

## CLIMATE CHANGE AND AGRICULTURE RESEARCH PAPER

# The effects of spatial variability of nitrous oxide emissions from grazed pastures on the sampling distribution of chamber measurements

D. L. GILTRAP<sup>1\*</sup> AND A. J. R. GODFREY<sup>2</sup>

<sup>1</sup> Landcare Research, Palmerston North, New Zealand

<sup>2</sup> Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

(Received 28 October 2014; revised 23 March 2015; accepted 4 May 2015;  
first published online 9 June 2015)

## SUMMARY

Chamber sampling is a common method for measuring nitrous oxide (N<sub>2</sub>O) emissions from agricultural soils. However, for grazed pastures, the patchy nature of urine deposition results in very high levels of spatial variability in N<sub>2</sub>O emissions. In the present study, the behaviour of the sample mean was examined by simulating a large number (9999) of random N<sub>2</sub>O chamber samples under different assumptions regarding the underlying N<sub>2</sub>O distribution. Using sample sizes of up to 100 chambers, the Central Limit Theorem did not apply. The distribution of the sample mean was always right-skewed with a standard deviation varying between 12.5 and 135% of the true mean. However, the arithmetic mean was an unbiased estimator and the mean of the sample mean distribution was close to the true mean of the simulated N<sub>2</sub>O distribution. The properties of the sample mean distribution (variance, skewness) were affected significantly by the assumed distribution of the emission factor, but not by distribution of the urine patch concentration. The geometric mean was also investigated as a potential alternative estimator. However, although its distribution had lower variance, it was also biased. Two methods for bias correcting the mean were investigated. These methods reduced the bias, but at the cost of increasing the variance. Neither of the bias-corrected estimators were consistently better than the arithmetic mean in terms of skewness and variance. To improve the estimation of N<sub>2</sub>O emissions from a grazed pasture using chambers, techniques need to be developed to identify urine patch and non-urine patch areas before sampling.

## INTRODUCTION

In New Zealand, nitrous oxide (N<sub>2</sub>O) emissions from agricultural soils account for 13.6% of anthropogenic greenhouse gas emissions on a CO<sub>2</sub>-equivalent basis (Ministry for the Environment 2014). Of these, 56.3% are direct emissions from grazing animals' excreta, 18.4% direct emissions from other sources of nitrogen (N) to soil (e.g. fertilizer, animal waste and crop residues) and 25.4% from indirect emissions. The present study focuses on direct emissions from urine patches, as the emissions from dung are comparatively small (van der Weerden *et al.* 2011).

Nitrous oxide emissions have a high degree of spatial and temporal variability due to the complexity of the various interacting processes and variability of the

driving factors (Mosier *et al.* 1996). A number of different factors, such as topography and soil properties, have been linked to spatial variability of N<sub>2</sub>O emissions (Parkin 1987; Pennock *et al.* 1992; Velthof *et al.* 1996, 2000; Ball *et al.* 1997; Grant & Pattey 2003; Yanai *et al.* 2003; Turner *et al.* 2008). In grazed pasture systems there is additional spatial variability due to the highly non-uniform nature of urine deposits (i.e. small areas of high N-concentration) (Bolan *et al.* 2004). Previous studies have looked at the distribution of urine patches both experimentally (Moir *et al.* 2011) and theoretically (Pleasant *et al.* 2007); however, these studies did not investigate N<sub>2</sub>O emissions. Giltrap *et al.* (2014) investigated the spatial variability of N<sub>2</sub>O emissions within a field and found differences in the emission response to applied urine at different locations.

Closed chambers are a relatively inexpensive method for measuring N<sub>2</sub>O emissions. In this method

\* To whom all correspondence should be addressed. Email: GiltrapD@landcareresearch.co.nz

a number of chambers are inserted into the soil. The tops are then sealed to form closed chambers and the air within the headspaces sampled at regular intervals. The  $\text{N}_2\text{O}$  flux is calculated from the change in  $\text{N}_2\text{O}$  concentration over time. Sealing and sampling of the chambers can be automated or performed manually. Manual chambers in particular are relatively cheap to deploy, although the labour requirements increase as more chambers are used. International researchers have published guidelines on the use of chambers for  $\text{N}_2\text{O}$  measurements (de Klein & Harvey 2012). These guidelines do not specify a minimum number of chambers, but do state that chambers should cover the largest total area practicable while providing information at the smallest scale required.

Integrative, micrometeorological methods are an alternative to chamber measurements. These methods measure emissions over a wider spatial scale and are therefore less affected by spatial variability; however, there can be limitations due to weather conditions and topography. Also atmospheric  $\text{N}_2\text{O}$  concentrations tend to be quite low, and may be close to the detection limit of many sensors.

Nitrous oxide emissions are frequently expressed in terms of an emission factor (EF). That is, the amount of N lost as  $\text{N}_2\text{O}$  (in excess of any background emission rate) as a fraction of the amount of N input. However, EF values vary with soil properties, weather conditions and other factors. Almost 90% of the uncertainty in New Zealand's inventory value for  $\text{N}_2\text{O}$  emissions from agricultural soil in 2002 (the relative error of the 95% confidence interval is -42% to +74%) was attributed to uncertainty in the value of EF (Ministry for the Environment 2014).

Studies for determining EF for animal urine in New Zealand used closed chambers to which known amounts of urine-N were applied (de Klein *et al.* 2003). This eliminated the variability due to uneven urine deposition, although there was still high variability in  $\text{N}_2\text{O}$  emissions due to variability in soil properties (CV ranged from 22 to 55% for different sites). However, these types of experiments may not be completely representative of what occurs during an actual grazing event (e.g. effects such as compaction caused by animal treading and differences in urine concentration between animals are neglected). Some studies have been conducted to measure  $\text{N}_2\text{O}$  emissions from actual grazed pastures (Saggar *et al.* 2004, 2007; Hoefft *et al.* 2012; Luo *et al.* 2013). These studies generally (although not always) used a larger number of chambers than the applied urine studies to account

for the additional spatial variability of urine patch distribution. The question is how well can such a sample estimate the mean emissions from a field?

Many common statistical techniques assume that the data come from a normal distribution. This is clearly not the case for sampling a grazed pasture as there is both the probability of any given chamber landing on a urine patch combined with the variability of emissions from a urine patch. Neither is there an obvious transformation to normalize the data. However, according to the Central Limit Theorem (CLT), the arithmetic mean of a number ( $n$ ) of independent random variables, each with a finite mean ( $\mu$ ) and variance ( $\sigma^2$ ), will be approximately normally distributed with mean  $=\mu$  and variance  $=\sigma^2/n$  as  $n \rightarrow \infty$  (Bain & Engelhardt 1992).

In the case of  $\text{N}_2\text{O}$  emissions from grazed pastures, the distribution is highly skewed and it cannot be taken for granted that the CLT can be applied for practical sample sizes. To estimate the probability density function of the sampling distribution, computer simulations were used to generate a large number of samples from the underlying  $\text{N}_2\text{O}$  emission distribution using a number of different sample sizes and assumptions about the properties of the underlying distribution. As the number of samples taken was large, the distribution of these samples should closely approximate the probability density function of the sampling distribution, and this distribution shall be referred to as the empirical sampling distribution (ESD). The properties of the ESDs were then examined to observe how they changed with different sample sizes and with changes to the underlying emission distribution. The arithmetic mean of each sample was initially used as the estimator of the mean of the underlying  $\text{N}_2\text{O}$  emissions; some alternative estimators were then examined for comparison.

While the current research has been framed in terms of  $\text{N}_2\text{O}$  measurements, the results may also be applicable to other situations where point measurements are made on a grazed field and the results are likely to be influenced by the presence or absence of a urine patch (e.g. nitrate leaching).

## MATERIALS AND METHODS

### Nitrous oxide emission model

For the simulations it was assumed that a number of chambers,  $n$ , were used to estimate the  $\text{N}_2\text{O}$  emissions from a field following a grazing event. Each chamber had a probability,  $p$ , of landing on a urine patch,

Table 1. *Parameters varied to investigate the effect on the sample mean distribution*

Parameter	Values
Chambers per sample ( <i>n</i> )	4, 8, 16, 20, 100
Proportion of field covered in urine patches ( <i>p</i> )	0.025, 0.05, 0.10, 0.15
Distribution of urine concentration ( <i>U</i> <sub>DIST</sub> )	Constant, normal
Distribution of EF ( <i>EF</i> <sub>DIST</sub> )	Constant, normal, gradient, log-normal

where *p* represents the proportion of the field covered by urine patches. This assumed that the total area of chambers was sufficiently small that *p* could be considered the same for all chambers. The possibility of overlapping urine patches or chambers having only part of their area on a urine patch was not considered.

As the present study was concerned with spatial rather than temporal variation, it was assumed that each chamber measured the total N<sub>2</sub>O emitted over time accurately and that the duration of the measurement campaign was long enough to capture all the additional N<sub>2</sub>O emissions due to the grazing event.

The following equations describe the N<sub>2</sub>O emissions measured in a single chamber:

$$\text{N}_2\text{O-N} = B, \quad (1)$$

if the chamber was not on a urine patch

$$\text{N}_2\text{O-N} = B + EF \times U, \quad (2)$$

if the chamber was on a urine patch

where N<sub>2</sub>O–N is the total N emitted as N<sub>2</sub>O (kg N/ha), *B* is the background N<sub>2</sub>O emissions over the measurement period (kg N/ha), *EF* is the emission factor (which is specific to a given location and time period, in kg/kg) and *U* is the N concentration of the urine patch (kg N/ha).

This formulation assumed that *EF* and *U* were independent. Some studies have suggested that *EF* may increase with increasing N concentration (Kim *et al.* 2013); however, de Klein *et al.* (2014) suggested that this dependency only occurs at low values, while Kelliher *et al.* (2014) found no dependency.

### Simulations

Empirical sample distributions were generated by simulating 9999 random samples in R (R Foundation for Statistical Computing, <http://www.r-project.org/>) and were generated for different sample sizes, urine coverages and assumptions about the distribution of *U* and *EF*. Table 1 shows the parameters investigated

and the range of values considered. The maximum value of *n* considered was 100 chambers, which is higher than generally used in practice but still potentially feasible. Increasing the maximum number of chambers will eventually result in the sample mean distribution approaching a normal distribution (by the CLT). However, in this case it would no longer be necessary to use the methods described in the present study.

The background emissions (*B*) were considered to have a constant value of 1 kg N/ha. While there is likely to be some variability in background emissions, this was not considered likely to be a major contributor to overall spatial variability.

Different underlying distributions were investigated for the urine concentration and *EF*. For urine, only the constant and normal distributions were considered, whereas for the *EF*, gradient and log-normal distributions were also examined.

In all cases the distributions were parameterized to have the same expected value (0.01 for *EF* and 1000 kg N/ha for urine concentration). The expected value of 0.01 for *EF* corresponds to the New Zealand specific value used in the national inventory for calculating N<sub>2</sub>O emissions from urine applied to soil (Ministry for the Environment 2014). However, measured *EF* values have a high degree of variability with a positive skew (de Klein *et al.* 2003). Therefore, for the normal distribution a standard deviation of 0.005 was selected to give the largest variability possible while keeping the *EF* positive within two standard deviations of the mean.

Few studies have measured the N loading in urine patches in terms of g N/ha (Selbie *et al.* 2015). Haynes & Williams (1993) give a range of 8–15 g/l for urine N concentration, with urination volumes in the range 1.6–2.2 l and patch size 0.16–0.49 m<sup>2</sup> for dairy cattle. The expected value of 1000 kg N/ha for urine patch N concentration in the present study is consistent with this and is a value often used in artificial urine patch experiments (e.g. Luo *et al.* 2008). The distributions considered are described briefly below:

*Constant distribution (Const)* – assumes there is no variability in the underlying quantity. While this is unrealistic, it may be a reasonable approximation when the variability of the quantity is negligible compared with other sources of variability. For the urine concentration the constant value was 1000 kg N/ha, while for EF the constant value was 0.01.

*Normal distribution (Norm)* – commonly occurs in nature and many standard statistical methods assume a normal distribution. For the urine concentration the normal distribution was assumed to have mean = 1000 kg N/ha and standard deviation = 200 kg N/ha. For the EF distribution the mean was 0.01 and the standard deviation 0.005.

*Gradient distribution (Grad)* – this distribution simulates the case where the variation in EF is due to correlation with some property (e.g. soil moisture) that varies linearly across the field. Sampling a random location is therefore equivalent to drawing a sample from a uniform distribution. In these simulations, the minimum and maximum EF values were 0 and 0.02, respectively.

*Log-normal distribution (LNorm)* – the skew in the chamber samples may not be due solely to the patchy nature of the urine deposits, but may also reflect a skewness in the EF distribution. Studies looking at N<sub>2</sub>O emissions from known applications of urine N often find positively skewed results, suggesting this may be the case (e.g. van der Weerden *et al.* 2011). Log-normal distributions frequently occur when several factors act in a multiplicative fashion. For example, N<sub>2</sub>O emission rates could be affected by soil moisture, temperature and availability of substrates. The expected value of the log-normal distribution is given by:

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad (3)$$

where  $\mu$  is the log-mean and  $\sigma$  is the log-standard deviation. For the log-normal EF simulations the log-mean was set to  $-5.105$  and the log-standard deviation to 1. There are an infinite number of combinations of  $\mu$  and  $\sigma$  that could be selected that satisfy  $E(X) = 0.01$ . The particular combination selected here resulted in a 95% confidence interval for EF of 0.00085–0.043. Note that due to rounding, the expected value of this distribution was 0.017% higher than the other distributions.

## Properties of the empirical distribution

Nitrous oxide emissions from grazed pastures are highly skewed and it cannot be taken for granted that the CLT can be applied for practical sample sizes. Empirical sample distributions were generated for a number of estimators (arithmetic mean, geometric mean and two methods of biased-corrected geometric mean of a sample) using different sample sizes and assumptions about the distributions of EF and  $U$  (e.g. constant, normal, gradient and log-normal). The influence of these changes on the properties of the sampling distribution was then investigated. The properties of the ESDs investigated were the mean, variance, relative bias ( $r\text{Bias}$ ), probability that given a sample was an underestimate of the true mean ( $p\text{Under}$ ) and skewness.

The true mean of the N<sub>2</sub>O emissions distribution is the expected value given by:

$$E(\text{N}_2\text{O}) = E(B + p\text{EF} \times U) = B + pE(\text{EF})E(U) \quad (4)$$

assuming that the EF is independent of the urine concentration. All the distributions examined were parameterized, so that  $E(\text{EF}) = 0.01$  and  $E(U) = 1000$  kg N/ha. Therefore, for an unbiased estimator the mean of the ESD should be  $1 + 10p$ .

When the CLT applies, the arithmetic mean of a sample should have variance =  $\sigma^2/n$ . However, even when the CLT does not apply it is desirable for an estimator to have a low variance.

$r\text{Bias}$  was defined as:

$$r\text{Bias} = (\text{Mean}(\text{ESD}) - E(\text{N}_2\text{O}))/E(\text{N}_2\text{O}) \quad (5)$$

where  $\text{Mean}(\text{ESD})$  is the mean of the empirical sampling distribution and  $E(\text{N}_2\text{O}) = 1 + 10p$  as shown above.

The  $p\text{Under}$  is the proportion of samples that underestimate the true mean. In an unbiased, symmetrical distribution this should be 0.50. A  $p\text{Under}$  that is much higher or lower than 0.50 indicates either asymmetry (skew) and/or bias. Note that in the absence of any variability in the urine concentration or EF,  $p\text{Under}$  could be calculated using the binomial distribution.

Finally, skewness is defined as the third standardized moment of a distribution. Skewness should be 0 for a symmetric distribution (such as the Normal), positive or negative values indicate positive (right) or negative (left) skew, respectively.

It should be noted that a skewed distribution is not the same as a biased distribution. A biased distribution is one in which the expected value (average over a large number of samples) differs from the true mean. However, a skewed distribution could still produce



the correct average over a large number of samples, even though there is a high probability of any individual sample being an under- (or over-) estimate.

### Modelling the distribution properties

Having generated ESDs, the next step was to model how the properties of the sample distributions varied with sample size and with the properties of the underlying population distribution. The obvious explanatory variables were  $p$  and  $n$ , as well as the distributions assumed for the urine patch concentration ( $U_{\text{dist}}$ ) factors and the emission factor ( $EF_{\text{dist}}$ ). However, some non-linear terms also suggested themselves. The sampling distribution was related to the binomial distribution as the number of urine patches sampled would follow a binomial distribution. The variance of a binomial distribution is  $npq = np(1 - p) = np - np^2$ . Therefore  $np$  and  $p^2$  were included. When the CLT applies the variance of the sample mean depends on  $1/n$ , so  $1/n$  was added to the list of explanatory variables. The terms  $n^2$ ,  $1/p$ ,  $n/p$  and  $p/n$  were also considered.

For each variable the interactions with  $EF_{\text{dist}}$ ,  $U_{\text{dist}}$  and the  $EF_{\text{dist}} : U_{\text{dist}}$  interaction were considered. Thus the full model was:

$$\text{Property} \sim EF_{\text{dist}} * U_{\text{dist}} * (n + p + np + n^2 + p^2 + 1/n + 1/p + p/n + n/p) \quad (6)$$

where '\*' indicates that both main and interaction effects were considered.

The best model was initially selected using stepwise selection (with the Akaike Information Criterion [AIC] as the selection criterion), in both the forward and backward directions. However, in many cases the selected model still contained a large number of explanatory variables. These models were replaced with simplified models when it could be demonstrated that explanatory variables could be dropped while retaining satisfactory model performance (i.e. good explanatory power, lack of structure and constant variance in the residuals).

A standard linear model was applied for all properties except  $p_{\text{Under}}$ , for which a binomial generalized linear model was used.

### Alternative estimators

Nitrous oxide emissions from grazed pastures are highly skewed and therefore sample sizes will probably need to be large before the CLT can be applied.

However, it is sometimes possible to use data transformations to improve the normality of the data. In particular, log-transformations are commonly used to reduce positive skew.

The geometric mean ( $g$ ) is the mathematical equivalent of back-transforming the mean of the log-transformed data and is given by:

$$g = \left( \prod_{i=1}^n x_i \right)^{1/n} = \exp \left( \frac{\sum_{i=1}^n \log(x_i)}{n} \right) \quad (7)$$

where  $x_i$  is the  $i$ th observation and  $n$  is the number of observations.

Therefore,  $g$  could potentially be used as an alternative estimator to the arithmetic mean ( $a$ ). Unfortunately  $g$  is a biased estimator, that is, its expected value is less than the true mean of the N<sub>2</sub>O distribution.

Assuming that the underlying N<sub>2</sub>O distribution was actually log-normal, then from Eqn (3) a reasonable correction factor might be  $\exp(\sigma^2/2)$ . Therefore the second alternative estimator investigated was:

$$g_{c1} = g \times \exp(s^2/2) \quad (8)$$

where  $s^2$  is the sample variance.

Neyman & Scott (1960) gave bias-correction formulae for a number of variable transformations. However, for the log-transformation this requires solving a Bessel function of an imaginary argument. Given that the N<sub>2</sub>O distribution is not precisely log-normal a slightly simpler formula was used to calculate the correction factor:  $\exp((1-1/n)s^2/2)$ . When the Taylor expansion is applied to this correction factor it only differs from that of Neyman & Scott (1960) for the third and higher terms. However, it does have the property that it approaches  $\exp(s^2/2)$  as  $n \rightarrow \infty$ . Therefore the final estimator considered was:

$$g_{c2} = g \times \exp \left( \left( 1 - \frac{1}{n} \right) \frac{s^2}{2} \right) \quad (9)$$

All simulations and analyses were performed using the R statistical software package version 3.0.2

## RESULTS

### Sample mean

Using the arithmetic mean ( $a$ ) as the estimator for each simulated N<sub>2</sub>O sample and then fitting the sample mean (the mean of the 9999 randomly generated samples) to the proportion of urine coverage gave

the following relationship:

$$\begin{aligned}\text{Mean}(a) &= 1.01(\pm 0.001) + 9.9(\pm 0.01)p, \\ R_{\text{adj}}^2 &= 0.997\end{aligned}\quad (10)$$

where  $\text{Mean}(a)$  is the sample mean (using the arithmetic mean as the estimator for each sample).

This agrees fairly well with the true mean as expressed in Eqn (4), indicating that the arithmetic mean is largely unbiased irrespective of the number of chambers used or the underlying distributions of the urine concentration or EF.

Figure 1 shows the modelled means (values calculated from Eqn (10)) plotted against the observed values (simulated sample means) as well as the residuals plotted against fitted values,  $n$  and  $p$ . There was an obvious decrease in the variability of the residuals as  $n$  increased. It was also apparent that the largest residuals correspond to the log-normal EF distribution. However, the largest relative residual was still quite small at only 3% of the fitted value.

## Variance

Figure 2 shows a plot of  $\text{Var}(a) \times n$  v.  $n$ . When the CLT applies  $\text{Var}(a) \rightarrow \sigma^2/n$ , in which case  $\text{Var}(a) \times n$  should be constant with respect to  $n$ . From Fig. 2(a) this appears to be the case for all but the log-normal EF distribution. The log-normal EF distribution shows variability in  $\text{Var}(a) \times n$  for different  $n$  (but not in any easily to discernible pattern, Fig. 2(b)).

The variance could be quite well explained by a model involving  $p/n$ ,  $\text{EF}_{\text{DIST}}$  and their interaction (Eqn (11)).

$$\begin{aligned}\text{Var}(a) &\sim p/n + \text{EF}_{\text{DIST}} + \text{EF}_{\text{DIST}} : p/n, \\ R_{\text{adj}}^2 &= 0.9753\end{aligned}\quad (11)$$

	Coefficients ( $\pm$ S.E.)
Intercept	$(2 \pm 4.6) \times 10^{-2}$
$\text{EF}_{\text{dist}} = \text{Norm}$	$(1 \pm 6.5) \times 10^{-2}$
$\text{EF}_{\text{dist}} = \text{Grad}$	$(6 \pm 65.4) \times 10^{-3}$
$\text{EF}_{\text{dist}} = \text{LNorm}$	$(5 \pm 6.5) \times 10^{-2}$
$p/n$	$90 \pm 4$ ( $P < 0.001$ )
$p/n : \text{EF}_{\text{dist}} = \text{Norm}$	$23 \pm 5.6$ ( $P < 0.001$ )
$p/n : \text{EF}_{\text{dist}} = \text{Grad}$	$33 \pm 5.6$ ( $P < 0.001$ )
$p/n : \text{EF}_{\text{dist}} = \text{LNorm}$	$169 \pm 5.6$ ( $P < 0.001$ )
Note ':' indicates an interaction term	

According to Eqn (11) the variance depends on the EF distribution and increases with  $p/n$ . The model is a reasonable approximation for the constant, normal and gradient EF distributions, but there are some large residuals for the log-normal EF distribution (Fig. 3). The variance was always highest for the log-normal EF distribution. For low  $p/n$  values the constant, normal, gradient EF distributions have similar variances. As  $p/n$  increases the variances of the different EF distributions diverge.

Equation (11) is useful for determining the level of precision with which the mean can be determined. For example, suppose  $P = 0.05$ ,  $n = 20$ . The variance of the sample mean would be 0.24 for a constant EF, corresponding to a standard deviation of 0.49 kg N/ha or 33% of the true mean. If the EF was normally distributed instead, then the standard deviation would be 38% of the true mean.

## Relationship between mean and variance

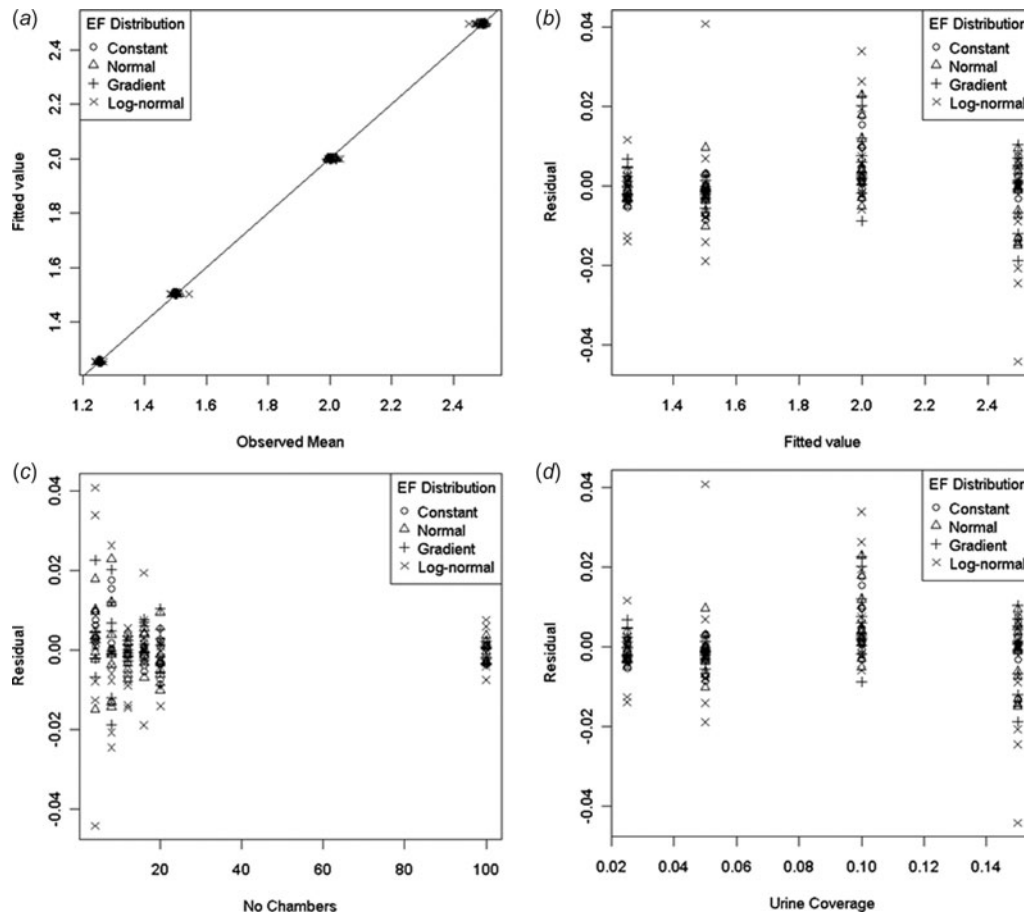
The mean and variance appear to behave in a manner consistent with the CLT (with the exception of the variance when using the log-normal EF distribution). However, another requirement for a normal distribution is that the variance and mean should be independent. This is also an important assumption of linear regression.

From inspection of models (10) and (11) it can be seen that both the mean and variance depend on  $p$ . Therefore, it seems likely that the mean and variance will be correlated. This is confirmed in Fig. 4. However, the slope of the relationship between variance and mean decreases as  $n$  increases, and for  $n = 100$  the variance is almost constant with respect to the mean.

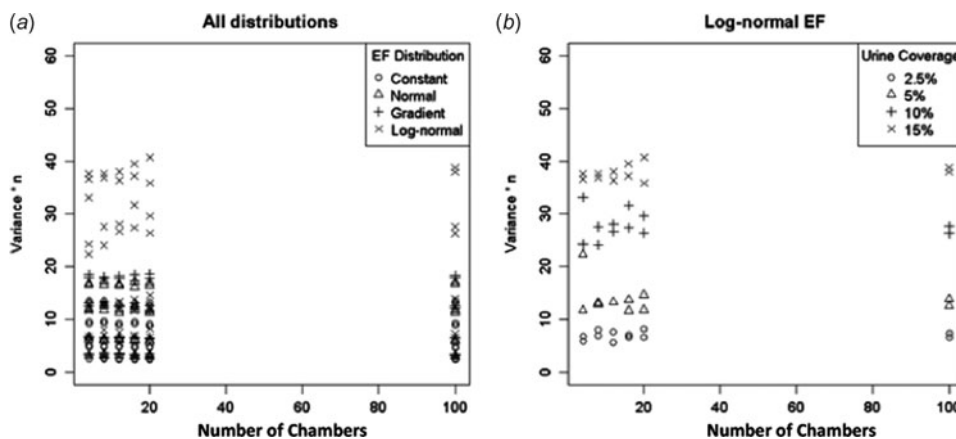
The significance of the slopes of the relationships between variance and mean were determined by fitting a linear regression model of the variance of the ESD. The predictors were the mean of the ESD, a 24-level factor ( $\text{EF}_n$ ) with each level corresponding to a combination of EF distribution and number of chambers, and their interactions (Eqn (12)). The following relationship was fitted:

$$\text{Var}(a) \sim \text{EF}_n + \text{EF}_n : \text{Mean}(a) - 1 \quad (12)$$

With the exception of the levels corresponding to  $n = 100$ , all the coefficients for the  $\text{EF}_n$  interaction with the ESD mean were significantly different from 0 (at a 0.01 confidence level). None of the  $\text{EF}_n : \text{Mean}$  interaction coefficients were significant for  $n = 100$



**Fig. 1.** Plots comparing values fitted using model (10) with the observed mean of the ESD for *a* (arithmetic mean of a sample): (a) Fitted value *v.* observed mean, (b) residuals *v.* fitted values, (c) residuals *v.* number of chambers (*n*), (d) residuals *v.* urine coverage (*p*).



**Fig. 2.** The variance of the ESD for  $a \times n$  (arithmetic mean times the number of chambers) plotted against the number of chambers (should be constant if CLT applies). (a) All EF distributions and (b) log-normal EF only.

(at a 0.1 confidence level). Therefore the variance of the sample mean depends on the sample mean for the chamber numbers used in practice. However, at

100 chambers the variance was independent of the mean. This relationship held across all EF distributions.

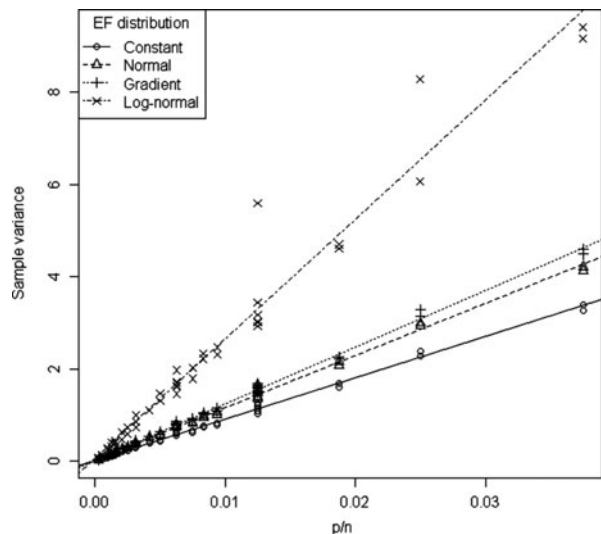


Fig. 3.  $\text{Var}(a)$  v.  $p/n$ . Straight lines indicate model (11) estimates.

As a consequence of this dependence of the variance on the mean (for  $n < 100$ ) ordinary linear regression methods (including ANOVA) should not be performed on samples of this type. However, this can be corrected using a weighted regression provided the relationship between the mean and the variance is known. If the variance can be written as:

$$\text{Var}(Y_i) = \sigma^2 f(\mu_i) \quad (13)$$

then the weighted regression can be performed using  $1/f(\mu_i)$  as weights. There are several options for  $f(\mu_i)$ . Using model (12) gives  $f(\mu_i) = \bar{x}$ . This assumes the same  $n$  and EF for all samples and will be slightly inaccurate due to neglecting the constant term. If the EF distribution is known, then the more precise form of  $f(\mu_i) = a + b\bar{x}$  can be used. Alternatively, model (11) gives  $f(\mu_i) = p/n$ . The constant terms in model (11) are not statistically significant, so can justifiably be neglected. If the urine coverage can be assumed to be constant, then it can be incorporated into  $\sigma^2$  leaving  $f(\mu_i) = 1/n$ .

### Relative bias

Earlier it was found that the mean of the ESDs were very close to the theoretical population means (Eqn (10)). Therefore bias should be small. Figure 5 shows the bias (relative to the mean) plotted against  $p$  and  $n$ .

It is difficult to discern any patterns from these plots, although the variability in the  $r\text{Bias}$  reduces with increasing  $n$ . However, the most extreme  $r\text{Bias}$  values

are only  $-2\%$  and  $+3\%$ , both of which occur for log-normal distributions.

Bias becomes important when it is high relative to the variability. One measure of the relative importance of the bias would be to take the ratio of the bias to the standard deviation. This is equivalent to the ratio of the  $r\text{Bias}$  to the CV. Figure 6 shows that the magnitude of the  $r\text{Bias}$  is at most  $2\%$  of the CV. That is, the absolute bias is at most  $2\%$  of the standard deviation of the sample mean and therefore not a major concern.

Probability that a given sample was an underestimate of the true mean ( $p\text{Under}$ )

With the exception of some constant EF distribution points,  $p\text{Under}$  was  $> 0.5$  (Fig. 7). Both  $p\text{Under}$  and the spread of  $p\text{Under}$  decreased with increasing number of chambers and urine coverage. Even with 100 chambers the probability of underestimating the mean can be much larger than  $0.5$ . Given that no evidence of bias was found in the sample mean distribution, these high  $p\text{Under}$  values indicate a right-skewed distribution.

$p\text{Under}$  is a probability and therefore bounded by 0 and 1. Therefore a binomial generalized linear model with a logit† link function (Eqn (14)) was used for modelling:

$$\text{logit}(p\text{Under}(a)) \sim 1/n \pm p/n + \text{EF}_{\text{dist}} \quad (83\% \text{ Deviance explained}) \quad (14)$$

	Coefficient ( $\pm$ S.E.)
Intercept	$-0.087 \pm 0.0036$ ( $P < 0.001$ )
$1/n$	$9.49 \pm 0.036$ ( $P < 0.001$ )
$p/n$	$-57.5 \pm 0.29$ ( $P < 0.001$ )
$\text{EF}_{\text{dist}} = \text{Norm}$	$0.177 \pm 0.0043$ ( $P < 0.001$ )
$\text{EF}_{\text{dist}} = \text{Grad}$	$0.219 \pm 0.0043$ ( $P < 0.001$ )
$\text{EF}_{\text{dist}} = \text{LNorm}$	$0.519 \pm 0.0044$ ( $P < 0.001$ )

Model (14) produced a reasonable fit for all but the constant EF distribution (Fig. 8). As an example, a sample from a normal EF distribution with  $n=20$  and  $P=0.05$  would have log-odds of  $0.42$  of being an underestimate. This corresponds to a  $p\text{Under}$  of  $60\%$ , indicating a slight right-skew in the sample mean distribution.

†  $\text{Logit}(p) = \ln(p/(1-p))$ .



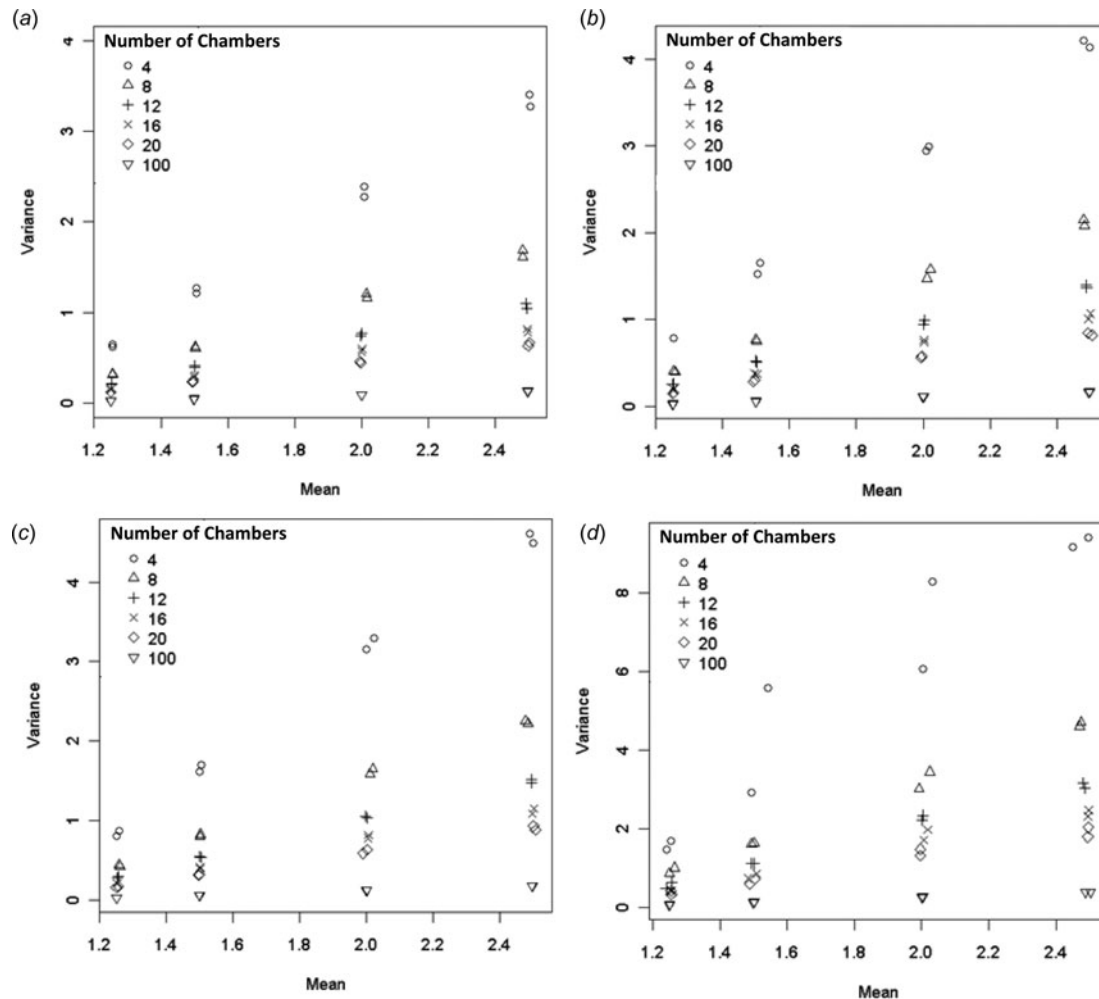


Fig. 4. Var(a) v. mean(a) for (a) constant, (b) normal, (c) gradient and (d) log-normal EF distributions.

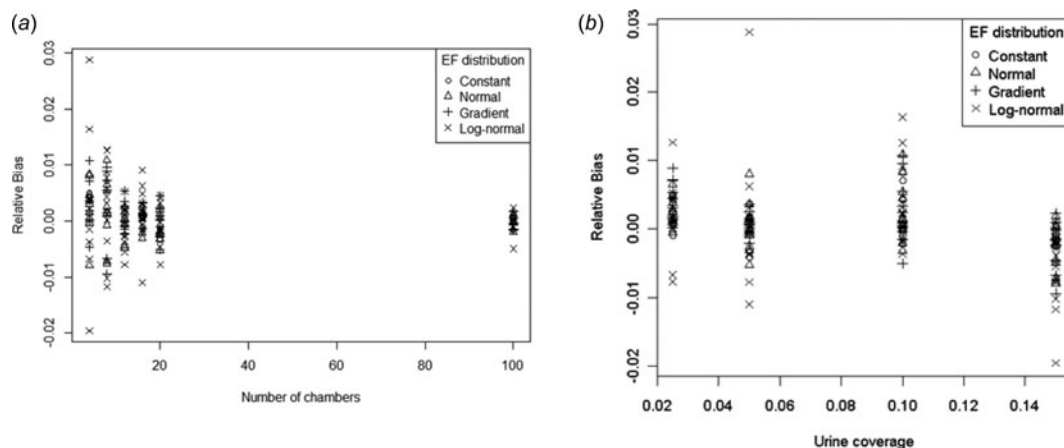


Fig. 5. Relative bias v. (a) number of chambers and (b) urine coverage.

### Skewness

Skewness is the third standardized moment and relates to the lack of symmetry in a distribution. Skewness

should be 0 for a symmetric distribution (such as the normal). Figure 9 shows the observed skewness of the sample mean distribution plotted as a function of

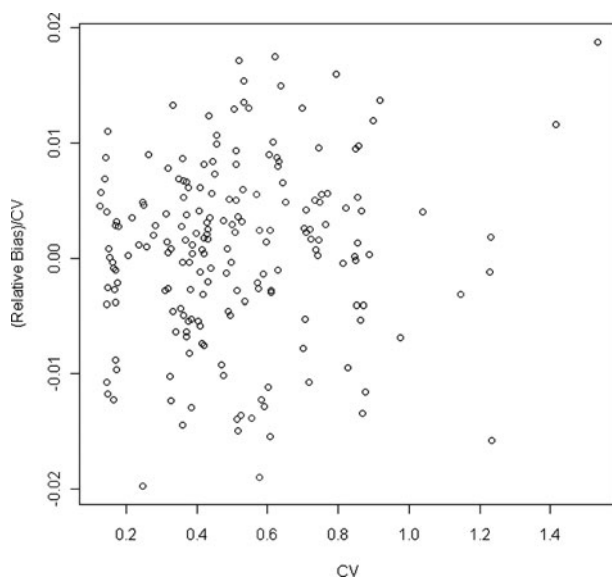


Fig. 6. Ratio of relative bias to CV plotted against CV.

number of chambers and urine coverage. Unsurprisingly, skewness was largest for the right-skewed log-normal EF distribution. Skewness reduced with increasing number of chambers, but remained positive over the range of conditions tested. There did not seem to be an obvious pattern to skewness with increasing urine coverage. Model (15) below produced a reasonable fit to the data, except for the log-normal EF distribution (Fig. 10.):

$$\begin{aligned} \text{Skewness}(a) &\sim \text{EF}_{\text{dist}} + 1/n + p/n + \text{EF}_{\text{dist}} \\ &\quad : 1/n + \text{EF}_{\text{dist}} : p/n, \\ R^2_{\text{adj}} &= 0.8075 \end{aligned} \quad (15)$$

	Coefficients ( $\pm$ S.E.)
Intercept	$0.5 \pm 0.25$ ( $P < 0.05$ )
$1/n$	$12 \pm 3.0$ ( $P < 0.001$ )
$p/n$	$-76 \pm 26.4$ ( $P < 0.01$ )
$\text{EF}_{\text{dist}} = \text{Norm}$	$0.2 \pm 0.36$
$\text{EF}_{\text{dist}} = \text{Grad}$	$0.2 \pm 0.36$
$\text{EF}_{\text{dist}} = \text{LNorm}$	$2.1 \pm 0.36$ ( $P < 0.001$ )
$1/n : \text{EF}_{\text{dist}} = \text{Norm}$	$3 \pm 4.2$
$1/n : \text{EF}_{\text{dist}} = \text{Grad}$	$3 \pm 4.2$
$1/n : \text{EF}_{\text{dist}} = \text{LNorm}$	$36 \pm 4.2$ ( $P < 0.001$ )
$p/n : \text{EF}_{\text{dist}} = \text{Norm}$	$-20 \pm 37$
$p/n : \text{EF}_{\text{dist}} = \text{Grad}$	$-20 \pm 37$
$p/n : \text{EF}_{\text{dist}} = \text{LNorm}$	$-190 \pm 37$ ( $P < 0.001$ )

In model (15) only the log-normal EF distribution terms were significantly different from the constant EF distribution. The variability of the residuals

increased with increasing skewness (Fig. 10.(b)). As the highest skewness occurred in the log-normal EF distribution, this distribution resulted in some large residuals and indicated that a better model of skewness may be required for log-normal EF distributions.

For the other EF distributions, model (15) indicates that for low values of  $p$  there will always be positive skewness regardless of the size of  $n$ . This threshold for  $p$  ranges from 0.168 to 0.178 depending on the EF distribution. This is higher than the maximum urine coverage considered in the present study (0.15), indicating that skewness cannot be eliminated for the range of urine coverages likely to be found in the field. Increasing  $p$  decreased skewness, but as  $n$  increases the skewness approaches a non-zero value (which depends on the EF distribution). By this model it is not possible to eliminate the skewness by increasing the sample size. While the CLT means the skewness should approach 0 for sufficiently high  $n$ , this must happen at  $n$  values well beyond those simulated in the present study, for which model (15) is valid.

Model (15) can be applied to constant, normal or uniform EF distributions, but is not recommended for log-normal EF-distributions. It also requires knowledge of the number of chambers and urine coverage. For 20 chambers and  $P = 0.05$ , a 20% increase in  $p$  resulted in a ~4% reduction in skewness, while a 20% increase in  $n$  resulted in a 7–8% decrease in skewness for the constant, normal and gradient EF distributions.

## ALTERNATIVE ESTIMATORS

### Mean

Fitting a linear regression of the alternative estimators on  $p$  gave the following relationships:

$$\begin{aligned} \text{Mean}(g) &= 0.984 (\pm 0.0051) + 3.02 (\pm 0.054)p, \\ R^2_{\text{adj}} &= 0.94 \end{aligned} \quad (16a)$$

$$\begin{aligned} \text{Mean}(g_{c1}) &= 0.94 (\pm 0.031) + 9.0 (\pm 0.33)p, \\ R^2_{\text{adj}} &= 0.80 \end{aligned} \quad (16b)$$

$$\begin{aligned} \text{Mean}(g_{c2}) &= 0.95 (\pm 0.016) + 8.0 (\pm 0.16)p, \\ R^2_{\text{adj}} &= 0.92 \end{aligned} \quad (16c)$$

The geometric mean seriously underestimates the  $\text{N}_2\text{O}$  emissions, with the coefficient of  $p$  being less

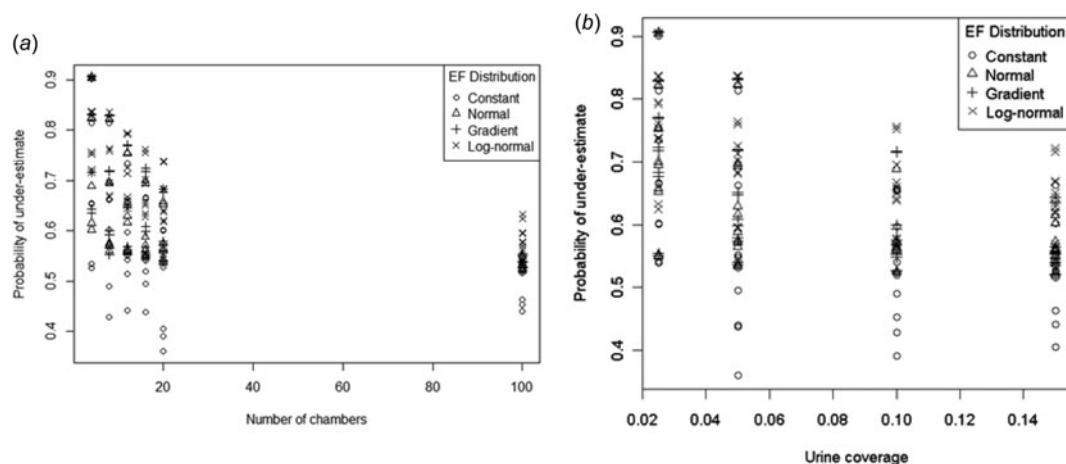


Fig. 7. Probability of a sample mean being less than the true mean by (a) number of chambers and (b) urine coverage.

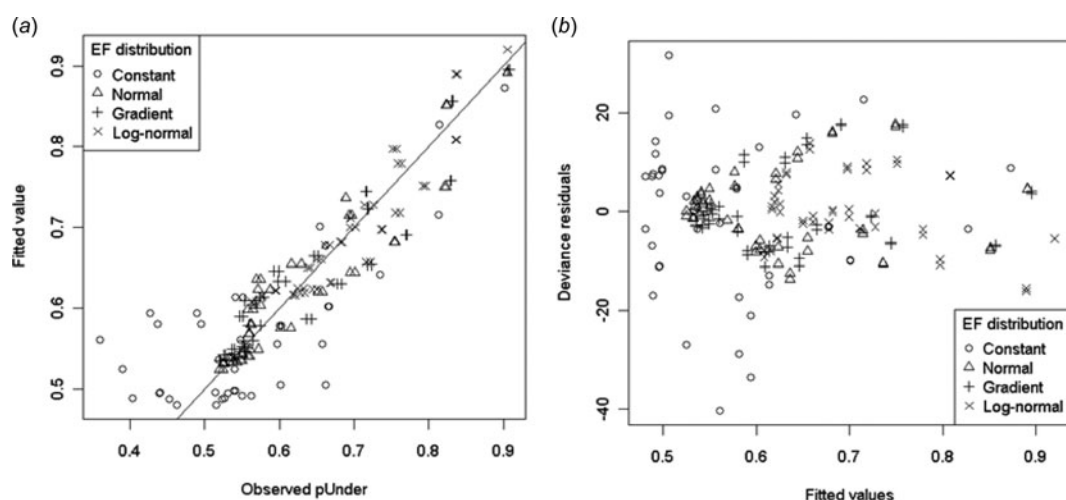


Fig. 8. (a) Modelled v. observed  $p_{\text{Under}}$  and (b) deviance residuals v. fitted values for model (14).

than one-third of what it should be. The first corrected geometric mean  $g_{c1}$  gets much closer to the true mean, but not quite as close as the arithmetic mean (Eqn (10)). The second corrected mean ( $g_{c2}$ ) also performs much better than the uncorrected  $g$  but still underestimates by more than  $g_{c1}$ .

For all estimators the variability increased as the mean increased (Fig. 11.).

#### $r\text{Bias}$

Figure 12 shows  $r\text{Bias}$  plotted against  $n$  and  $p$  for each of the three alternative estimators. The three estimators generally had negative bias, although  $g_{c1}$  actually had a strong positive bias for  $n = 4$ . The bias was strongest in the uncorrected geometric mean. The  $r\text{Bias}$  for  $g_{c2}$  was always more negative than  $g_{c1}$  but the difference

approached 0 as the number of chambers increased. However, in terms of the magnitude of the bias,  $g_{c2}$  was actually less biased than  $g_{c1}$  for  $n = 4$ , but more biased at higher  $n$ .  $r\text{Bias}$  decreased slightly (became more negative) then stabilized with increasing  $n$  for all three estimators.  $g$  showed a strong negative trend (i.e. became more biased) with increasing  $p$ , whereas the two bias-corrected geometric means had increasing variability in  $r\text{Bias}$  as urine coverage increased. The range of  $r\text{Bias}$  was from  $-13$  to  $-46\%$  for  $g$ ,  $-26$  to  $+29\%$  for  $g_{c1}$ , and  $-26$  to  $+1.3\%$  for  $g_{c2}$ . These are large compared with the relative biases of  $a$ .

#### Variance

The uncorrected geometric mean consistently had a lower variance than the arithmetic mean (variance

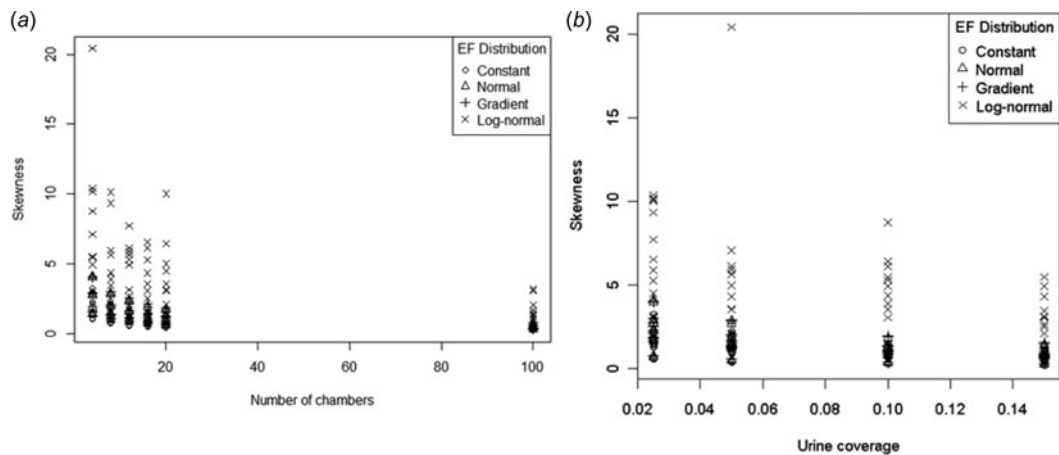


Fig. 9. ESD Skewness as a function of (a) number of chambers and (b) urine coverage.

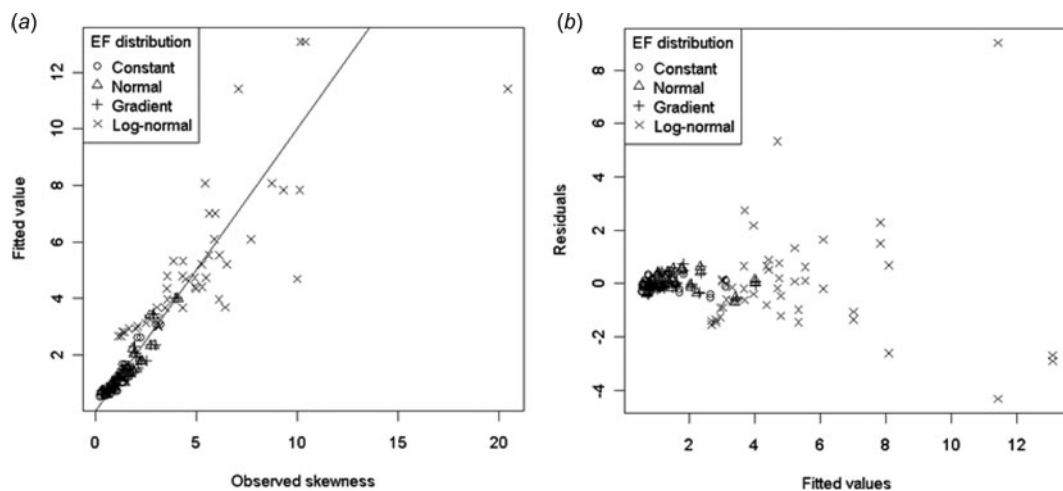


Fig. 10. (a) Modelled *v.* observed skewness and (b) residuals *v.* fitted values for model (15).

ratio  $< 1$ , Fig. 13.). Given that low variance is a desirable property, the relative advantage (from a variance perspective) of  $g$  relative to  $a$  increases as  $n$  increases but decreases with increasing  $p$ . In Fig. 2, it was seen that for most cases  $\text{Var}(a)$  was inversely proportional  $n$ . Therefore the variance of  $g$  must decrease even more rapidly with  $n$ .

The bias corrections for the geometric mean use the sample standard deviation of the log-transformed data to estimate the population log-standard deviation. Unfortunately, this increases the variance relative to the uncorrected geometric mean. For low sample sizes the corrected geometric means can sometimes have larger variances than the arithmetic mean. Both the corrected geometric means show similar trends in variance relative to  $\text{Var}(a)$  with increasing  $n$  and

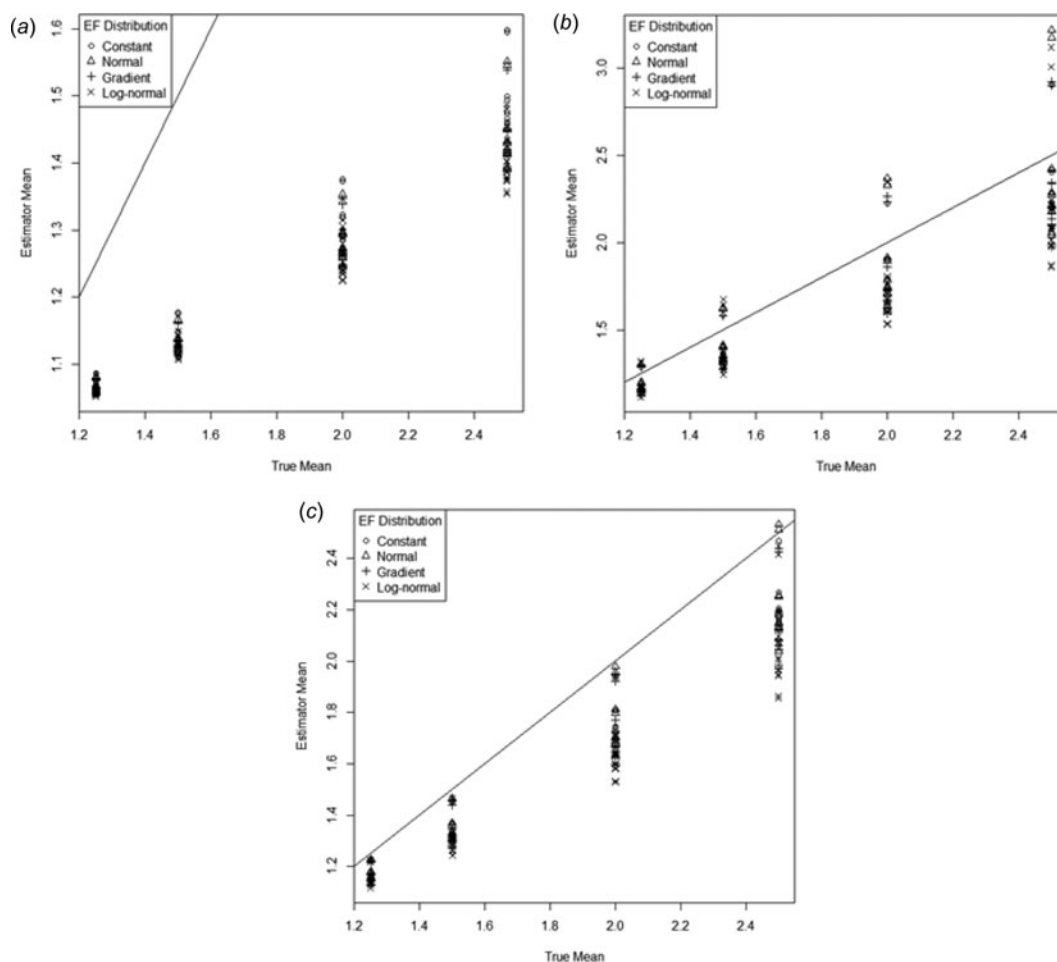
$p$ . The second bias correction method always produced a lower variance than the first, although the difference decreased as the number of chambers increased.

For all three alternative estimators the relative improvement in variance over the arithmetic mean was greatest when the EF distribution was log-normal.

### Skewness

None of the estimators consistently reduced (or increased) the skew relative to the arithmetic mean (Fig. 14.). In general, for log-normal EF the geometric mean (and the corrected versions) tended to have lower skew than the arithmetic mean. For the other





**Fig. 11.** Mean of the ESD plotted against the true mean for (a) geometric mean ( $g$ ), (b) first corrected geometric mean  $g_{c1}$  and (c) second corrected geometric mean  $g_{c2}$ . The 1–1 line is also indicated.

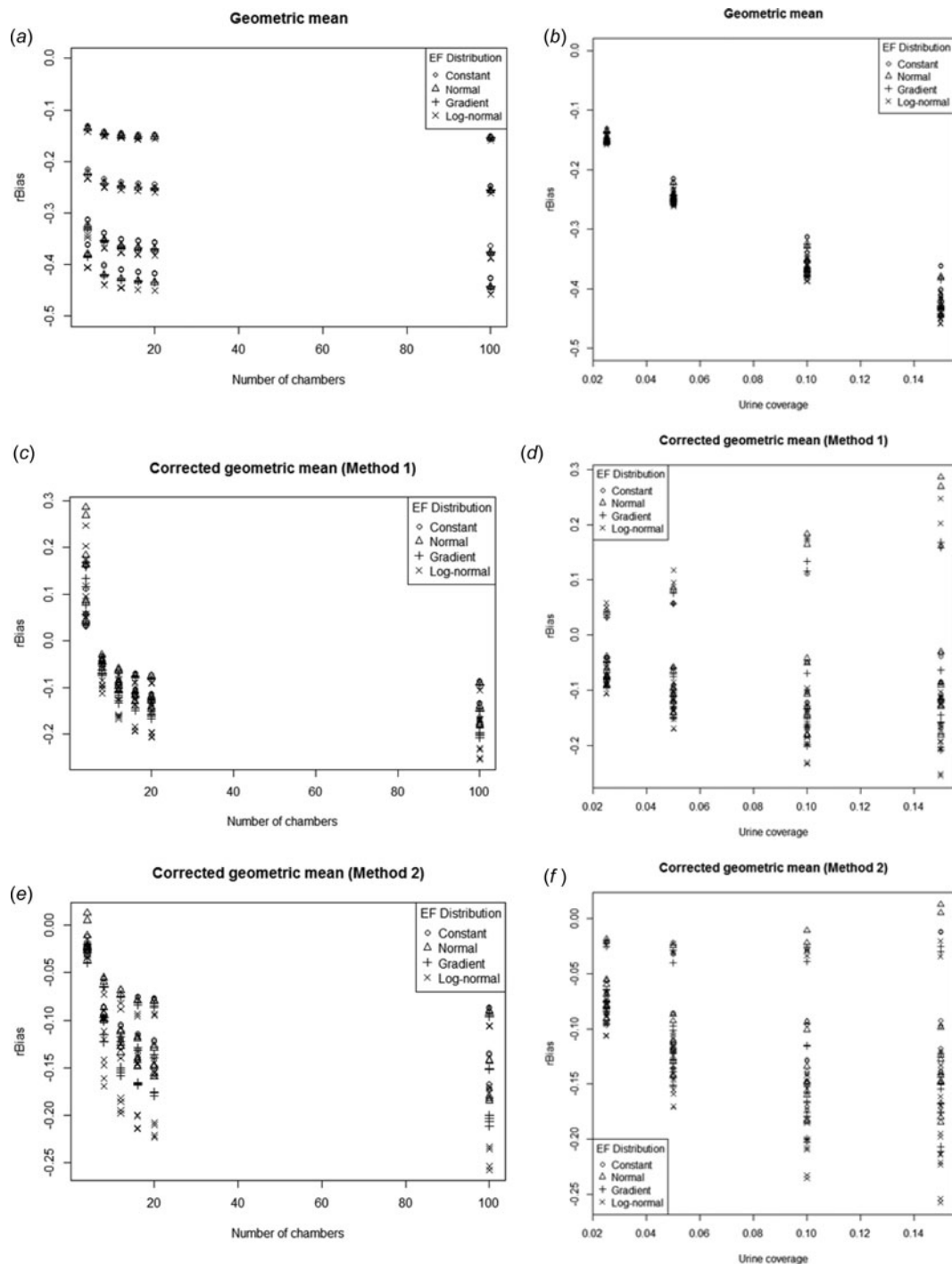
EF distributions the skew tended to be higher than the arithmetic mean.

## DISCUSSION

The arithmetic mean of chamber measurements was an unbiased estimator of the N<sub>2</sub>O emissions from a grazed pasture. However, the CLT did not apply over the range of conditions examined (even with 100 chambers) and the sample mean distribution was right-skewed. The variance was relatively high (with standard deviation ranging from 12 to 135% of the actual mean) but did appear to be proportional to  $1/n$  except when the EF distribution was log-normal. For  $n < 100$  the variance was proportional to the mean, therefore attempts to use regression modelling on chamber data should include weights to account for the non-constant variance. The lack of normality means that statistical tests that assume

normality of the sample mean (e.g.  $t$ -test and  $F$ -test) should not be used for this type of data. With knowledge of the urine coverage and EF distribution, equivalent tests could be created by generating empirical sampling distributions as we have done here. However these tests would probably have lower power due to the high probability of not sampling any urine patches.

Formulae were derived that gave reasonable estimates for variance,  $p_{\text{Under}}$  and skewness of the sample distribution of the arithmetic mean. These formulae required knowledge of the number of chambers, the urine coverage and the EF distribution. Whether the urine concentration was constant or normally distributed had little impact on the sample mean distribution. An experimenter will know the number of chambers being used, but will not know the urine coverage or the EF distribution. However, there are methods for estimating the urine coverage if the

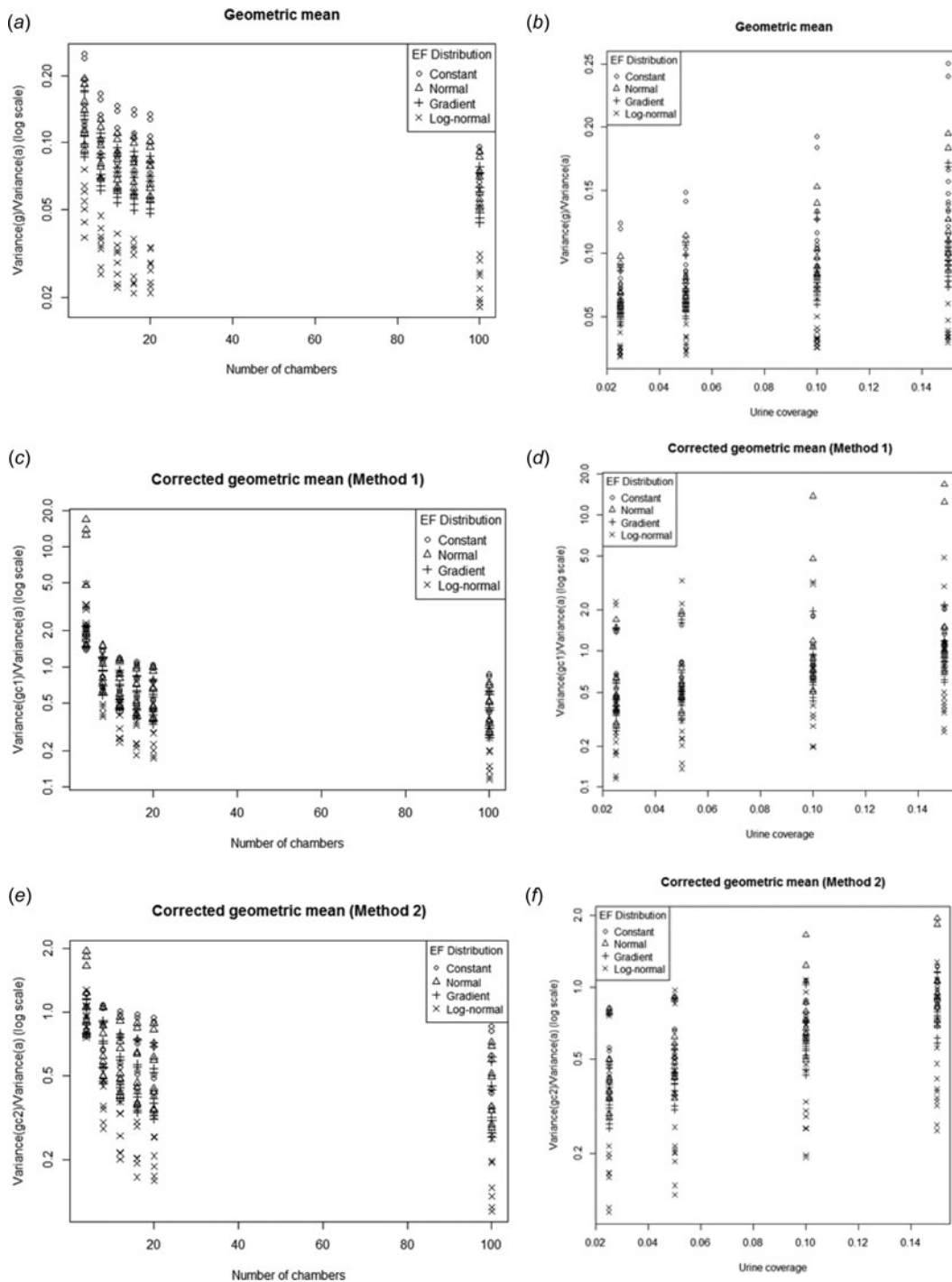


**Fig. 12.**  $rBias$  plotted against number of chambers (a, c, e) and urine coverage (b, d, f) for geometric mean (a, b), first corrected geometric mean (c, d), second corrected geometric mean (e, f).

number of animals grazed and the duration of the grazing event are known (e.g. Pleasants *et al.* 2007) and there have recently been some studies looking at measuring urine patch concentration experimentally (Moir *et al.* 2011). The EF distribution is trickier. In particular, the formulae for variance and skewness

are poor predictors when the EF distribution is log-normal. Therefore determining the nature of the EF distribution should be a priority.

Some alternative estimators were also examined. Log-transformations are commonly used to make right-skewed data more symmetrical. The geometric mean

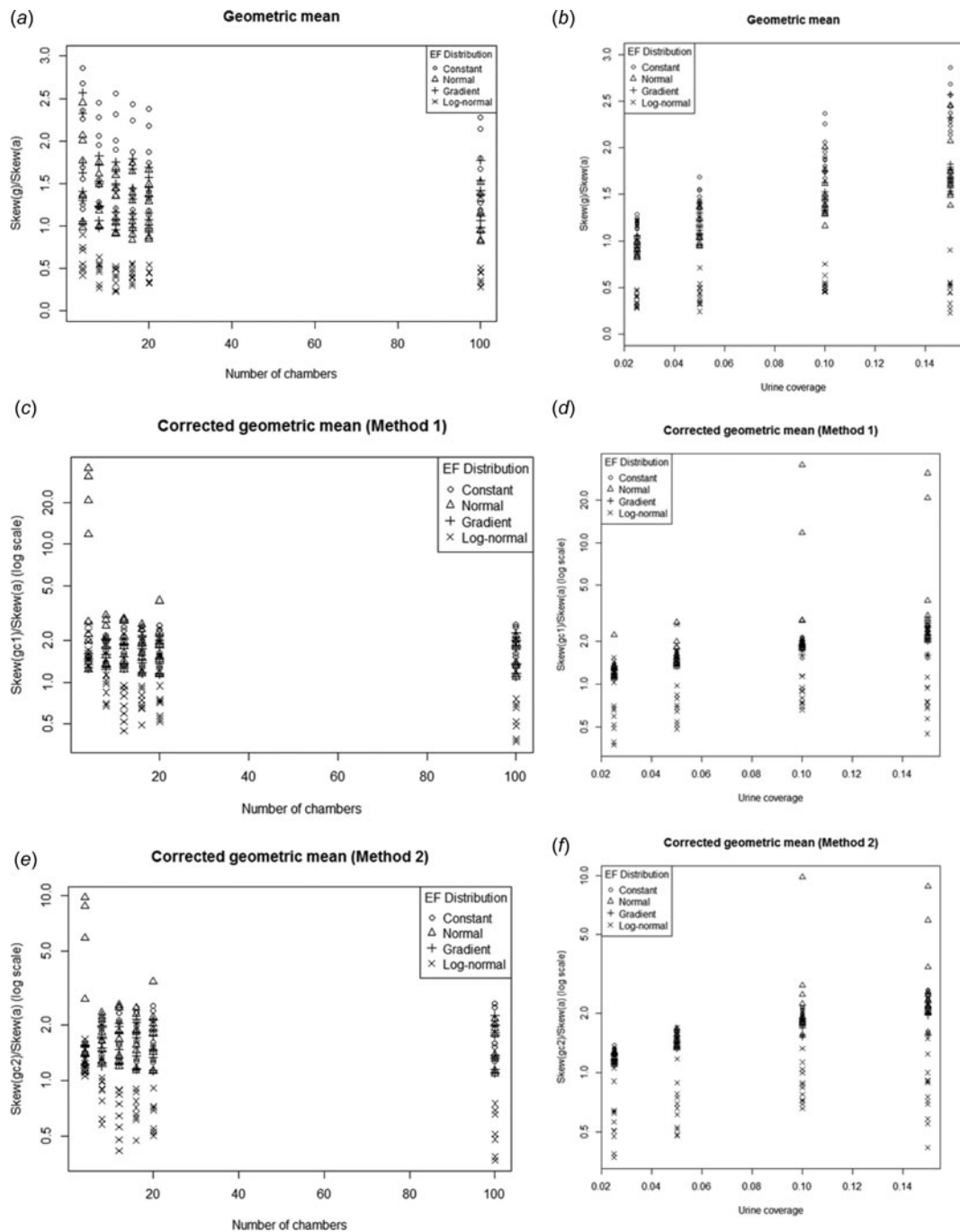


**Fig. 13.** Ratio of the variance of  $g$ ,  $g_{c1}$  and  $g_{c2}$  to the variance of  $a$  plotted against the number of chambers (a, c, e) and urine coverage (b, d, f). Note log scale on the y-axis for the corrected geometric means.

is the mathematical equivalent of back-transforming the mean of the log-transformed data. Unfortunately, the geometric mean was a biased estimator that consistently underestimated the true mean, and although the geometric mean did have a lower variance than the arithmetic mean, it did not necessarily have less skew. Still,

the lower variance of the geometric mean could make it a useful estimator for situations where the aim is to compare N<sub>2</sub>O emissions from different treatments, rather than to estimate the actual emissions.

Two methods for bias-correcting the geometric mean were also investigated. These methods both



**Fig. 14.** Ratio of the skewness of  $g$ ,  $g_{c1}$  and  $g_{c2}$  to the skewness of  $a$  plotted against number of chambers (a, c, e) and urine coverage (b, d, f). Note log scale on the y-axis for the corrected geometric means.

used the sample standard deviation of the log-transformed data to calculate the bias correction. While these methods did reduce the bias they also had the unfortunate side effect of increasing the variance. None of the alternative estimators universally outperformed the arithmetic mean. However, in some circumstances (e.g. when the EF distribution is log-normal) a bias-corrected geometric mean might

make a better estimator, but this would need to be determined on a case-by-case basis.

A number of simplifying assumptions were made for the current simulations, which included that the urine patches did not overlap and that the EF was independent of the urine concentration. The first assumption is more likely to be satisfied when stocking rates are low and grazing durations short. According to the method



of Pleasants *et al.* (2007) with the mean patch area (0.37 m<sup>2</sup>) and urination rate (0.53 events/cow/hour) measured by Moir *et al.* (2011), one day grazing at a stocking density of 180 cows/ha would result in 0.081 of the area being covered by urine patches, but 0.96 of the urine patch area would be single urine patches. So overlapping urine patches would account for only 0.003 of the total area. However, this assumes that the distribution of urine patches is completely random. In reality, animal behaviour is likely to result in non-random urine patch distribution. For example, urine patches are likely to be more frequent in areas where animals congregate (e.g. near water troughs or shelter), but there could also be a tendency to avoid grazing near existing urine patches. The measurements of Moir *et al.* (2011) found urine patch coverage in a season ranged from 0.035 to 0.076 with cow-days/ha/season ranging from 120 to 450. This lower coverage indicates a higher rate of overlap than estimated using the Pleasants method.

The second assumption requires that the relationship between N<sub>2</sub>O emissions and urine N concentration be linear. Kim *et al.* (2013) hypothesized that the relationship between N<sub>2</sub>O emissions and applied N could have linear and non-linear phases depending on the limiting process. Kelliher *et al.* (2014) found a strong linear relationship between N<sub>2</sub>O emissions and applied N, but no significant quadratic term for urea applied at rates of 25 to 1500 kg N/ha. However, in a field study, de Klein *et al.* (2014) found an increasing trend of EF with urine N concentration in a free-draining soil, but not on a poorly draining soil. In addition to a direct relationship between urine concentration and EF, there could also be an indirect effect if the spatial variability of EF was driven by some factor (such as soil moisture or slope) that also affects the amount of time the animals spend in a given part of the field.

The present study also did not consider the effect of using different means and standard deviations for the EF and U concentrations. This is an area that could be further investigated if more data regarding these distributions were available. However, completely random sampling on a grazed pasture is always going to be problematic due to the large difference between the emissions from urine and non-urine patches and the lack of knowledge of the number of urine patches included in a given sample. If non-destructive methods for determining the urine patch locations shortly after a grazing event can be developed then a stratified sampling scheme could be

used to give more robust estimates of emissions. Alternatively, if the presence or absence of a urine patch in each chamber could be conclusively determined after the fact (e.g. from soil properties or grass growth), then the relative contribution from the urine and non-urine containing chambers could be scaled according to the overall urine coverage. The method of applying a known amount of urine-N to each chamber avoids the errors introduced by not knowing the location of the urine patches, but requires the total amount of urine-N excreted during grazing to be estimated, which may miss effects caused by animal behaviour (e.g. treading).

## CONCLUSIONS

In the present study formulae were derived that can be used to calculate the mean, bias, variance and skewness of the sample distribution when measuring N<sub>2</sub>O emissions using chambers on a grazed pasture. These formulae require knowledge of the fractional area covered by urine patches, number of chambers, and the nature of the EF distribution.

The arithmetic mean was an unbiased estimator, and over a large number of trials the mean of all the samples will approach the true value. Unfortunately, the sample mean distribution had a high variance and skew. This meant for any particular trial the sample mean could be a long way from the true value and there was a >50% chance that any given sample would be an underestimate. The CLT did not apply for realistic numbers of chambers, therefore many common statistical techniques requiring normal data (e.g. *t*-tests and *F*-tests) should not be applied to data of this type. The behaviour of the sample mean distribution did not depend on whether the urine concentration was constant or normally distributed, but did depend on the distribution of the EF.

The geometric mean had lower variance, but high bias compared with the arithmetic mean. Two bias corrections for the geometric mean were tried. These methods did reduce the bias of the geometric mean, but as the corrections used estimates of the standard deviation, the variances were high. The corrected geometric means were not a universal improvement on the arithmetic mean, although in some circumstances they had lower variances and skewness (e.g. when the EF was log-normally distributed and  $n > 4$ ). However, all these alternative estimators still suffered from non-normality.

The best way to improve the estimation of N<sub>2</sub>O emissions from a grazed pasture would be to sample separately from known urine and non-urine patch areas. This, however, requires the development of a non-disruptive method for identifying urine patches shortly after a grazing event. In the absence of such a method there will always be some inherent problems (e.g. skewness) in the sample mean distribution. The distribution of the EF was also an important factor in the properties of the sample mean distribution. Knowledge of how the EF varies spatially is therefore useful for determining the properties of the sample mean distribution.

This research was supported by Core funding for Crown research institutes from the Ministry of Business, Innovation and Employment's Science and Innovation Group. The authors thank Andrew McMillan for his helpful comments on the draft paper and Anne Austin for internal editing.

## REFERENCES

- BAIN, L. J. & ENGELHARDT, M. (1992). *Introduction to Probability and Mathematical Statistics*, 4th edn. Pacific Grove, CA, USA: Duxbury Classic.
- BALL, B. C., HORGAN, G. W., CLAYTON, H. & PARKER, J. P. (1997). Spatial variability of nitrous oxide fluxes and controlling soil and topographic properties. *Journal of Environmental Quality* **26**, 1399–1409.
- BOLAN, N. S., SAGGAR, S., LUO, J., BHANDRAL, R. & SINGH, J. (2004). Gaseous emissions of nitrogen from grazed pastures: processes, measurements and modelling, environmental implications, and mitigation. *Advances in Agronomy* **84**, 37–120.
- DE KLEIN, C. A. M. & HARVEY, M. (2012). *Nitrous Oxide Chamber Methodology Guidelines*. Global Research Alliance on Agricultural Greenhouse Gases. Wellington, New Zealand: Ministry for Primary Industries. Available from: <http://www.globalresearchalliance.org/research/livestock/activities/nitrous-oxide-chamber-methodology-guidelines/> (accessed April 2015).
- DE KLEIN, C. A. M., BARTON, L., SHERLOCK, R. R., LI, Z. & LITTLEJOHN, R. P. (2003). Estimating a nitrous oxide emission factor for animal urine from some New Zealand pastoral soils. *Australian Journal of Soil Research* **41**, 381–399.
- DE KLEIN, C. A. M., LUO, J., WOODWARD, K. B., STYLES, T., WISE, B., LINDSEY, S. & COX, N. (2014). The effect of nitrogen concentration in synthetic cattle urine on nitrous oxide emissions. *Agriculture Ecosystems & Environment* **188**, 85–92.
- GILTRAP, D. L., BERBEN, P., PALMADA, T. & SAGGAR, S. (2014). Understanding and analysing spatial variability of nitrous oxide emissions from a grazed pasture. *Agriculture, Ecosystems & Environment* **186**, 1–10.
- GRANT, R. F. & PATTEY, E. (2003). Modelling variability in N<sub>2</sub>O emissions from fertilized agricultural fields. *Soil Biology & Biochemistry* **35**, 225–243.
- HAYNES, R. J. & WILLIAMS, P. H. (1993). Nutrient cycling and soil fertility in the grazed pasture ecosystem. *Advances in Agronomy* **49**, 119–199.
- HOEFT, I., STEUDE, K., WRAGE, N. & VELDKAMP, E. (2012). Response of nitrogen oxide emissions to grazer species and plant species composition in temperate agricultural grassland. *Agriculture, Ecosystems & Environment* **151**, 34–43.
- KELLIHER, F. M., LI, Z. & NOBEL, A. D. L. (2014). Nitrogen application rate and nitrous oxide flux from a pastoral soil. *New Zealand Journal of Agricultural Research* **57**, 370–376.
- 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.
- LUO, J., LINDSEY, S. B. & LEDGARD, S. F. (2008). Nitrous oxide emissions from animal urine application on a New Zealand pasture. *Biology and Fertility of Soils* **44**, 463–470.
- LUO, J., LEDGARD, S. F. & LINDSEY, S. B. (2013). Nitrous oxide and greenhouse gas emissions from grazed pastures as affected by use of nitrification inhibitor and restricted grazing regime. *Science of the Total Environment* **465**, 107–114.
- Ministry for the Environment (2014). *New Zealand's Greenhouse Gas Inventory 1990–2012*. Wellington, New Zealand: Ministry for the Environment. Available from: <http://www.mfe.govt.nz/publications/climate-change/new-zealands-greenhouse-gas-inventory-1990%E2%80%932012> (accessed April 2015).
- MOIR, J. L., CAMERON, K. C., DI, H. J. & FERTSAK, U. (2011). The spatial coverage of dairy cattle urine patches in an intensively grazed pasture system. *Journal of Agricultural Science, Cambridge* **149**, 473–485.
- MOSIER, A. R., DUXBURY, J. M., FRENEY, J. R., HEINEMEYER, O. & MINAMI, K. (1996). Nitrous oxide emissions from agricultural fields: assessment, measurement and mitigation. *Plant & Soil* **181**, 95–108.
- NEYMAN, J. & SCOTT, E. L. (1960). Correction for bias introduced by a transformation of variables. *Annals of Mathematical Statistics* **31**, 643–655.
- PARKIN, T. B. (1987). Soil microsites as a source of denitrification variability. *Soil Science Society of America Journal* **51**, 1194–1199.
- PENNOCK, D. J., VAN KESSEL, C., FARRELL, R. E. & SUTHERLAND, R. A. (1992). Landscape-scale variations in denitrification. *Soil Science Society of America Journal* **56**, 770–776.
- PLEASANTS, A. B., SHORTEN, P. R. & WAKE, G. C. (2007). The distribution of urine deposited on a pasture from grazing animals. *Journal of Agricultural Science, Cambridge* **145**, 81–86.
- SAGGAR, S., ANDREW, R. M., TATE, K. R., HEDLEY, C. B., RODDA, N. J. & TOWNSEND, J. A. (2004). Modelling nitrous oxide emissions from dairy-grazed pastures. *Nutrient Cycling in Agroecosystems* **68**, 243–255.
- SAGGAR, S., HEDLEY, C. B., GILTRAP, D. L. & LAMBIE, S. M. (2007). Measured and modelled estimates of nitrous oxide

- emission and methane consumption from a sheep-grazed pasture. *Agriculture, Ecosystems & Environment* **122**, 357–365.
- SELBIE, D. R., BUCKTHOUGHT, L. E. & SHEPHERD, M. A. (2015). The challenge of the urine patch for managing nitrogen in grazed pasture systems. *Advances in Agronomy* **129**, 229–292.
- TURNER, D. A., CHEN, D., GALBALLY, I. E., LEUNING, R., EDIS, R. B., LI, Y., KELLY, K. & PHILLIPS, F. (2008). Spatial variability of nitrous oxide emissions from an Australian irrigated dairy pasture. *Plant & Soil* **309**, 77–88.
- VAN DER WEERDEN, T. J., LUO, J., DE KLEIN, C. A. M., HOOGENDOORN, C. J., LITTLEJOHN, R. P. & RYS, G. J. (2011). Disaggregating nitrous oxide emission factors for ruminant urine and dung deposited onto pastoral soils. *Agriculture, Ecosystems & Environment* **141**, 426–436.
- VELTHOF, G. L., JARVIS, S. C., STEIN, A., ALLEN, A. G. & OENEMA, O. (1996). Spatial variability of nitrous oxide fluxes in mown and grazed grasslands on a poorly drained clay soil. *Soil Biology & Biochemistry* **28**, 1215–1225.
- VELTHOF, G. L., VAN GROENIGEN, J. W., GEBAUER, G., PIETRZAK, S., JARVIS, S. C., PINTO, M., CORRÉ, W. & OENEMA, O. (2000). Temporal stability of spatial patterns of nitrous oxide fluxes from sloping grassland. *Journal of Environmental Quality* **29**, 1397–1407.
- YANAI, J., SAWAMOTO, T., OE, T., KUSA, K., YAMAKAWA, K., SAKAMOTO, K., NAGANAWA, T., INUBUSHI, K., HATANO, R. & KOSAKI, T. (2003). Spatial variability of nitrous oxide emissions and their soil-related determining factors in an agricultural field. *Journal of Environmental Quality* **32**, 1965–1977.