zealand

[2] INIA, Fidel Oteíza 1956, piso 12, Providencia, Santiago, Chile

[3] Maanaki Whenua Landcare Research, Palmerston North, New Zealand

[4] SRUC, West Mains Rd. Edinburgh, Scotland EH9 3JG, UK

[5] AgResearch, Lincoln Research Centre, Lincoln, New Zealand

Correspondence
Cecile A. M. de Klein, AgResearch, Invermay Agricultural Centre, Mosgiel 9053, New Zealand.
Email: cecile.deklein@agresearch.co.nz

Assigned to Associate Editor Søren Petersen.

Funding information
New Zealand Government, in support of the objectives of the Livestock Research Group of the Global Research Alliance on Agricultural Greenhouse Gases; Scottish Government's Strategic Research Programme

Abstract
Static chambers are often used for measuring nitrous oxide (N_2O) fluxes from soils, but statistical analysis of chamber data is challenged by the inherently heterogeneous nature of N_2O fluxes. Because N_2O chamber measurements are commonly used to assess N_2O mitigation strategies or to determine country-specific emission factors (EFs) for calculating national greenhouse gas inventories, it is important that statistical analysis of the data is sound and that EFs are robustly estimated. This paper is one of a series of articles that provide guidance on different aspects of N_2O chamber methodologies. Here, we discuss the challenges associated with statistical analysis of heterogeneous data, by summarizing statistical approaches used in recent publications and providing guidance on assessing normality and options for transforming data that follow a non-normal distribution. We also recommend minimum requirements for reporting of experimental and metadata of N_2O studies to ensure that the robustness of the results can be reliably evaluated. This includes detailed information on the experimental site, methodology and measurement procedures, gas analysis, data and statistical analyses, and approaches to generate EFs, as well as results of ancillary measurements. The reliability, robustness, and comparability of soil N_2O emissions data will be improved through (a) application, and reporting, of more rigorous methodological standards by researchers and (b) greater vigilance by reviewers and scientific editors to ensure that all necessary information is reported in scientific publications.

1 | INTRODUCTION

Static (or non-steady-state) chambers are the most commonly used method for measuring nitrous oxide (N_2O)

Abbreviations: EF, emission factor; EFDB, emission factor database; GAMM, generalized additive mixed model; IPCC, Intergovernmental Panel on Climate Change; REML, restricted maximum likelihood.

fluxes from agricultural soils, as this method is relatively inexpensive, versatile in the field, and very easy to adopt (Holland et al., 1999; Hutchinson & Livingston, 1993; Kanemasu, Powers, & Sij, 1974; Mosier, 1989; Rochette & Eriksen-Hamel, 2008). Nitrous oxide chamber measurements are often used to assess emissions from a nitrogen (N) source to determine cumulative emissions, to provide data for calculating country- or region-specific N_2O emission factors (EFs), to test mitigation strategies, or to investigate drivers of emissions. All of these objectives require reliable results that can be compared among different studies, yet many of the challenges of chamber methodologies have the potential to bias results, or third-party interpretation of those results. This in turn limits interstudy comparisons and assessment of the reliability and uncertainty associated with the results of the individual studies. This paper is part of a special section of the *Journal of Environmental Quality* that considers the challenges associated with all key aspects of chamber methodologies, including design (Clough et al., 2020), deployment (Charteris et al., 2020), gas analysis (Harvey et al., 2020), automated chambers (Grace et al., 2020), flux calculation methods (Venterea et al., 2020), modeling (Giltrap et al., 2020), and gap-filling procedures (Dorich et al., 2020). These papers collectively provide guidance on best practices and factors that need to be considered when adopting N_2O chamber methodologies. Here, we discuss some of the pitfalls of statistical analysis of N_2O emissions that are inherently variable and rarely normally distributed. This violates assumptions of many statistical tests, and data transformation is often required prior to statistical analysis. The large variability in N_2O emissions derived from static chamber measurements results in large CVs in flux data: 13–57% (Yamulki, Goulding, Webster, & Harrison, 1995), 60–81% (Khalil, van Cleemput, Rosenani, & Schmidhalter, 2007) 8–108% (Chadwick et al., 2014), and 31–168% (Matthias, Yarger, & Weinbeck, 1978). Calculation of mean fluxes and statistical analyses of N_2O fluxes from replicated experiments must adequately represent this variability. Kravchenko and Robertson (2015) discussed the statistical challenges of N_2O chamber measurements when comparing treatment differences at individual time points. Here, we focus on the statistical challenges for comparing cumulative N_2O emissions and N_2O EFs between treatments, and their driving variables.

Core Ideas

- N_2O chamber measurements are often used for estimating N_2O emission factors (EFs).
- Accurately determining EFs is challenging due to the inherently variable nature of N_2O fluxes.
- We provide guidance on statistical analysis and EF calculation of N_2O chamber measurements.
- Accurate reporting of data ensures the robustness of N_2O results can be reliably evaluated.
- Authors should adhere to minimum requirements for reporting of experimental and metadata.

Emission factors represent the proportion or percentage of the N applied as a given N source (e.g., urine, manure, or fertilizer) that is emitted as N_2O in excess of background N_2O emissions. The Intergovernmental Panel on Climate Change (IPCC) methodology for estimating national greenhouse gas inventories (Eggleston, Buendia, Miwa, Ngara, & Tanabe, 2006) relies heavily on the use of EFs, and these are most commonly calculated from N_2O field measurements using chamber methodologies (Cardenas et al., 2010; Chadwick et al., 2018; de Klein, Smith, & Monaghan, 2006; Luo, Saggar, van der Weerden, & de Klein, 2019; Saggar et al., 2015; van der Weerden et al., 2016). To ensure that the soundness and robustness of the results can be verified, and that derived EFs can be reliably evaluated, it is also important that there is a consistent approach to calculating EFs. This includes how to calculate EFs from single experiments, how to estimate country- or region-specific EFs using results from multiple experiments, and/or how to conduct meta-analyses from published data. As mentioned, this paper is part of a series on guidelines for using N_2O chamber methodologies, and here we focus mainly on minimum requirements for calculating EFs from single experiments. However, we also provide some examples of meta-analyses of results from multiple experiments for estimating country- or region-specific EFs. The intended audience of this paper is soil and environmental scientists that use N_2O chamber methodologies, and the paper does not provide an in-depth analysis of statistical methods but instead aims to provide an overview of some of the pitfalls associated with statistical analyses and EF calculation when using static chamber methodologies.

For EF results to be accepted by the IPCC into the EF database (EFDB; https://ghgprotocol.org/Third-Party-Databases/IPCC-Emissions-Factor-Database), the results must be published in refereed journals. Thus, as to obtain reliable information for publication purposes, and to allow comparison of results across the globe, a minimum set of information must be provided together with the scientific results of specific experiments. In addition, reporting the results with experimental data and metadata also allows researchers around the world to compare the results of studies that have generated EFs and determined treatment effects. Buckingham et al. (2014) found that a poor description of experimental and environmental data in published studies restricted their meta-analysis of N_2O emission data. Finally, the development and evaluation of process-based models also relies on robust reporting of (meta)data and supporting information (Giltrap et al., 2020).

This paper reviews (a) statistical considerations for inherently heterogeneous data; (b) calculation of EFs for single experiments, as well as current and emerging statistical approaches for analyzing results from multiple experiments to provide national or regional average EFs; and (c) requirements for reporting of experimental data, environmental data, and metadata for verifying and accepting EFs estimates, or for modeling purposes.

2 | STATISTICAL CONSIDERATIONS FOR HETEROGENEOUS DATA

Accurately determining N_2O emissions from agricultural soils is a major challenge, due to the large variability of the environmental variables that affect microbial processes responsible for the emissions. Charteris et al. (2020) provide recommendations for experimental design and deployment of chambers to reduce the uncertainty associated with the spatial, temporal, and experimental variability in N_2O fluxes. These authors also discuss the use of power analysis to determine the required number of replicates. Here, we consider issues relating to the statistical analysis of spatially heterogenous cumulative N_2O flux estimates. This spatial variability has been considered log-normal at different scales, from plot to paddock to farm to landscape (Oenema, Velthof, Yamulki, & Jarvis, 1997; van Cleemput, Vermoesen, de Groot, & van Ryckeghem, 1994), although normal distributions have also been reported (Petersen, 1999).

2.1 | Assessment of normality and transformation

Before applying any standard ANOVA procedures, several assumptions must be established concerning the underlying error structure of the data. Among these is the assumption of normality. The effects that violations of the assumption of normality have on the efficacy of parametric statistical tests, such as the *t* test, have long been known (Cochran, 1947; Hey, 1938). Non-normality will influence the ability of a statistical test to perform at the stated α level—an effect that Cochran (1947) refers to as the validity of the test. Non-normality also affects the power of a statistical test to detect differences when real differences in the data actually exist.

The high variability of N_2O emissions often manifests as positively skewed distributions, and for individual experiments with a relatively small number of observations (e.g., <50), you can use statistical packages to generate diagnostic plots to get a visual assessment of your data (Albanito et al., 2017). There are also some simple quantitative ways to detect skewed distributions data. For example, for positive data that are bounded by zero, the distribution is likely to be skewed if the mean is less than twice the standard deviation. Also, if the standard deviation across different groups in a population increases as the mean increases, the constant variance assumption required for linear models is violated. Finally, there are also statistical tests that can verify normality, such as Kolmogorov–Smirnov or Shapiro–Wilk tests (Mohd Razali & Yap, 2011).

If a dataset follows a non-normal distribution, data transformations are commonly used to stabilize variances prior to statistical analysis. Those often used are logarithmic, square root, cube root, and Box–Cox transformations (Albanito et al., 2017; Chadwick et al., 2018; Luo et al., 2019; van der Weerden et al., 2016, 2020). The transformation options can either be selected based on the (biophysical) theory that suggests a certain distribution is most likely for the data, or on an empirical approach to ensure an approximately normal distribution of residuals. As N_2O measurement data are typically log-normally distributed, log-transformation is a commonly used method (Stehfest & Bouwman, 2006). However, the downside of log-transformation (and also square root and Box–Cox transformations) is that values ≤ 0 cannot be transformed. Some researchers deal with this by setting a minimum detectable flux limit and setting all smaller measured fluxes to this limit (Stehfest & Bouwman, 2006), but this will bias the outcome of any analysis. Others use a $\log(x + a)$ transformation, where a values are chosen by selecting the largest negative value and then adding a very small amount to this to ensure all values are >0 (IPCC, 2019; van der Weerden et al., 2016, 2020). Alternatively, cube root transformations can be used as this can be applied to values ≤ 0 (Albanito et al., 2017).

When estimating confidence intervals for log-normally distributed data, the standard approach is to calculate the interval on the log-scale and then back-transform the extremes. However, this poses challenges, as there are situations where the back-transformed mean value lies outside the back-transformed confidence interval (Olsson, 2005). In such cases, other approaches can be used, including the Cox method or estimating generalized confidence intervals (Olsson, 2005). Another downside is that if this back-transformed mean is used as an estimate of the mean on the original scale, it will be biased (Rothery, 1988). This effect is especially well reported for log-transformation (Sprugel, 1983), whereas more limited information is available for other types of transformations (Strimbu, Amarioarei, McTague, & Paun, 2018). Kelliher et al. (2014) used multiplicative scaling to bias-correct their back-transformed estimates and get their weighted mean to be the same as the overall mean value (N. Cox, personal communication, 2013).

To deal with these challenges of data transformation, some researchers have used *transformed* data for statistical analysis to assess treatment effects, but *untransformed* data to calculate mean values and uncertainty ranges for assessing country-specific EFs (Chadwick et al., 2018; IPCC, 2019, van der Weerden et al., 2020). Although this may appear to be an inconsistent approach to data analysis, it is important to note that there is a subtle but important difference between the statistical analysis to identify significant differences between treatments and other categorical variables (e.g., N sources, regions, animal types, climates), and estimating N_2O EF values. The former can help justify the disaggregation of EFs in different categories but requires data transformation if the data are skewed. However, the latter needs to provide the best possible estimate of the true EF for each treatment, N source, and variable for use in a country's inventory calculations. Furthermore, calculation of the means on the untransformed scale can be augmented with bootstrapped confidence intervals (Efron & Tibshirani, 1986), which would be more robust than the alternative of back-transforming the confidence limits from the analysis of the transformed data.

2.2 | Statistical modeling of nitrous oxide datasets

Statistical or empirical modeling is often used for identifying factors that significantly influence N_2O emissions, in order to estimate country- or region-specific N_2O emissions (Giltrap et al., 2020), or aid decisions on disaggregation of EFs for different N sources and/or soil, climatic, or landscape features (e.g., soil drainage class, season, or slope; Chadwick et al., 2018; Saggar et al., 2015; Shrestha et al., 2014). Statistical modeling can also be used to estimate missing data values in N_2O time series datasets (gap filling). For a detailed description on gap-filling procedures, please refer to Dorich et al. (2020).

To analyze results from individual experiments, researchers generally apply linear fixed-effects models (e.g., fixed regression parameter models, multiple linear regression models, generalized linear models; Albanito et al., 2017). In contrast, linear mixed-effects models are now more widely used for analyzing datasets from multiple experiments (Hafner et al., 2018; IPCC, 2019; van der Weerden et al., 2020). There are advantages in mixed-effect models, as they can include parameters that have either a fixed (i.e., nonrandom) effect or a random effect on the N_2O emission estimates. These random effects may have a significant effect on the outcome of the analysis, even if the random variables are generally of little interest to the researcher. For example, in a single experiment, random effects could include blocking variables or the site number, whereas in an analysis of multiple experiments, the "experiment ID number" or "research institute" could be considered a random effect. Fixed effects are assumed to be correlated with independent variables, whereas for random effects, the assumption is that the individual-specific effects are uncorrelated with the independent variables. In a recent analysis of ammonia emission data, Hafner et al. (2018) used "institute" (i.e., research institute that conducted the measurements) as a random effect, and their results showed large apparent difference among institutes. The authors found that, for their dataset, the inclusion of random predictors resulted in a significantly better model. They found that a model with only one fixed-effect predictor and two random-effect predictors gave better results than a model with many fixed-effect predictors but no random-effect predictors (Hafner et al., 2018). It is yet to be shown if a similar "institute" effect is important for N_2O emission data. Linear mixed-effect models are relatively complex, and the number and specification of the fixed- and random-effect factors that are included can potentially affect the outcomes and interpretability of these models (Philibert, Loyce, & Makowski, 2012). Albanito et al. (2017) used a generalized additive mixed model (GAMM) to statistically analyze their N_2O emissions data in tropical agricultural systems. Because a GAMM has fewer assumptions (e.g., no assumption of normality or constant variances), it provides a more flexible procedure to describe the variability across different parameters.

Due to recent advances in statistics and computational power, the literature is now providing many alternative models to analyze N_2O data. The advantage or disadvantage of using a specific approach for statistically analyzing N_2O datasets depends on the intended purpose of the analysis, and any data- or experiment-specific conditions are likely to require a tailored approach. It is recommended that when conducting (meta-)analyses with datasets from multiple experiments, researchers base their choice of model on published examples that match their specific purpose and circumstances (e.g., environmental conditions and methodologies) or, where possible, consult a statistician for advice on the most appropriate models to use.

3 | CALCULATING NITROUS OXIDE EMISSION FACTORS

In single experiments, EF values are calculated by subtracting the cumulative N_2O emissions occurring in a control treatment where no N was added (N_2O_0) from the cumulative N_2O emissions in a given experimental treatment where N was added (N_2O_x), then dividing this by the amount of N applied (N_{Appx}):

$$\mathrm{EF}\,(\%) = \left(\frac{N_2O_x - N_2O_0}{N_{APPx}}\right) \times 100 \qquad (1)$$

where the units of N_2O and N inputs need to be the same (e.g., N_2O in kg N_2O-N ha^{-1} yr^{-1} and N_{Appx} in kg N ha^{-1} yr^{-1}).

For estimating EFs from an experimental trial, a control treatment should therefore always be included. Ideally, the control and N treatments of the experiments are blocked, so that independent EFs can be calculated for each replicate block, which allows the calculation of mean and variance of each treatment EF value, and differences in the EF among treatments to be examined. Taking pretreatment N_2O flux measurements can provide information on preexisting spatial patterns of N_2O emissions and thereby assist with experimental design and blocking (Charteris et al., 2020). If blocking is not possible, the mean cumulative emission from the control can be subtracted from the individual replicates of the N treatments, thereby maintaining replicate-level data, although this reduces some of the statistical power provided by true replication. Alternatively, EF values can be calculated using the mean cumulative emissions of the control treatments and the mean cumulative emissions for the N treatments over all replicates, but this is not advisable as it will result in a single EF value for each experimental treatment with no indication of variance.

Robust estimates of EF values rely on accurate estimates of cumulative N_2O emissions from experimental trials. Estimating cumulative emissions from noncontinuous chamber measurements is challenged by the high temporal variability in N_2O fluxes. This requires careful consideration of the frequency of sampling throughout the measurements period, as well as of the method of interpolating between sampling points (Charteris et al., 2020; Dorich et al., 2020). Charteris et al. (2020) recommends that, as a minimum, N_2O measurements should be taken at least twice per week when higher emissions are likely to be occurring, and at even higher intensities around events (e.g., rainfall, fertilization, cultivation). When conditions are conducive to near-zero N_2O fluxes (e.g., in dry or cold soils), weekly or biweekly sampling may be adequate. The most common method for estimating cumulative emissions from noncontinuous chamber measurements is linear interpolation (Dorich et al., 2020). Although this is a simple method, it relies on a sampling frequency that can adequately capture the temporal variability in fluxes. The longer the time gap between sampling points, the more uncertain the estimates will be. More advanced approaches to estimating cumulative emissions include "gap-filling" techniques that estimate daily N_2O fluxes when sampling was not conducted. Dorich et al. (2020) provides detailed guidance on approaches to estimate cumulative emissions using "gap-filling."

It is normally expected that N_2O emissions from treatments receiving added N will be greater than emissions from no-N control treatments. However, in cases where cumulative N_2O emissions are greater in the control than in the treatment replicate (resulting in a negative EF value), we do not recommend simply substituting EF = 0 for these values as this would upwards bias the calculated mean EF. Rather, we suggest including the actual value in the subsequent statistical analysis, unless excluding that value as an outlier is justified. It should also be noted that some studies have found nonlinear relationships between N_2O emissions and synthetic N fertilizer addition rate (Hoben, Gehl, Millar, Grace, & Robertson, 2011; Philibert et al., 2012; Shcherbak, Millar, & Robertson, 2014). This implies that a fertilizer EF calculated for a single rate of N addition should not necessarily be generalizable to other fertilizer N addition rates, even within the same management and cropping system. In contrast, a recent meta-analysis of EF values for urine and dung found no relationship between N loading rate and EF value (van der Weerden et al., 2020). Nevertheless, it is important that in experiments aimed at estimating EFs, the treatment application rates of both synthetic N fertilizer as well as animal excreta represent typical agricultural practices and conditions (Charteris et al., 2020).

A key criterion for EF values to be included in the IPCC EFDB is the length of time of the measurement period. Although a recent meta-analysis of global data did not find a relationship between EF value and length of the experiment (IPCC, 2019), all experimental information of greenhouse gas emissions for EF calculation should ideally be based on at least a year of data, as EFs for the whole year can be significantly higher than those reported for the growing season (Shang et al., 2020). In situations that include different cropping rotations, or productive seasons, emissions from the full year (including fallow periods) or the full crop rotation should be calculated. Where a study involves estimating EFs for a specific N source on a specific land use (e.g., N fertilizer application or a urine deposition onto permanent pasture), the duration of measurements could be shorter than a year, as long as the full N-induced "emission envelope" has been captured (Charteris et al., 2020). This means that measurements should continue until the N_2O emissions, as well as the soil mineral N levels in the treatment plots, are no longer significantly different from those in the control plots. Considering differences in soil mineral N as well is important, as N_2O emissions might be at control levels due to other limiting factors, such as low soil moisture. If soil mineral N levels in the treatment plots are still higher than in the control plots, measurements will need to continue to capture any emission difference that might occur going forward (e.g., after rainfall). It is also possible that some years or seasons may have atypical meteorological conditions, leading to drier or wetter, or warmer or cooler conditions compared with long-term averages. In such cases, the estimated EF may not necessarily provide a representative average for that system or region, and multiyear measurements may be required.

To estimate country- (or region-) specific EFs, a meta-analysis that combines N_2O EF results from a range of trials can be conducted (Buckingham et al., 2014; Chadwick et al., 2018; Kelliher et al., 2014; van der Weerden et al., 2020; Wang et al., 2018). The determination of the average EF for a country or region is challenging, as different studies may use different treatments of N sources and/or collate different sets of ancillary data. This can result in unbalanced datasets and gaps in the dataset of explanatory variables. For example, Kelliher et al. (2014) used the results of 185 field trials from a wide range of studies to estimate the mean EF of N_2O from N sources applied to pastoral soils across New Zealand. However, for some N sources, N_2O EF results were only available for certain land types (e.g., lowland sites for some N sources and hill country sites for other sources). To deal with such issues, Kelliher et al. (2014) used a mixed-effects model with a restricted maximum likelihood (REML) method on log(EF) data using trial, site, N source, topographic type, drainage class, rain ratio, season, and the interaction between N source and topographic type as random effects. The mixed-effects linear model was fitted to log-transformed EF values, and the results were reported as bias-corrected best linear unbiased predictions. The bias correction involved scaling the back-transformed estimates by the amount required to get their weighted mean to be the same as the overall mean of the EF value. Chadwick et al. (2018) also used a mixed-effects model with REML but used Box–Cox transformations to normalize their data.

4 | DATA REPORTING REQUIREMENTS

Reporting of experimental data and metadata allows researchers around the world to compare the results of studies that have generated EFs and/or determined treatment effects. Furthermore, the development and evaluation of process-based models at different scales also requires the reporting of experimental and environmental conditions of N_2O emission studies. This section summarizes the minimum requirements for reporting of data for calculating EFs and for development of process-based models (Table 1). The recommendations are a combination of those listed in the earlier version of the Global Research Alliance guidelines (refer to de Klein et al., 2020) and those suggested by Buckingham et al. (2014).

4.1 | Reporting requirements for calculating emission factors

The experimental setup should include a control treatment, so that EF can be calculated. Treatments, including the control, should be reported in detail, indicating the number of replicates, N and, where relevant, carbon (C) application rates; the application date; the gas sampling starting date; and the length of the measurement period. The N and C loading rates should be determined from N and C analysis of the N source (e.g., slurry, manure, urine, dung) from samples that are taken at the time of application onto the soil (as it may differ from the N and C loading rates that may have been determined from samples taken during the collection of the N source). When using animal manures, the application method should be reported (e.g., subsurface, slurry surface broadcast, trailing shoe, incorporated). Reporting on control and treatment plots should also include previous management history of grazing, nutrient applications, and crops, including winter cover crops.

Reporting of the chamber methodology should include details on the design and use of the chambers as recommended in these guidelines (Charteris et al., 2020; Clough et al., 2020). This should include details on treatments, replication and trial design; gas sampling procedure (including t_0 measurements; Charteris et al., 2020), deployment time, and timing and duration of the experiment; and N_2O flux results for each sampling date, with indication of variability (standard error or standard deviation). For studies using animal urine or dung patches, information of the treatment application method should be provided, including details on the treated area inside the chamber base (e.g., centrally applied only, thus leaving a diffusional area; spread over the entire area; one larger patch vs. a number of smaller patches; see Charteris et al., 2020)

Reporting of ancillary data should include key crop and soil information, including crop yield and N content. If the experiment is related to mown pasture or crops, information should be provided on the height and frequency of cut, and whether the residues were left or

TABLE 1 Recommended minimum requirements for reporting of experimental and metadata of N_2O studies for calculating emission factors and for development of process-based models

Parameter	Desirable for emission factors (EFs)	Desirable for process-based model development
Experimental site and setup	Latitude, longitude, altitude	All data from the previous column, plus:
	Soil type and classification	Historic information on site and soil management and climatic variables for at least 3 yr
	Previous and current site and soil management, including crop and catch crop type, and fertilizer management; going back at least 1 yr, preferably over 3 yr	Site management: weed and pest control; drainage limitations or other relevant aspect
	Initial soil chemical conditions: available N (NH_4 and NO_3), total N and C contents, pH; at relevant depths	Other relevant aspects of soil fertility
	Initial soil physical characteristics: texture, bulk density; macro- and microporosity, at relevant depths	Field capacity or soil moisture release curve, soil conductivity wilting point, soil hydraulic conductivity
		Number and depth of soil layers
Methodology	Details of chamber design (see Clough et al., 2020) and deployment (see Charteris et al., 2020)	All data from the previous column, plus:
	Treatment details: rates of application; total N and total C inputs; NH_4–N and NO_3–N inputs; dates of application and method of application; fertilizer or manure type and composition (C content, total N and inorganic N contents, pH, and dry matter content); and indication of how representative these are for their circumstances	For control treatments, all information described for treatment plots should be clearly provided
	Trial (statistical) design and replication, number of chambers per plot	
	Duration of experiment (see Charteris et al., 2020)	
	Number of samples taken to estimate the flux from a single chamber	
	Chamber closure period	
	Time elapsed between measurements	
	Number of background control measurements	
	Average concentration of background control	
	measurements over the measurement period	
	N_2O emissions for each sampling date, with	
	indication of variability and associated errors for treatment results	
Ancillary measurements[a]	Average (max. and min.) soil temperature, at relevant depth for crops or pastures, for each sampling date	All data from the previous column, plus:
	Average (or max. and min.) air temperature, at relevant height, for each sampling date	Daily min. and max. temperatures
	Total daily rainfall and irrigation	Daily rainfall intensity information
	Average (max. and min.) soil and air temperature within the chamber, (when applicable) for each sampling date, at relevant height and depth	Daily solar radiation
	Soil moisture content at relevant depth, for each sampling date	Daily wind speed
	Soil available N (NH_4 and NO_3) at relevant depth as frequently as possible (e.g., weekly)	Daily relative humidity
	Bulk density in arable soils at key stages throughout the season (cultivation effects)	Soil drainage, if available

(Continues)

TABLE 1 (Continued)

Parameter	Desirable for emission factors (EFs)	Desirable for process-based model development
	Total yield and dry matter production for each component of the crop (e.g., straw and grain)	Seeding system for crops (no tillage, conventional tillage, other)
	Total N export in yield or dry matter production	Planting date
		Harvest date for crops and cutting and grazing dates for pastures
		In the case of pastures, an indication of dominant plant species
		If possible, an indication of material left on the field and its composition should be also noted, in relation to the seeding system used
Gas analysis	Equipment details including detector and precision of analyzer	Same as previous column
	For gas chromatography determinations, information on column used, temperatures in detector and oven	
	Detection limit for the method	
	Quality control information for gas analysis	
Data analysis	Flux calculation method	Same as previous column
	Statistical analysis procedures and the software package used	
	Uncertainty ranges of the results, including standard errors and the number of replicates	

[a] Measurements of soil water content, bulk density, and temperature allows researchers to apply the chamber bias correction (CBC) flux calculation method as discussed in Venterea et al. (2020).

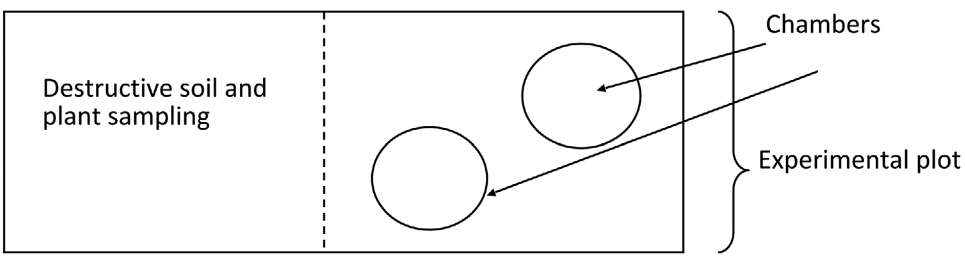

FIGURE 1 Example of experimental plot layout for N_2O flux measurements

removed from the area. If the experiment includes grazing, information on animal type and category, stocking rate, forage species, and the number of grazing days should also be included. Key climate and soil characteristics should be monitored and reported for the trial (see Table 1 for details). It is important to note that measurements of soil water content, bulk density, and temperature will allow researchers to apply the chamber bias correction (CBC) flux calculation method, as discussed in Venterea et al. (2020). These measurements should be taken as close as possible to, but outside, the chamber area to avoid any soil disturbance for gas determinations (Charteris et al., 2020). Soil samples taken for nutrients determination should also be taken outside the chamber area but should be representative of the chamber area and receive the same treatments. For example, when small experimental plots are used, split the plot in half so that one side can be used for gas determinations, whereas the other side can be used for destructive sampling (Figure 1).

Reporting of the gas analysis procedure should include details on the analytical equipment and its precision and detection limit (Harvey et al., 2020). The data analysis reporting should include details on the flux calculation methods used (see also Venterea et al., 2020), the statistical analysis performed (see sections above), and adequate information on the uncertainty of the N_2O emission

TABLE 2 Summary of guidance on statistical analysis, calculating emission factors (EFs), and data reporting of N_2O measurements using chamber methodologies

Topic or issue	Guidance and recommendations
Statistical analysis of inherently heterogeneous data	Assess the distribution of the dataset.
	If required, transform the data (most commonly log transformation).
	If negative values, use $(x + a)$ for log, square root, or Box–Cox transformations; or use cube root transformation.
	Back-transform the data and bias correct the means (e.g., by adjusting so that their weighted mean is the same as the overall mean value).
Calculating EFs	Always include control plots (ideally for each replicated block; otherwise ensure sufficient replication for estimating mean cumulative emission from control treatment).
	Ensure adequate sampling frequency, trial duration, and interpolation methods (Charteris et al., 2020; Dorich et al., 2020) for estimating cumulative emissions.
	Calculate EFs as indicated in Equation 1.
	Do not substitute negative EF values with zero but use the actual value in the statistical analysis.
	Mean EF values that are calculated from untransformed data can be augmented with bootstrapped confidence intervals.
	For individual trial data, use fixed-effect models to analysis treatment effects and estimate best average EF.
	For a meta-analysis of data from a range of different trials, use mixed-effects models to provide the best estimate of country- or region-specific EFs.
Data reporting	Provide sufficient details (see Table 1) on the experimental setup, methodology, and statistical analysis to enable:
	Inclusion of the results in the IPCC EF database
	Assessment of reliability and robustness of results
	Comparison with results obtained under similar conditions and with similar treatments
	Incorporation of the results in EF meta-analyses
	Use of the results for developing or refining process-based models

estimates (i.e., the standard error and the number of replicates; missing values). All replicated measurements should be recorded and reported on an individual plot basis (not averaged).

4.2 | Information required to evaluate process-based models

Because models must be sensitive enough to account for temporal and spatial variability, they require more detailed information than reporting of EF results. A summary of the additional information that is required for evaluating models is included in Table 1. Information requirements vary between models, so users should check with model developers, or documentation, for the necessary model-specific data. If sensitivity analyses have been previously published for a particular model, these can provide a useful guide regarding parameter values that can have a large influence on the model output. On the other hand, if a given model has not previously been calibrated for the system and conditions being measured, then your data may be used to calibrate the model. In this case, it is helpful if the data can be divided into two independent subsets (e.g., different treatments or different years) so that one may be used for model calibration and one for validation. Before testing any specific model, appropriate data requirements and underlying model assumptions should be discussed among modelers and empirical researchers. Giltrap et al. (2020) includes an overview of some of the key points for three commonly used process-based models.

As stated above, general information on the experimental site location and similar details should be reported explicitly. At least one full year of experimental information should be reported, to account for temporal variability, unless the objective of the study allows for shorter periods of analysis. Ideally, a number of experimental sites would be established over different soil and weather conditions, to provide variability for the model to be tested and make it applicable to different conditions. If possible, this should include a history of site and soil management for the previous 3 yr. Information on weed and pest control, drainage limitations, and other site-specific characteristics is also valuable.

5 | CONCLUSION

Statistical analysis of N_2O data from static chambers is challenged by the inherently heterogeneous nature of N_2O fluxes, which often requires data transformation. It is recommended that researchers conduct a careful visual examination of the data and the residuals, to assess the appropriateness of any transformation applied. Back-transformed means should be bias corrected, by multiplicative scaling, if their weighted mean is not the same as the overall mean. Alternatively, researchers can use *transformed* data to assess significant differences between treatments and use *untransformed* data to calculate mean values and uncertainty ranges for country-specific EFs (Table 2). For statistical analysis of multiple experiments, it is recommended to use mixed-effects models. To calculate N_2O EFs from single experiments, control plots need to be included, and the sampling frequency, trial duration, and interpolation methods need to be appropriate for estimating cumulative emissions (Table 2). The reliability, robustness, and comparability of soil N_2O emissions data will be improved through (a) application, and reporting, of more rigorous methodological standards by researchers, and (b) greater vigilance by reviewers and scientific editors to ensure that all necessary information is reported in scientific publications (Table 1).

ACKNOWLEDGMENTS

This paper was prepared with financial support from the New Zealand Government, in support of the objectives of the Livestock Research Group of the Global Research Alliance on Agricultural Greenhouse Gases (GRA); and the Scottish Government's Strategic Research Programme. We also thank Dr. Tim Parkin for invaluable contributions to the original version of the GRA N_2O Chamber Methodology Guidelines.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Cecile A. M. de Klein https://orcid.org/0000-0001-7840-0934
Marta A. Alfaro https://orcid.org/0000-0002-1881-4745
Donna Giltrap https://orcid.org/0000-0001-7919-0414
Cairistiona F. E. Topp https://orcid.org/0000-0002-7064-638X
Priscila L. Simon https://orcid.org/0000-0003-0391-6690
Alasdair D. L. Noble https://orcid.org/0000-0003-2292-8216
Tony J. van der Weerden https://orcid.org/0000-0002-6999-2584

REFERENCES

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*(1). https://doi.org/10.1038/srep44235

Buckingham, S., Anthony, S., Bellamy, P. H., Cardenas, L. M., Higgins, S., McGeough, K., & Topp, C. F. E. (2014). Review and analysis of global agricultural N_2O emissions relevant to the UK. *Science of the Total Environment*, *487*, 164–172. https://doi.org/10.1016/j.scitotenv.2014.02.122

Cardenas, L. M., Thorman, R., Ashlee, N., Butler, M., Chadwick, D., Chambers, B., … Scholefield, D. (2010). Quantifying annual N_2O emission fluxes from grazed grassland under a range of inorganic fertiliser nitrogen inputs. *Agriculture, Ecosystems & Environment*, *136*, 218–226. https://doi.org/10.1016/j.agee.2009.12.006

Chadwick, D. R., Cardenas, L. M., Dhanoa, M. S., Donovan, N., Misselbrook, T., Williams, J. R., … Rees, R. M. (2018). The contribution of cattle urine and dung to nitrous oxide emissions: Quantification of country specific emission factors and implications for national inventories. *Science of the Total Environment*, *635*, 607–617. https://doi.org/10.1016/j.scitotenv.2018.04.152

Chadwick, D. R., Cardenas, L. M., Misselbrook, T., Smith, K. A., Rees, R. M., Watson, C. J., … Dhanoa, M. S. (2014). Optimizing chamber methods for measuring nitrous oxide emissions from plot-based agricultural experiments. *European Journal of Soil Science*, *65*, 295–307. https://doi.org/10.1111/ejss.12117

Charteris, A. F., Chadwick, D. R., Thorman, R. E., Vallejo, A., de Klein, C. A. M., Rochette, P., & Cardenas, L. M. (2020). Global Research Alliance N_2O chamber methodology guidelines: Recommendations for deployment and accounting for sources of variability. *Journal of Environmental Quality*. https://doi.org/10.1002/jeq2.20126

Clough, T. J., Rochette, P., Thomas, S. M., Pihlatie, M., Christiansen, J. R., & Thorman, R. (2020). Global Research Alliance N_2O chamber methodology guidelines: Design considerations. *Journal of Environmental Quality*. https://doi.org/10.1002/jeq2.20117

Cochran, W. G. (1947). Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, *3*, 22–38. https://doi.org/10.2307/3001535

de Klein, C. A. M., Harvey, M. J., Clough, T. J., Petersen, S. O., Chadwick, D. R., & Venterea, R. T. (2020). Global Research Alliance N_2O chamber methodology guidelines: Introduction, with health and safety considerations. *Journal of Environmental Quality*. https://doi.org/10.1002/jeq2.20131

de Klein, C. A. M., Smith, L. C., & Monaghan, R. M. (2006). Restricted autumn grazing to reduce nitrous oxide emissions from dairy pastures in Southland, New Zealand. *Agriculture, Ecosystems & Environment*, *11*, 192–199. https://doi.org/10.1016/j.agee.2005.08.019

Dorich, C., Conant, R., Grace, P. R., Barton, L., de Rosa, D., Wagner-Riddle, C., & Fehr, B. (2020). Global Research Alliance N_2O

chamber methodology guidelines: Guidelines for gap-filling missing measurements. *Journal of Environmental Quality*. https://doi.org/10.1002/jeq2.20138

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, *1*, 54–75. https://doi.org/10.1214/ss/1177013815

Eggleston, H. S., Buendia, L., Miwa, K., Ngara, T., & Tanabe, K., (Eds.). (2006). *2006 IPCC guidelines for national greenhouse gas inventories*. Hayama, Japan: Institute for Global Environmental Strategies.

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

Grace, P. R., van der Weerden, A. 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

Hafner, S. D., Pacholski, A., Bittman, S., Burchill, W., Bussink, W., Chantigny, M., … Sommer, S. G. (2018). The ALFAM2 database on ammonia emission from field-applied manure: Description and illustrative analysis. *Agricultural and Forest Meteorology*, *258*, 66–79. https://doi.org/10.1016/j.agrformet.2017.11.027

Harvey, M. J., Sperlich, P., Clough, T. J., Kelliher, F. M., Martin, R. J., & Moss, R. (2020). Global Research Alliance N_2O chamber methodology guidelines: Recommendations for air sample collection, storage, and analysis. *Journal of Environmental Quality*. https://doi.org/10.1002/jeq2.20129

Hey, G. B. (1938). A new method of experimental sampling illustrated on certain non-normal populations. *Biometrika*, *30*, 68–80. https://doi.org/10.1093/biomet/30.1-2.68

Hoben, J. P., Gehl, R. J., Millar, N., Grace, P. R., & Robertson, G. P. (2011). Nonlinear nitrous oxide (N2O) response to nitrogen 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

Holland, E. A., Robertson, G. P., Greenberg, J., Groffman, P. M., Boone, R. D., & Gosz, J. R. (1999). Soil CO_2, N_2O, and CH_4 exchange. In G. P. Robertson et al. (Eds.), *Standard soil methods for long-term ecological research* (pp. 185–201). New York: Oxford University Press.

Hutchinson, G. L., & Livingston, G. P. (1993). Use of chamber systems to measure trace gas fluxes. In L. A. Harper et al. (Eds.), *Agricultural ecosystem effects on trace gases and global climate change* (pp. 63–78). Madison, WI: ASA, CSSA, and SSSA.

IPCC (2019). *2019 refinement to the 2006 IPCC guidelines for national greenhouse gas inventories*. Intergovernmental Panel on Climate Change. Retrieved from https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch11_Soils_N2O_CO2.pdf

Kanemasu, E. T., Powers, W. L., & Sij, J. W. (1974). Field chamber measurements of CO_2 flux from soil surface. *Soil Science*, *118*, 233–237.

Kelliher, F. M., Cox, N., van der Weerden, T. J., de Klein, C. A. M., Luo, J., Cameron, K. C., … Rys, G. (2014). Statistical analysis of nitrous oxide emission factors from pastoral agriculture field trials conducted in New Zealand. *Environmental Pollution*, *186*, 63–66. https://doi.org/10.1016/j.envpol.2013.11.025

Khalil, M. I., van Cleemput, O., Rosenani, A. B., & Schmidhalter, U. (2007). Daytime, temporal, and seasonal variations of N_2O emissions in an upland cropping system of the humid tropics. *Communications in Soil Science and Plan Analysis*, *38*, 189–204. https://doi.org/10.1080/00103620601094122

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

Luo, J., Saggar, S., van der Weerden, T., & de Klein, C. (2019). Quantification of nitrous oxide emissions and emission factors from beef and dairy cattle excreta deposited on grazed pastoral hill lands. *Agriculture, Ecosystems & Environment*, *270–271*, 103–113. https://doi.org/10.1016/j.agee.2018.10.020

Matthias, A. D., Yarger, D. N., & Weinbeck, R. S. (1978). A numerical evaluation of chamber methods for determining gas fluxes. *Geophysical Research Letters*, *5*, 765–768. https://doi.org/10.1029/GL005i009p00765

Mohd Razali, N., & Yap, B. (2011). Power comparisons of Shapiro–Wilk, Kolmogorov–Smirnov, Lilliefors, and Anderson–Darling tests. *Journal of Statistical Modeling and Analytics*, *2*, 21–33.

Mosier, A. R. (1989). Chamber and isotope techniques. In M. O. Andreae & D. S. Schimel (Eds.), *Exchange of trace gases between terrestrial ecosystems and the atmosphere* (pp. 175–187). New York: John Wiley & Sons.

Oenema, O., Velthof, G. L., Yamulki, S., & Jarvis, S. C. (1997). Nitrous oxide emissions from grazed grassland. *Soil Use and Management*, *13*, 288–295. https://doi.org/10.1111/j.1475-2743.1997.tb00600.x

Olsson, U. (2005). Confidence intervals for the mean of log-normal distribution. *Journal of Statistics Education*, *13*(1). https://doi.org/10.1080/10691898.2005.11910638

Petersen, S. O. (1999). Nitrous oxide emissions from manure and inorganic fertilizers applied to spring barley. *Journal of Environmental Quality*, *28*, 1610–1618. https://doi.org/10.2134/jeq1999.00472425002800050027x

Philibert, A., Loyce, C., & Makowski, D. (2012). Assessment of the quality of meta-analysis in agronomy. *Agriculture, Ecosystems & Environment*, *148*, 72–82. https://doi.org/10.1016/j.agee.2011.12.003

Rochette, P., & Eriksen-Hamel, N. S. (2008). Chamber measurements of soil nitrous oxide flux: Are absolute values reliable? *Soil Science Society of America Journal*, *72*, 331–342. https://doi.org/10.2136/sssaj2007.0215

Rothery, P. (1988). A cautionary note on data transformation: Bias in backtransformed means, *Bird Study*, *35*, 219–221, https://doi.org/10.1080/00063658809476992

Saggar, S., Giltrap, D. L., Davison, R., Gibson, R., de Klein, C. A. M., Rollo, M., … Rys, G. (2015). Estimating direct N_2O emissions from sheep, beef, and deer grazed pastures in New Zealand hill country: Accounting for the effect of land slope on the N_2O emission factors from urine and dung. *Agriculture, Ecosystems & Environment*, *205*, 70–78. https://doi.org/10.1016/j.agee.2015.03.005

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*. https://doi.org/10.1016/j.envpol.2019.113864

Shcherbak, I., Millar, N., & Robertson, G. P. (2014). Global metaanalysis 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*, 9199–9204. https://doi.org/10.1073/pnas.1322434111

Shrestha, N., Leta, O. T., Fraine, B., Garcia-Armisen, T., N.K, O., Servais, P., … Bauwens, W. (2014). Modelling *Escherichia coli* dynamics in the river Zenne (Belgium) using an OpenMI based integrated model. *Journal of Hydroinformatics*, *16*, 354–374. https://doi.org/10.2166/hydro.2013.171

Sprugel, D. G. (1983). Correcting for bias in log-transformed allometric equations. *Ecology*, *64*, 209–210. https://doi.org/10.2307/1937343

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. https://doi.org/10.1007/s10705-006-9000-7

Strimbu, B. M., Amarioarei, A., McTague, J. P., & Paun, M. M. (2018). A posteriori bias correction of three models used for environmental reporting. *Forestry*, *91*, 49–62. https://doi.org/10.1093/forestry/cpx032

van Cleemput, O., Vermoesen, A., de Groot, C.-J., & van Ryckeghem, K. (1994). Nitrous oxide emission out of grassland. *Environmental Monitoring and Assessment*, *31*, 145–152. https://doi.org/10.1007/BF00547190

van der Weerden, T. J., Cox, N., Luo, J., Di, H. J., Podolyan, A., Phillips, R. L., … Rys, G. (2016). Refining the New Zealand nitrous oxide emission factor for urea fertiliser and farm dairy effluent. *Agriculture, Ecosystems & Environment*, *222*, 133–137. https://doi.org/10.1016/j.agee.2016.02.007

van der Weerden, T. J., Noble, A. N., Luo, J., de Klein, C. A. M., Saggar, S., Giltrap, D., … Rys, G. (2020). Meta-analysis of nitrous oxide emission factors from ruminant excreta deposited onto New Zealand pastures. *Science of the Total Environment*, *732*. https://doi.org/10.1016/j.scitotenv.2020.139235

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

Wang, X., Zou, C., Gao, X., Guan, X., Zhang, W., Zhang, Y., … Chen, X. (2018). Nitrous oxide emissions in Chinese vegetable systems: A meta-analysis. *Environmental Pollution*, *239*, 375–383. https://doi.org/10.1016/j.envpol.2018.03.090

Yamulki, S., Goulding, K. W. T., Webster, C. P., & Harrison, R. M. (1995). Studies on no and N_2O fluxes from a wheat field. *Atmospheric Environment*, *29*, 1627–1635. https://doi.org/10.1016/1352-2310(95)00059-8

How to cite this article: de Klein CAM, Alfaro MA, Giltrap D, et al. Global Research Alliance N_2O chamber methodology guidelines: Statistical considerations, emission factor calculation, and data reporting. *J. Environ. Qual*. 2020;1–12. https://doi.org/10.1002/jeq2.20127