Boundary line analysis of the effect of water-filled pore space on nitrous oxide emission from cores of arable soil

R.M. Lark[a] & A.E. Milne[b]

[a]*British Geological Survey, Keyworth, Nottinghamshire NG12 5GG, U.K. and* [b]*Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, U.K.*

Correspondence: R.M. Lark. E-mail: mlark@bgs.ac.uk

Running title: Water-filled pore space and nitrous oxide emission

Key words: Greenhouse gases, anaerobicity, limiting factor, emission factor

Summary

The boundary line has been proposed as a model of the effects of some variable on a biological response, when this variable might limit the response in only some of a set of observations. It is proposed that the upper boundary (in some circumstances the lower boundary) represents the response function of interest. Boundary line analysis is a method to estimate this response function from data. The approach has been used to model the emission of N_2O from soil in response to various soil properties. However, the methods that have been used to identify the boundary are based on somewhat *ad hoc* partitions of the data. A statistical model that we have presented previously has not been applied to this problem in soil science, and we do so here to represent how the water-filled pore space (WFPS) of the soil affects the rate of N_2O emission. We derive a boundary line response that can be shown to be a better model for the data than an unbounded alternative by statistical criteria. Furthermore, the fitted boundary response model is consistent with past empirical observations and modelling studies with respect to both the WFPS at which the potential emission rate is largest and the measurement error for the emission rates themselves. We show how the fitted model might be used to interpret data on soil volumetric water content with respect to seasonal changes in potential emissions, and to compare potential emissions between soil series that have contrasting physical properties.

- We obtain a boundary model of the effect of water-filled pore space on soil nitrous oxide emission
- The boundary model can be fitted by maximum likelihood allowing for measurement error.
- The boundary model indicates a maximum emission rate with water-filled pore space from 0.7–0.8
- The model can be used to compare potential emission rates of soil with different properties

Introduction

Nitrous oxide (N_2O) is produced in soil by nitrification and denitrification. Microbial denitrification occurs when soil becomes anaerobic. Facultative anaerobic bacteria use nitrate as the electron acceptor in their respiration, reducing it to various forms dominated by N_2O and N_2. The proportions of these two products depend on factors such as soil redox potential and pH (Delwiche, 1981). Nitrous oxide is also produced as a by-product of nitrification, the oxidation of ammonium to nitrate. The relative importance of these two sources of N_2O depends on local conditions (Stevens *et al.,* 1997).

Nitrous oxide is an important greenhouse gas, and it has been estimated that a CO_2 equivalent of 97 Tg C year^{-1} is emitted from agricultural sources across continental Europe (Schulze *et al.,* 2009). This, together with methane, is more or less balanced by the net sink for carbon provided by Europe's grassland and forest. With the intensification of agriculture and forestry a net flux of greenhouse gases to the atmosphere can be expected from agricultural and forest land of Europe (Schulze *et al.,* 2009). We must be able to predict N_2O emissions from soil under different conditions to formulate policy and to design interventions to mitigate this effect.

Various factors determine the rate of N_2O emission from soil (Dobbie & Smith, 2003). Soil organic carbon as a substrate for respiration is unlikely to be limiting on denitrification but the consumption of oxygen by aerobic microflora, stimulated by a supply of organic carbon, might promote the development of anaerobic conditions in which denitrification can occur (Groffman *et al.,* 1987). Bacteria require a supply of nitrate and ammonium to sustain denitrification and nitrification, respectively, and the form in which nitrogen is available in soil affects the rate of N_2O emission (e.g. Bayer *et al.,* 2015). Both nitrification and denitrification respond to temperature (Smith *et al.,* 1998). Soil pH has an effect on most microbially-mediated processes, and it influences the proportions of N_2O and N_2 in denitrification products (Delwiche, 1981). One of the most important factors that affects the rate of denitrification in soil is the development of anaerobic centres where the process can take place. This depends on factors that affect the rate of gaseous diffusion into soil such as compaction (Ball *et al.,* 2000) and the proportion of the soil's pore space that is filled with water (water-filled pore space; WFPS; Smith *et al.,* 1998). Because nitrification is an aerobic process, the relative contributions of the two processes to N_2O emission also depends on the WFPS, see, for example, Bateman & Baggs

(2005).

These factors must be considered in any quantitative model for nitrous oxide emissions from soil. Some progress has been made towards process-based models including the DNDC (Li, 2000) and DAYCENT (Del Grosso *et al.,* 2006) models. Process modelling has also been used to investigate particular factors, such as the effect of WFPS. Rabot *et al.* (2015) used simulation modelling to investigate the effects of WFPS on gas transport and implications for denitrification and the emission of denitrification products from the soil. Their model showed a bell-shaped response to WFPS with the maximum emission rate at a WFPS in the interval [0.76, 0.79]. The WFPS that gave the maximum emission rate drifted from the bottom to the top of this interval with time during an experiment because of the increase in the N_2O concentration gradient between the soil surface and the atmosphere. At smaller WFPS the rate of emission of N_2O is limited by small denitrification rates because anaerobicity is reduced. At larger WFPS the rate of emission is limited by the rate of gaseous diffusion through the soil.

Process models give insight into the factors that contribute to N_2O emissions from soil. Note, for example, how the modelling by Rabot *et al.* (2015) helps us to understand the factors that contribute to the non-linear effect of WFPS. However, process models may be challenging to use in practice because of the need for information on many soil properties, the propagation of error in model parameters and inputs and uncertainty about the model structure. Conen *et al.* (2000) considered that empirical models might be more useful in some circumstances, at least as submodels within broader process models.

Various empirical models have been used to predict N_2O emission rates from agricultural soil. Conen *et al.* (2000) used a model based on soil mineral N content, WFPS and soil temperature, and the same variables were used in a similar approach by Smith & Massheder (2014). Although these models are empirical they are not simple regressions. Rather they use the boundary line concept of Webb (1972). Webb (1972) proposed that, for many biological processes, the boundary (typically the upper boundary) in a scatter plot of the biological response (on the ordinate) against an environmental variable of interest (on the abscissa) expresses best the effect of the environmental variable. Specifically, it represents a maximum response given the value of the environmental variable, which will be expressed only if other factors are not limiting. The two empirical models cited above use boundary line responses to express the effects of WFPS and soil temperature on N_2O emissions. The boundary line method was used by Schmidt *et al.* (2000) to examine the effects of temperature, soil nitrate content and WFPS. In earlier research the boundary line method was used to model the effects of soil properties specifically on denitrification rates (Elliot & de Jong, 1993; Bergstrom & Beauchamp, 1993).

Farquharson & Baldock (2008) suggest that boundary line models might be particularly appropriate for modelling N_2O emissions from soil because of the many factors which affect this process and cannot be controlled in an observational study, and the plausibility of the limiting factor interpretation of the boundary line. Nevertheless, they noted some limitations with the methods that had been used for the boundary line analysis (BLA). For example, Schmidt *et al.* (2000) obtained a boundary line model for the response of N_2O emission rate to some factor by dividing the range of values of the factor into eight equal intervals, and then extracting the observation from within each interval that corresponded to the 99th percentile of N_2O emission rates in that interval. A continuous function was then fitted to the resulting eight data points by ordinary least squares. This is a reasonable heuristic approach, and is similar to other methods published at the time of their study (Schnug *et al.,* 1996) and since (Shatar & McBratney, 2004). However, as Farquharson & Baldock (2008) pointed out, these methods provide no statistical evidence that the boundary line is a plausible model of the particular data. Furthermore, they either disregard measurement error in the response variable or deal with it in an arbitrary way. Farquharson & Baldock (2008) referred to previous research that we had undertaken with colleagues to develop exploratory methods to examine the plausibility of the boundary line interpretation of data (Milne *et al.,* 2006a) and a statistical model for the boundary line which can be fitted by maximum likelihood (Milne *et al.,* 2006b). The suggested that these methods be applied in studies on N_2O production from soil stating that 'The adoption of BLA to define relationships could be of considerable benefit to model development as it provided a more appropriate way to define bivariate relationships where other factors cannot be controlled.' The papers by Milne *et al.* (2006a,b) demonstrated our BLA methods in examples from plant physiology, agronomy and studies on soil carbon. We are not aware of any studies that have applied our method to the study of N_2O emission from soil. Therefore we decided to use it to investigate the effect of WFPS on N_2O emission rate with data from a previous study on arable soil (Lark *et al.,* 2004).

The proposed methodology for BLA has two stages. In an initial exploratory analysis the evidence for an upper boundary line model, provided by a concentration of observations near the upper limit of the scatter plot, is examined by counting the number of upper vertices in the first few convex hull 'peels' (Eddy, 1982) of the scatter plot of the response variable against the environmental variable of interest. The convex hull of a set of data in a plane is the subset of points that are the vertices of the convex polygon which includes exactly all the data. The convex hull of a bivariate data set is its first peel. The convex hull of the remaining data after the first peel is removed is the second peel, and so on. In the exploratory analysis of data, these are compared with the expected number of vertices in the null case represented by a bivariate normal joint distribution of the two variables. We expect to see more vertices than are expected in the null case if the upper boundary of the scatter plot represents the limiting response to the variable on the abscissa of the plot. Milne *et al.* (2006a) describe the method. Second, the boundary line is then modelled as a function that censors a joint bivariate normal distribution of the underlying response variable, y, and the measured environmental factor, x, on the abscissa of the plot (Milne *et al.,* 2006b). In summary, if the boundary line is described by $b(x)$, then a variate from the joint distribution $\{y, x\}$ where $y > \bar{y} = b(x)$ is replaced by $\{\bar{y}, x\}$. However, the response variable might be measured with error, and so observed variates $\{\breve{y}, x\}$ might occur above the boundary line. The model is fitted by finding maximum likelihood estimates of the bivariate normal distribution of $\{y, x\}$, parameters of the boundary function $b(x)$ and the variance of the measurement error, assumed to be a normal random variable with a mean of zero. By comparing the maximized likelihood for this distribution with the maximized likelihood of a bivariate normal joint distribution one may assess the weight of evidence for the boundary line model.

In this paper we use the methods of Milne *et al.* (2006a,b) to analyse a data set on rates of emission of N_2O from cores of arable soil, and their WFPS. We used this variable so that we could compare the WFPS at the maximum rate of emission rate in the fitted boundary model with the results of the process modelling reported by Rabot *et al.* (2015). We use a somewhat different formulation of the censored model for the boundary line to that presented by Milne *et al.* (2006b). We used conditional densities, which allows a more straightforward treatment of measurement error for the response variable. This is presented in the next section, followed by an account of the data and specific analyses.

Theory

The boundary line model

The boundary line model is a bivariate distribution of an observed response variable, $\breve{y}$ and an independent covariate x. This model is based on a latent normal random variate $\mathbf{z} = \{y, x\}^{\mathrm{T}}$ with joint density function

$$f(y, x) \;=\; \phi_2\left(\mathbf{z}|\boldsymbol{\mu}, \mathbf{C}\right), \tag{1}$$

where $\phi_2()$ denotes the bivariate normal density function for a random variate with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. The variate, $\mathbf{z}$, is censored by a boundary function $b(x|\boldsymbol{\beta})$ with parameters in $\boldsymbol{\beta}$, to give a censored variate $\bar{\mathbf{z}} = \{\bar{y}, x\}^{\mathrm{T}}$. In the case of an upper boundary:

$$\bar{\mathbf{z}} = \;\{\min\left(y, b(x|\boldsymbol{\beta})\right), x\}^{\mathrm{T}}. \tag{2}$$

We assume that the independent variable is known without error, as in the general linear model, and that the observed response variable, $\breve{y}$, arises from the observation of $\bar{y}$ with a normal error of mean zero and standard deviation σ_{e}, such that the distribution of the observed value conditional on $\bar{y}$ is:

$$\breve{y}|\bar{y} \;\sim\; \mathcal{N}\left(\bar{y}, \sigma_{\mathrm{e}}\right). \tag{3}$$

The boundary line model has three sets of parameters. These are the parameters of the censoring function, in $\boldsymbol{\beta}$, the parameters (means and covariances) of the latent bivariate normal random variate, and the observation error σ_{e}. Our objective is to estimate these parameters by maximum likelihood, given the observed values $\{\breve{y}_1, \breve{y}_2, ..\breve{y}_n\}$ and $\{x_1, x_2, ..x_n\}$. To obtain the appropriate likelihood function we require the joint density for $\breve{y}$ and x conditional on the parameters:

$$f\left(\breve{y}, x|\boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{C}, \sigma_{\mathrm{e}}\right). \tag{4}$$

For brevity we drop the parameters from the density functions. Following familiar properties of conditional densities we may write

$$f\left(\breve{y}, x\right) \;=\; f\left(\breve{y}|x\right) f\left(x\right), \tag{5}$$

where $f(x)$ is the probability density function for x. From the assumptions made about the measurement error we may write the conditional density in Equation (5) as

$$f(\breve{y}|x) \;=\; f(\bar{y}|x) * f_{\mathrm{N}}(v|0,\sigma_e), \tag{6}$$

where $f * g$ denotes a convolution of two functions and $f_{\mathrm{N}}(v|\mu,\sigma)$ denotes a normal density with specified parameters.

The conditional density $f(\bar{y}|x)$ in Equation (6) above is the censoring of conditional density $f(y|x)$ which may be written as:

$$f(y|x) \;=\; f_{\mathrm{N}}(y|\mu_{y|x},\sigma_{y|x}), \tag{7}$$

where $\mu_{y|x}$ and $\sigma_{y|x}$ are the conditional mean and standard deviation respectively of y:

$$\mu_{y|x} \;=\; \mu_y + \left(x - \mu_x\right)\frac{\mathrm{Cov}\{x,y\}}{\sigma_x^2}, \tag{8}$$

and

$$\sigma_{y|x} \;=\; \sigma_y\sqrt{1-\rho^2}, \tag{9}$$

where the means μ_y and μ_x are elements of the mean vector $\boldsymbol{\mu}$ in Equation (1) and the correlation and (co)variances are from the covariance matrix $\mathbf{C}$ in Equation (1). The censored conditional density, right-censored which implies an upper boundary, can be written, therefore, as

$$\begin{aligned} f(\bar{y}|x) &= f_{\mathrm{N}}(y|\mu_{y|x},\sigma_{y|x}), \qquad y < b(x) \\ &= \int_{b(x)}^{\infty} f_{\mathrm{N}}(y|\mu_{y|x},\sigma_{y|x})\,\mathrm{d}y, \qquad y = b(x) \\ &= 0 \qquad y > b(x). \end{aligned} \tag{10}$$

Following Turban (2010), we may now obtain the density of the observed variable, the convolution of the right-censored density in Equation (10) for an upper-boundary line with the observation error density as:

$$f(\breve{y}|x) \;=\; \zeta_{\mathrm{U}}\,\gamma_{\mathrm{U}}\,\eta_{\mathrm{U}}\exp\left\{-\frac{\left(\breve{y}-\mu_{y|x}\right)^2}{2\left(\sigma_{\mathrm{e}}^2+\sigma_{y|x}^2\right)}\right\} + (1-\zeta_{\mathrm{U}})\,f_{\mathrm{N}}\left(\breve{y}|b(x),\sigma_{\mathrm{e}}\right), \tag{11}$$

where

$$\gamma_{\mathrm{U}} \;=\; \frac{\beta\sqrt{2\pi}}{2\pi\sigma_{y|x}\sigma_{\mathrm{e}}\left\{1-\Phi\left(\frac{\mu_{y|x}-b(x)}{\sigma_{y|x}}\right)\right\}}, \tag{12}$$

$\Phi()$ denotes the standard normal distribution function,

$$\eta_{\mathrm{U}} = 1 - \Phi\left(\frac{\breve{y} - b(x) - \alpha}{\beta}\right), \quad (13)$$

$$\zeta_{\mathrm{U}} = \Phi\left(\frac{b(x)}{\sigma_{y|x}}\right), \quad (14)$$

$$\alpha = \frac{\sigma_{\mathrm{e}}^2\left(\bar{y} - \mu_{y|x}\right)}{\sigma_{y|x}^2 + \sigma_{\mathrm{e}}^2}, \quad (15)$$

and

$$\beta^2 = \frac{\sigma_{y|x}^2 \sigma_{\mathrm{e}}^2}{\sigma_{y|x}^2 + \sigma_{\mathrm{e}}^2}. \quad (16)$$

In the case of a left-censored conditional density for a lower boundary line, the same expression may be used but with γ_{L}, η_{L} and ζ_{L} substituted for γ_{U}, η_{U} and ζ_{U} respectively, where

$$\gamma_{\mathrm{L}} = \frac{\beta\sqrt{2\pi}}{2\pi\sigma_{y|x}\sigma_{\mathrm{e}}\Phi\left(\frac{\mu_{y|x} - b(x)}{\sigma_{y|x}}\right)}, \quad (17)$$

$$\eta_{\mathrm{L}} = \Phi\left(\frac{\breve{y} - b(x) - \alpha}{\beta}\right), \quad (18)$$

and

$$\zeta_{\mathrm{L}} = 1 - \Phi\left(\frac{b(x)}{\sigma_{y|x}}\right). \quad (19)$$

For some proposed set of parameters $\boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{C}$ and σ_{e}, and a pair of observed values $\breve{y}$ and x one may compute the density from Equation (11). Treating each of a set of n observations as independent, one may compute the negative log-likelihood for a set of parameter values, given the observations, as

$$\ell = -\sum_{i=1}^{n} \log f(\breve{y}_i|x_i). \quad (20)$$

Materials and methods

Data collection

Our data are drawn from a study on the spatial variation of N_2O emissions from soil cores from a regular transect (Lark *et al.* 2004). The measurements were made on incubated intact cores so that temperature was fixed, but other factors (such as water content) varied between soil cores to reflect variation in the field. Ideally, measurements would be made *in situ* in the field. However, spatial analysis and BLA require large data sets, and it is difficult to collect these in the field without confounding spatial with temporal variation. We chose, therefore, to make measurements on incubated intact cores, given previous experience of making useful measurements of denitrification and mineralization this way (Ryden *et al.*, 1987; Webster & Goulding, 1989; Hatch *et al.*, 1990).

A full account of the data collection is given by Lark *et al.* (2004), but we provide an outline here. Soil samples were taken within a period of seven hours in the autumn of 2000 on a straight transect, with a spacing of 4 m and to give 256 sample points across the farm of the former Silsoe Research Institute in Bedfordshire in Eastern England. All fields traversed by the transect had been under a cereal crop in the summer of 2000, and had been either recently drilled with an autumn-sown crop or were under stubble. At each site a gouge auger of length 150 mm and diameter 44 mm was pushed fully into the soil, twisted and removed. Four cores were taken in this way at each site. The cores were transported with a minimum of delay to a cold room at 4°C, and were kept at this temperature until they were analysed.

In subsequent laboratory analysis one core was selected from each site, and its fresh weight and length were recorded. Cores were placed in a 1-litre Kilner jar and pre-incubated at 15°C for 17–24 hours with the jar lids in position to prevent desiccation of the core, but unsealed to allow some gas transfer from the jar. After pre-incubation the jars were flushed with laboratory air and re-sealed with a rubber gasket and clamped in position so that they were gas-tight. An initial sample (20 ml) of the gas headspace was collected and injected into an evacuated vaco tube. The jars were incubated at 15°C for 24 hours, and then two further 20-ml samples from the headspace were collected. Within a few days the gas samples were injected into an Ai93 gas chromatograph which analysed them for N_2O by an electron capture detector (ECD). The rate of emission was determined from the change in concentrations of N_2O in the headspace of the incubation jar.

After incubation the core was cut in half and the moisture content of one half of the core was determined by oven-drying to constant weight at 105°C allowing the determination of volumetric water content. The dry bulk density was determined. The rate of N_2O emission could then then expressed on an area basis as in previously-cited modelling studies (here g N ha^{-1} day^{-1}). Other analyses were undertaken on the soil, including soil organic carbon content (SOC) by a combustion method following Tabatabai & Bremner (1991).

Ideally the saturated water content of the soil would be determined directly by saturating the core and then determining its volumetric water content. This value for the saturated water content would then be used to compute WFPS of the field-moist soil. However, this was not required in the original study and further destructive analyses were done on the cores so that material was not available to do this measurement subsequently. For this reason we chose to compute the total pore space of the soil (cm^3 cm^{-3}) from the measured bulk density on the assumption of a particle density of 2.65 g cm^{-3} (Hall *et al.,* 1977), as proposed by Minasny *et al.* (1999) when no soil physical data other than bulk density are available. We recognize that this introduces an approximation into our data on WFPS because soil particle density may vary, and, furthermore, the total porosity might differ from saturated water content. Our soil samples were from arable sites only (excluding waste ground and field boundaries), therefore the variation in SOC was small (see Table 1) and so this approximation seems reasonable. The approach has been used to determine WFPS for modelling microbial processes in soil in a range of studies (e.g. Wu *et al.,* 2015; Franzluebbers, 1989; Linn & Doran, 1984). The WFPS was computed from this total porosity and the measured volumetric water content of the field-moist soil.

It is useful to note two findings from a later study in which soil cores were taken from a longer transect over more heterogeneous land uses (Haskard *et al.,* 2010). All protocols were identical, the exception was that the head space samples from the incubation jars were placed in vials on a Perkin Elmer Turbo Matrix 110 Headspace autosampler (Perkin Elmer, Waltham, MA). The autosampler was linked to a Perkin Elmer Clarus 500 gas chromatogram (GC) (Perkin Elmer) by a fused silica transfer line to allow the automatic analysis of samples. The samples in this latter study were allocated at random to batches for measurement of rates of N_2O emission. Analysis of these data showed that the effect of batch (and so the length of storage time of the sample) was negligible and statistically insignificant (Haskard *et al.,* 2010). Duplicate measurements were made on 78 cores from this latter study, which have been analysed subsequently to estimate the measurement error standard deviation. The estimated standard deviation of measurement error was 0.46 log g

N ha^{-1} day^{-1}.

The objective of the present study is to examine the effect of WFPS on rates of N_2O emission from arable soil. The original sampling transect was regular, so included 10 cores from tracks, ditches and waste ground under rough vegetation. Data from these cores were not used in the analysis reported here. We also excluded the first two cores from the headland of the northernmost field on the transect. This field was on the lightest soil formed over the Lower Greensand, and both cores were markedly compacted by the auger.

Data analysis

Exploratory analysis. In the boundary line statistical model observations are assumed to be drawn from a bivariate normal variable with an upper censor on the values of the response variable that depends on the value of the environmental variable. To make these assumptions plausible transformation of the variables may be necessary. The data on rates of N_2O emission were transformed to natural logarithms for analysis given their markedly skewed distribution. The data on WFPS were not markedly skewed, but, as proportions, they cannot be regarded as normally distributed and so were transformed to logits (natural logarithms) before analysis. The response variable in our boundary line analysis, variable $\breve{y}$ in Equation (11), was log N_2O emission rate, and the independent variable, x, in Equation (11), was the logit of WFPS.

In circumstances where a boundary line model is appropriate, and the sampling is sufficiently wide-ranging to cover a range of conditions with different limiting factors, one would expect to find a concentration of observations in a scatter plot of $\breve{y}$ against x near the boundary. This might be evident when the plot is examined, but we do not want to rely on visual assessment. It would be difficult to compare consistently the density of observations near a putative boundary and the concentration that would be expected under an unbounded model. Therefore, we prefer to use an objective statistical test of the density of observations near the boundary of interest. This was first described by Milne *et al.* (2006a). In this test the number of upper vertices in the first few convex hull peels of the scatter plot of the transformed flux and WFPS data are counted and compared with expected numbers under the null hypothesis of a bivariate normal distribution. The first convex hull peel of a scatter plot corresponds to the observations that are on the convex polygon that exactly encloses all the data points. The points on the convex hull are vertices of this polygon. If one removes the points on the convex hull of the first data set (these are called the first peel of the data) a convex hull can be determined for the remaining points. The observations that are on the vertices of this second convex hull constitute the second peel of the data.

Consider a peel of the scatter plot of our response variable (y) on the ordinate against the potentially limiting variable (x) on the abscissa. We number the vertices from 1 to N clockwise on the convex hull where the first vertex is $\mathbf{v}_1 = \{x_1, y_1\}$ such that if

$$x_1 = \min\{x_i\}_{i\in\{1,\ldots,N\}} \quad \text{and if} \quad (x_1 = x_j, j \in \{2,\ldots,N\}) \quad \text{then } y_1 \leq y_j. \tag{21}$$

The upper convex hull is defined as the ordered set of vertices $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_k$ where

$$x_k = \max\{x_i\}_{i\in\{1,\ldots,N\}}. \tag{22}$$

If an upper boundary is a plausible model of the relation between $\breve{y}$ and x then we would expect to find more vertices in the first few peels of the scatter plot than are expected under the bivariate normal null model. We followed Milne *et al.* (2006a) by counting the number of vertices of the upper convex hull in the first 5 to 10 peels of the data. Hueter (1994) showed that the asymptotic distribution of the number of vertices in the first peel of a bivariate normal random variate is normal. Milne *et al.,* (2006a) conjectured that the numbers of vertices in successive peels can be approximated as a normal random variable; Monte Carlo simulations supported this. They used the output from these simulations to find emulators for the number of vertices in successive peels of bivariate normal variates, and their variances. In the procedure that we used here the number of vertices in the first 5 peels was compared to the number expected under a null hypothesis of bivariate normality, and this was repeated for the first $6, 7 \ldots 10$ peels. This was done because a concentration of vertices is expected where the peel is close to any boundary function, but this might not be true of early peels of the data in the presence of measurement error. Because these six hypotheses constituted a multiple hypothesis test, and the hypotheses are not independent (the number of vertices in the first $n + 1$ peels is the number in the first n plus the number in the $n + 1$th), we applied the false discovery rate control procedure for hypothesis testing with non-independent hypotheses proposed by Benjamini & Yekeutieli (2001). Details of this procedure are given by Milne *et al.* (2006a).

Boundary line model. Schmidt *et al.* (2000) used a bell-shaped function for the boundary response of N_2O emission to WFPS:

$$\mathcal{F}_{\mathrm{N_2O}}(\vartheta) \;=\; \varphi_{\max} \exp\left\{ \frac{-2\,(\vartheta - 0.72)^2}{0.074} \right\}, \tag{23}$$

where the WFPS (a dimensionless proportion) is denoted by ϑ and $\mathcal{F}_{\mathrm{N_2O}}(\vartheta)$ denotes the boundary N_2O flux at this value of WFPS with a maximum value of $\varphi_{\max}$ when $\vartheta = 0.72$. Note that this equation is rescaled from Schmidt *et al.* (2000) who specified WFPS as a percentage, and that they define this function for $0.3 < \vartheta < 0.93$. Rabot *et al.* (2015) used the same function. In this study, we selected an equivalent expression for the boundary function on the log scale and with WFPS on the logit scale:

$$\log\left\{\mathcal{F}_{\mathrm{N_2O}}(\vartheta)\right\} \;=\; \beta_0 - \beta_2\left(\mathrm{logit}(\vartheta) - \beta_1\right)^2. \tag{24}$$

In this formulation β_0 is the logarithm of $\varphi_{\max}$; the maximum flux occurs when $\mathrm{logit}(\vartheta) = \beta_1$ and β_2 is a scaling parameter, which is zero if the boundary is a constant (not dependent on WFPS). Other functions have been used to model this effect, including a quadratic (Wu *et al.,* 2015), and could be used to model the boundary function.

We fitted this model as a boundary line to our data by finding values of the parameters $\beta_0, \beta_1,$ and β_2 that maximized the likelihood computed with Equation (20) where the function in Equation (24) is substituted for the general boundary function $b(x)$ in Equation (11). This was done on the R-platform for statistical computing (R Core Team, 2014) with the optim procedure and the quasi-Newton BFGS algorithm for optimization (Broyden, 1970). When the set of parameters that maximized the likelihood in Equation (20) was found, we evaluated the Hessian matrix of the likelihood with respect to the parameters and obtained from it a covariance matrix for the estimation error of the parameters (see Dobson, 1990). We recognize that the likelihood computed with Equation (20) treats the observations as independent, which requires independent random sampling. Our data were not collected this way, but we make the assumption as a first approximation, and use the independent estimate of the standard deviation of the measurement error (referred to above) to check for evidence of bias in the variance parameters of the model that could arise from the lack of independence.

One way to evaluate the boundary line model is to compare its fit with a simpler alternative in which the two variables considered are modelled as a bivariate normal random variate (Milne *et al.,* 2006b). A multivariate normal model was fitted to the observations by minimizing the negative log-likelihood. The minimized negative log likelihood, ℓ, for the boundary model can be compared with that for the multivariate normal model. However, the latter model has five parameters (the means for the two variables, their variances and their correlation). The boundary model has the same parameters (for the underlying bivariate normal process) in addition to three parameters of the censoring boundary line and the standard deviation of observation error. These extra parameters mean that the boundary line model must fit at least as well as the multivariate normal, as judged by the value of ℓ, and might be expected to fit better, even in a case where the multivariate normal model holds. It is necessary, therefore to account for the additional parameters in the boundary line model when making the comparison. This comparison can be made with Akaike's information criterion (AIC) (Akaike, 1973). The AIC is computed by

$$\text{AIC} = 2\ell + 2\mathcal{P}, \tag{25}$$

where $\mathcal{P}$ is the number of parameters in the model. The second term in Equation (25) is a penalty for model complexity. In any comparison the model with the smallest AIC is selected. Although the AIC is not a formal significance test, selection of the model with the smallest AIC minimizes the expected information loss through the selection decision (Verbeke & Molenberghs, 2000).

Results

Exploratory statistics are given in Table 1. Note that the number of vertices in the upper convex hulls of the first 5 to 10 peels all exceed the numbers expected under a null hypothesis of multivariate normality. All six of these null hypotheses can be rejected with the false discovery rate controlled at 0.05, i.e. the expected proportion of rejected null hypotheses that are actually true is no larger than 0.05. Summary plots are shown in Figure 1.

Table 2 gives the results from the fitting of the boundary line. The negative log-likelihood is markedly smaller for the boundary line model than for the alternative multivariate normal model. There are four more parameters in the boundary line model, but the AIC is still substantially smaller for the boundary line model indicating that it is to be preferred to the multivariate normal alternative. This is consistent with the results for the vertices in the convex hull peels of the data. The fitted boundary model is shown in Figure 2 together with a 95% confidence interval for the boundary line obtained from the covariance matrix of the boundary function parameters assuming the estimation errors are normal.

The parameter β_2 of the boundary model is positive for physically plausible cases (i.e. with the boundary convex upward) and zero if the boundary line is flat. The estimate of this parameter for the data analysed in this paper, given in Table 2, is 0.54 with a 95% confidence interval $[0.11, 0.97]$, which is consistent with a physically plausible non-constant boundary line function.

The parameter β_1 of the fitted model, 1.19, (Table 2) is the logit of the WFPS at which the maximum boundary emission rate occurs. On back-transformation this is equivalent to a WFPS of 0.77. The 95% confidence interval for the back-transformed parameter, assuming a normal estimation error, is $[0.69, 0.83]$. This is consistent with the process model results reported by Rabot *et al.* (2015) who found maximum fluxes at WFPS values between 0.757 and 0.798; the variation was attributed to temporal effects during the period when the soil is wetted and variation in soil bulk density. It is also comparable with the boundary line model reported by Schmidt *et al.* (2000) for which the maximum flux was at WFPS $= 0.72$.

Figure 2 shows the fitted boundary line model on the scatter plot of the transformed rate of N_2O emission against WFPS. This shows the maximum value where logit WFPS is 1.19. Also shown, as two dotted lines, is the 95% confidence interval for the boundary line obtained by sampling from the estimated distribution of the three parameters β_0, β_1 and β_2.

The estimated standard deviation of the measurement error, given in Table 2, is 0.53 log g N ha^{-1} day^{-1} with a 95% confidence interval $[0.35, 0.71]$. This is consistent with the standard deviation of duplicate measurements from the second study (Haskard *et al.,* 2010) reported above (0.46 log g N ha^{-1} day^{-1}). Figure 3 shows the profile likelihood for this parameter (the maximized value of the likelihood with this particular parameter fixed at different values). The profile likelihood is smooth with a minimum near to the estimate.

Case Study

In this section we give two examples to demonstrate the insight and information that the BLA model can provide, using the model with parameter estimates given in Table 2. In the first example we use data from sensors that measure volumetric water content (VWC) of the soil (5TE sensors, Decagon Devices, Pullman, WA). A cluster of 12 sensors was installed at a depth of 10 cm as part of a larger sensor network on a grassland site at Hollin Hill in Yorkshire, Northern England. Although this was a grassland site and our BLA model was estimated for arable soil, the data are used here for illustration. Two measurements of soil bulk density were made from soil removed when the cluster of sensors was installed, and these were used to estimate total porosity assuming a mineral particle density of 2.65 g cm^{-3} (Hall *et al.,* 1977). We computed the mean VWC for all sensors in the cluster for each day from 1st January 2013 to the end of July of the same year. We then scaled the mean VWC to mean WFPS given the estimate of total porosity. Figure 4(a) shows these values. The horizontal line on the graph is at WFPS = 0.77, the value at which the boundary line model for N_2O emissions is largest. We call the term

$$\mathcal{W} \;=\; \exp\left\{-\beta_2 \left(\text{logit}(\vartheta) - \beta_1\right)^2\right\}, \tag{26}$$

the WFPS factor, a dimensionless quantity which is 1 when the WFPS, denoted by ϑ, allows maximum N_2O flux and less than 1 otherwise. This is plotted for each day in Figure 4(b). Solid symbols indicate that the soil is wetter than the optimum for N_2O flux, and open symbols where it is drier. Note the initial increase in the factor, which is caused by drying of the soil. Within the first 50 days there were some heavy rain events, which increased WPFS above 0.77. This caused marked transient reductions in the WPFS factor, but much of the time it was close to 1. From about day 80 there was an overall decline in the WFPS factor because of drying of the soil with a few episodic increases in the factor that resulted from heavy rain. This analysis illustrates how potential N_2O emissions from soil vary temporally. During the winter months illustrated here, the WFPS factor was mostly close to 1, with episodic reductions because of inadequate air-filled pore space to allow gaseous diffusion out of the soil. Applied fertilizer N during this period is more likely to be lost to denitrification than during the spring when the WFPS factor shows a downward trend as the soil becomes too dry for the development of anaerobic centres (open symbols in Figure 4(b)). This provides a basis for the interpretation of sensor data for improved nitrogen management, or to develop generalized regional guidance on timing of applications or refinement of emission factors to account for regional weather patterns and soil conditions.

In our second example we used the BLA model, presented in Table 2, to examine the variation between potential rates of N_2O emission in contrasting soil series. We assumed in all cases that the soil was at field capacity (assumed to be a tension of 5kPa) and we used values of porosity and VWC at field capacity reported in soil survey memoirs. For the Cuckney series, a coarse loamy sand (Jarvis *et al.,* 1984) the WFPS is 0.38 at field capacity. The corresponding value of the WFPS factor, computed with Equation (26), is 0.2. We can infer, therefore, that the soil is too-well aerated at field capacity because of its coarse texture for the widespread development of anaerobic centres where denitrification can take place. In contrast the Formby series, a loamy medium sand (Fordham, 1986) has a WPFS of 0.75 at field capacity. The corresponding value of the WFPS factor is 0.99; the soil is close to the optimum for N_2O emission with respect to water content, wet enough for the development of anaerobic centres, but not so wet that it would inhibit diffusion of N_2O out of the soil. The Ragdale series, a stagnogley in chalky clay drift (Burton, 1986) has a WFPS of 0.86 at field capacity, the WFPS factor is 0.80. Because the WFPS is larger than 0.77, we can infer that the emission of N_2O is somewhat limited by the slow rate of gaseous diffusion from the soil. These calculations give insight into how physical differences between soil types affect their potential for N_2O emission when they are all at a standard water potential. Although the WFPS factor itself is not an emission factor for the various soil series (other potential limiting factors may apply), it could be used to rescale standard IPCC emission factors for soil to reflect variation between soils with different hydraulic properties that result from differences in texture inherited from contrasting parent materials.

Discussion

Our analysis shows that the boundary line is a plausible model for our data. The convex hull test provided evidence to reject the null hypothesis that the data were from a bivariate normal distribution, and the AIC for the boundary line model was smaller than that for the bivariate normal model. As noted in the previous section, the parameters for the fitted model were consistent with those reported elsewhere for the WFPS at which the maximum rate of emission occurs, and with an estimate of measurement error for our laboratory protocol. This consistency with previous empirical results and results from process modelling is encouraging and indicates that the BLA concept and our methods are plausible ways to model this particular response of the soil system.

Our analysis provides empirical support for the process model developed by

Rabot *et al.* (2015). It could also be used for the development of empirical models such as those of Conen *et al.* (2000) and Smith & Massheder (2014) which use the BLA concept. The calculation of N_2O emissions in the current methodology of the Intergovernmental Panel on Climate Change (IPCC) does not take account of variation in climate or soil, which limits its usefulness because, for example, interventions to reduce the application of fertilizer nitrogen to soil with a WFPS close to the optimum for N_2O emission would not affect inventory calculations. The use of process models to improve this is limited by their requirement for soil information. The BLA methodology presented here to estimate parameters of models in the style of Conen *et al.* (2000) could provide a basis for modelling in the IPCC framework because it takes better account of soil variation and our understanding of its effects. The case studies presented in the previous section illustrate how the BLA models could be used to explore how soil with contrasting physical properties might differ with respect to the likely rate of N_2O emission, and how to interpret temporal data on soil moisture content with respect to the likely effect on the rate of N_2O emission. This could provide a basis for refined advice on fertilizer use and timing, and improved emission rates for IPCC inventory.

Our boundary line model describes the limiting effect of a single explanatory variable, but more than one variable might be potentially limiting on a soil response. For example, temperature and pH might limit the rate of N_2O emission from soil. In a situation where we consider more than one limiting factor, the response of the soil system might be described by von Liebig's law of the minimum (von Liebig, 1863)

$$y = \min\{f_1(x_1), f_2(x_2), ..., f_n(x_n)\}, \tag{27}$$

where y is the response variable and x_i are the independent variables that limit y according to the functions f_i where $i = 1, ..., n$. Under the law of the minimum one of these factors is limiting in any one case, and the jth factor is limiting in some case if

$$y = \min\{f_1(x_1), f_2(x_2), ..., f_n(x_n)\} = f_j(x_j). \tag{28}$$

This hypothesis could be tested for a suite of candidate limiting factors to determine which is limiting for each observation of the response variable. This would require an appropriate inferential method to compare the limiting factors for any observation in terms of the distance of the observation from each respective boundary line. This is a topic for further work.

One way to develop the law of the minimum is to model interactions between different variables we observe that determine possible limiting effects on the soil response. For example, the WFPS and soil organic carbon content might interact to determine the 'anaerobiosis' limiting factor, whereas the nitrate and soil organic carbon might interact to determine a 'substrate' limiting factor. In this case the law of the minimum might be written as

$$y = \min\left\{f_{\text{anaerobiosis}}(x_1, x_2), f_{\text{substrate}}(x_3, x_4), ...\right\}. \quad (29)$$

In this latter case, the boundary-line model must be extended to three or more dimensions with a boundary plane described by each function within the braces on the right-hand side of Equation (29).

In the log-likelihood function for the BLA model given in Equation (20) the observations are treated as independent. This assumption requires independent random sampling to be fully justified, which was not the case in our example, and will not be true for many studies where data are obtained on regular grids or transects. The extension of the boundary line model to the situation with spatial dependence could be based on a linear model of coregionalization (Journel & Huijbregts, 1978) for the latent normal variate $\{y, x\}$, which underlies the boundary model. However, the derivation of the likelihood function for the model parameters under this joint distribution remains a challenge for further research.

Conclusions

The BLA approach with the censored normal model is an attractive method to estimate the limiting effects of soil factors on rates of N_2O emission, and the BLA model that we have fitted and presented is consistent with previous modelling, and experimental results. Our BLA method has a theoretical basis that enables us to test the evidence for a boundary response, and to quantify the uncertainty of model parameters with confidence intervals. These give us an insight into the precision of the parameter estimates and so can be used to assess the uncertainty in any predictions of rates of N_2O emission.

We have shown how the WFPS factor, derived from the BLA model and with parameters estimated from data, can be used to interpret real-time data on soil water content, and to indicate whether and how this variable can be expected to limit rates of N_2O emission at particular times. We have also shown the possibility of using the model to scale the emission factors for N_2O from contrasting types of soil, which offers a way to improve the greenhouse gas inventory in the IPCC framework. Further work is needed on the BLA methodology to account for spatial dependence and to fit more general models for potential limiting effects of several factors.

Acknowledgements

This paper is published with the permission of the Director of the British Geological Survey (NERC). We are grateful to Dr Barry Rawlins for the sensor data from Hollin Hill. We have no conflicts of interest to declare.

References

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory* (ed. B.N. Petov & F. Csaki), pp. 267–281. Akademia Kiado, Budapest.

Ball, B.C., Horgan, G.W. & Parker, J.P. 2000. Short-range spatial variation of nitrous oxide fluxes in relation to compaction and straw residues. *European Journal of Soil Science*, **51**, 607–616.

Bateman, E.J. & Baggs, E.M. 2005. Contributions of nitrification and denitrification to N2O emissions from soils at different water-filled pore space. *Biology and Fertility of Soils*, **41**, 379–388.

Bayer, C., Gomes, J., Zanatta, J.A., Costa, F., Viera, B., de Cassia Piccolo, M., Dieckow, J. & Six, J. 2015. Soil nitrous oxide emissions as affected by long-term tillage, cropping systems and nitrogen fertilization in Southern Brazil. *Soil and Tillage Research*, **146**, 213–222.

Benjamini, Y. & Yekutieli, D. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165–1188.

Bergstrom, D.W. & Beauchamp, E.G. 1993. Relationship between denitrification rate and determinant soil properties under barley. *Canadian Journal of Soil Science*, **73**, 567–578.

Broyden, C. G. 1970. The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and its Applications*, **6**, 76–90.

Burton, R.G.O. 1986. *Soils in Cambridgeshire III*, Sheet TL54 (Linton), Soil Survey Record No. 94, Soil Survey of England and Wales, Harpenden.

Conen, F., Dobbie, K.E. & Smith, K.A. 2000. Predicting N2O emissions from agricultural land through related soil parameters. *Global Change Biology*, **6**, 417–426.

Delwiche, C.C. 1981. The nitrogen cycle and nitrous oxide. In: *Denitrification, Nitrification and Atmospheric Nitrous Oxide* (ed. C.C. Delwiche), pp 1–15, John Wiley & Sons, New York.

Del Grosso, S.J., Parton, W.J., Mosier, A.R., Walsh, M.K., Ojima, D.S., & Thorton, P.E. 2006. DAYCENT National-Scale Simulations of Nitrous Oxide Emissions from Cropped Soils in the United States. *Journal of Environmental Quality*, **35**, 1451–1460.

Dobbie, K. E. & Smith, K. A. (2003), Nitrous oxide emission factors for agricultural soils in Great Britain: the impact of soil water-filled pore space and other controlling variables. *Global Change Biology*, **9**, 204–218.

Dobson, A.J. 1990. *An Introduction to Generalized Linear Models*, Second Edition Chapman and Hall, London.

Eddy, W.F. 1982. Convex hull peeling. In: *Compstat 1982, Part 1 Proceedings in Computational Statistics* (eds H. Caussinus, P. Ettinger, R. Tomassone), pp. 42–47. Physica Verlag, Heidelberg.

Elliot, J.A. & de Jong, E. 1993. Prediction of field denitrification rates: A boundary-line approach. *Soil Science Society of America Journal*, **57**, 82–87.

Farquharson, R. & Baldock, J. 2008. Concepts in modelling N_2O from land use. *Plant and Soil*, **309**, 147–167.

Fordham, S.J. 1986. *Soils in Surrey I*, Sheet TQ05 (Woking), Soil Survey Record No. 90, Soil Survey of England and Wales, Harpenden.

Franzluebbers, A.J. 1999. Microbial activity in response to water-filled pore space of variably eroded southern Piedmont soils. *Applied Soil Ecology*, **11**, 91–101.

Groffman, P.M., Tiedje, J.M., Robertson, G.P. & Christensen, S. 1988. Denitrification at different temporal and geographical scales: proximal and distal controls. In: *Advances in Nitrogen Cycling in Agricultural Ecosystems* (ed. J.R. Wilson), pp. 174–192. Commonwealth Agricultural Bureaux, Perth, Australia.

Hall, D.G.M., Reeve, M.J., Thomasson, A.J. & Wright, V.F. 1977. *Water retention, porosity and density of field soils. Soil Survey Technical Monograph No. 9.* Soil Survey of England and Wales, Harpenden.

Haskard, K.A., Welham, S.J. & Lark, R.M. 2010. A linear mixed model with spectral tempering of the variance parameters for nitrous oxide emission rates from soil across an agricultural landscape. *Geoderma*, **159**, 358–370.

Hatch, D.J., Jarvis, S.C. & Philipps, L. 1990. Field measurement of nitrogen mineralization using soil core incubation and acetylene inhibition of nitrification. *Plant and Soil*, **124**, 97–107.

Hueter, I. 1994 The convex hull of a normal sample. *Advances in Applied Probability*, **26**, 855–875.

Jarvis M.G., Allen, R.H., Fordham, S.J., Hazelden, J., Moffat, A.J. & Sturdy, R.G. 1984. *Soils and their Use in South East England. Soil Survey of England and Wales Bulletin No 15.* Lawes Agricultural Trust (Soil Survey of England and Wales), Harpenden.

Journel, A.G. & Huijbregts, Ch.J. 1978. *Mining Geostatistics.* Academic Press, San Diego CA.

Lark, R.M., Milne, A.E., Addiscott, T.M., Goulding, K.W.T, Webster, C.P. & O'Flaherty, S. 2004. Scale- and location-dependent correlation of nitrous oxide emissions with soil properties: an analysis using wavelets. *European Journal of Soil Science*, **55**, 611–627.

Li, C. 2000. Modeling trace gas emissions from agricultural ecosystems. *Nutrient Cycling in Agroecosystems*, **58**, 259–276.

Linn, D.M., & Doran, J.W. 1984. Effect of water-filled pore space on carbon dioxide and nitrous oxide production in tilled and nontilled soils. *Soil Science Society of America Journal*, **48**, 1267–1272.

von Liebig J. 1863. *The Natural Laws of Husbandry.* Walton and Maberly, London.

Milne, A.E., Wheeler, H.C. & Lark, R.M. 2006a. On testing biological data for the presence of a boundary. *Annals of Applied Biology* **149**, 213–222.

Milne, A.E., Ferguson, R.B. & Lark, R.M. 2006b. Estimating a boundary line model for a biological response by maximum likelihood. *Annals of Applied Biology* **149**, 223–234.

Minasny, B., McBratney, A.B. & Bristow, K.L. 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. *Geoderma*, **93**, 225–253.

R Core Team 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rabot, E., Cousin, I. & Hénault, C. 2015. A modeling approach of the relationship between nitrous oxide fluxes from soils and the water-filled pore space. *Biogeochemistry*, **122**, 395–408.

Ryden, J.C., Skinner, J.H. & Nixon, D.J. 1987. Soil core incubation system for the field measurement of denitrification using acetylene inhibition. *Soil Biology & Biochemistry*, **19**, 753–757.

Schmidt, U., Thöni, H. & Kaupenjohann, M. 2000. Using a boundary line approach to analyze N_2O flux data from agricultural soils. *Nutrient Cycling in Agroecosystems*, **57**, 119–129.

Schnug E., Heym J. & Murphy D.P. 1996. Establishing critical values for soil and plant analysis by means of the Boundary Line Development System (BOLIDES). *Communications in Soil Science and Plant Analysis*, **27**, 2739–2748

Shatar T.M. & McBratney A.B. 2004. Boundary-line analysis of field-scale yield response to soil properties, *Journal of Agricultural Science*, **142**, 1–7.

Schulze, E.D., Luyssaert, S. , Ciais, P., Freibauer, A., Janssens, I.A. *et al.* 2009. Importance of methane and nitrous oxide for Europe's terrestrial greenhouse-gas balance. *Nature Geoscience*, **2**, 842–850.

Smith, K.A. & Massheder, J. 2014. Predicting nitrous oxide emissions from N-fertilized grassland soils in the UK from three soil variables, using the B-LINE 2 model. *Nutrient Cycling in Agroecosystems*, **98**, 309–326.

Smith, K.A., Thomson, P.E., Clayton, H., McTaggart, I.P. & Conen, F. 1998. Effects of temperature, water content and nitrogen fertilisation on emissions of nitrous oxide by soils. *Atmospheric Environment*, **32**, 3301–3309.

Stevens, R.J., Laughlin, R.J., Burns, L.C., Arah, J.R.M. & Hood, R.C. 1997. Measuring the contributions of nitrification and denitrification to the flux of nitrous oxide from soil. *Soil Biology & Biochemistry*, **29**, 139–151.

Tabatabai, M.A. & Bremner, J.M. 1991. Automated instruments for determinations of total carbon, nitrogen, and sulfur in soils by combustion techniques. In: *Soil analysis: modern instrumental techniques*, 2nd Edition (ed. K.A. Smith), pp. 261–286. Marcel Dekker, New York.

Turban, S. 2010. Convolution of a truncated normal and a centred normal variable. `http://www.columbia.edu/~st2511/notes/Convolution%20of%20truncated%20normal%20and%20normal.pdf` Accessed 25th November 2015.

Verbeke, G. & Molenberghs, G. 2000. *Linear Mixed Models for Longitudinal Data.* Springer, New York.

Webb R.A. 1972. Use of the boundary line in analysis of biological data. *Journal of Horticultural Science*, **47**, 309–319

Webster, C.P. & Goulding, K.W.T. 1989. Influence of soil carbon content on denitrification from fallow land during autumn. *Journal of the Science of Food and Agriculture*, **49**, 131–142.

Wu, L., Rees, R.M., Tarsitano, D., Zhang, X., Jones, S.K. & Whitmore, A.P. 2015. Simulation of nitrous oxide emissions at field scale using the SPACSYS model. *Science of the Total Environment*, **530–531**, 76–86.

Table 1. Exploratory statistics.

	N_2O emission rate/ g N ha^{-1} day^{-1}	log N_2O emission rate/ log g N ha^{-1} day^{-1}	Water-filled pore space/ logit (cm^3 cm^{-3})	Soil organic carbon/ % by mass
Mean	65.8	3.56	0.74	2.40
Median	43.0	3.76	0.69	2.46
Quartile 1	15.0	2.71	0.42	2.09
Quartile 3	89.5	4.49	1.04	2.75
Minimum	0.5	−0.69	−0.33	1.21
Maximum	333.0	5.81	2.46	4.68
SD	67.8	1.29	0.47	0.50
skewness	1.6	−0.66	0.61	−0.05

Number of upper vertices in successive convex hull peels

Hulls	Expected number of upper vertices	Observed number	P-value*
1–5	37	44	3.0E–3
1–6	45	56	8.0E–5
1–7	53	63	1.0E–3
1–8	62	71	2.5E–3
1–9	70	80	2.0E–3
1–10	78	87	6.7E–3

* Null hypothesis: vertices arise from a bivariate normal process. With false discovery rate control at 0.05 all null hypotheses are rejected.

Table 2. Boundary line fitting results. The parameters β_0, β_1 and β_2 are parameters of the boundary line model given in Equation (24), and σ_{e} is the standard deviation of the measurement error, given in Equation (3). The negative log residual-likelihood and number of parameters in each model are ℓ and $\mathcal{P}$ respectively. The AIC is defined in Equation (25).

Boundary model

Parameter	Estimate	Standard error	95% confidence interval
β_0	4.99	0.24	$[4.52, 5.46]$
β_1	1.19	0.21	$[0.79, 1.61]$
β_2	0.54	0.22	$[0.11, 0.97]$
σ_{e}	0.53	0.09	$[0.35, 0.71]$

Comparison with multivariate normal model

	Boundary model	Multivariate normal
ℓ	560.3	580.2
$\mathcal{P}$	9	5
AIC	1138.7	1170.5

Figure Captions

1. Summary plots of (transformed) rates of emission and water-filled pore space. Scatter plots of log rate of N_2O emission against (a) WFPS on the original scale and (b) WFPS on the logit scale. Histograms of (c) log rate of N_2O emission and (d) logit of WFPS

2. Fitted boundary model for (transformed) rates of emission and water-filled pore space. The dotted line shows the 95% confidence interval for the boundary line

3. Profile likelihood for standard deviation of measurement error. Values $\hat{\sigma}_\mathrm{e}$ and $\hat{\sigma}_\mathrm{d}$ are, respectively, the maximum likelihood estimate and the estimated between-duplicate standard deviation of data collected in the study of Haskard *et al.* (2010). The vertical dotted line shows the 95% confidence interval of $\hat{\sigma}_\mathrm{d}$

4. (a) Daily mean water-filled pore space for a cluster of 12 sensors at Hollin Hill, N. Yorkshire, from 1st January 2013 to late July in the same year. The horizontal line is at 0.77 at which the WFPS factor, see Equation (26), is largest. (b) WFPS factor, see Equation (26), plotted with open symbols where the water- filled pore space is less than 0.77 and closed symbols where it is larger or equal to 0.77.

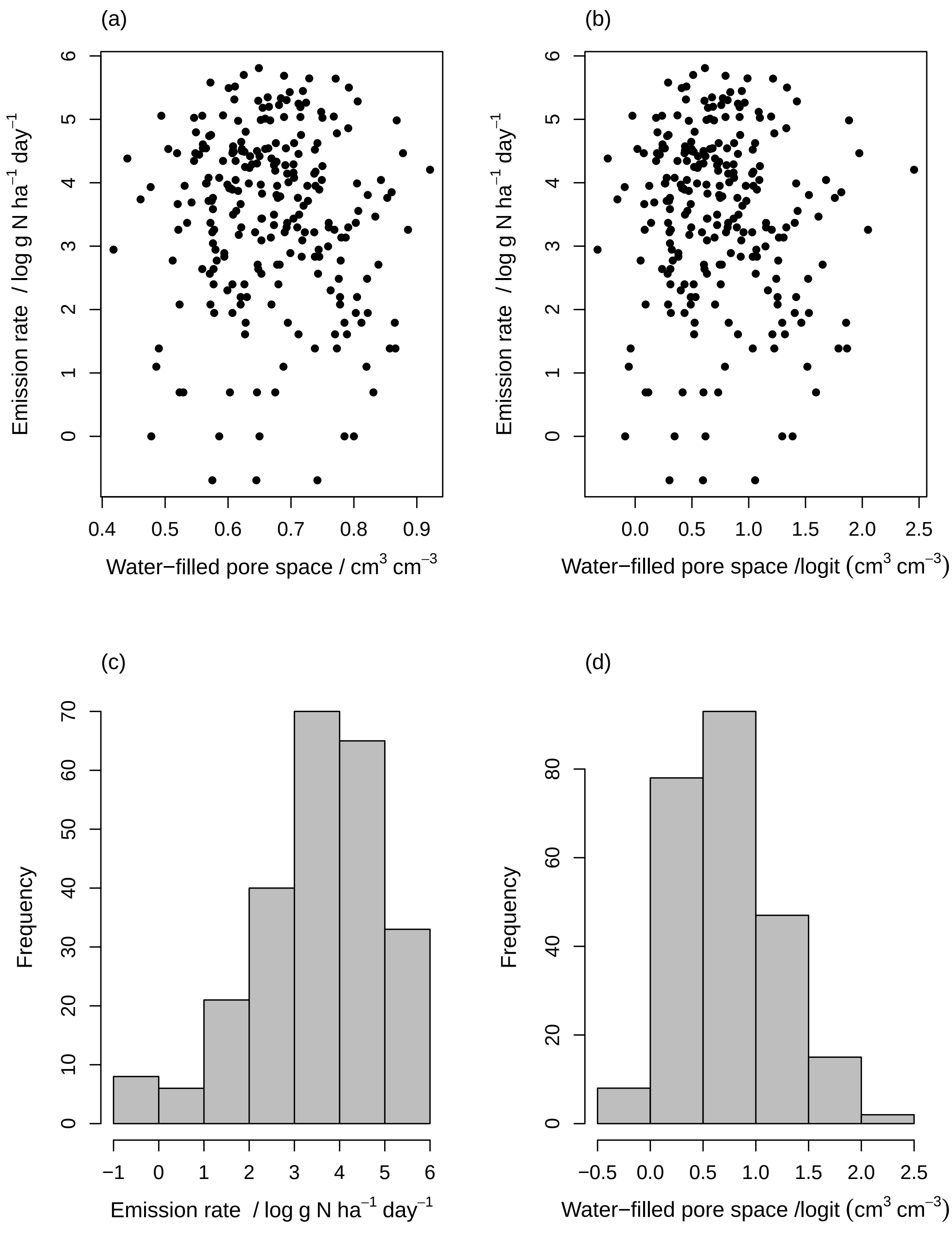
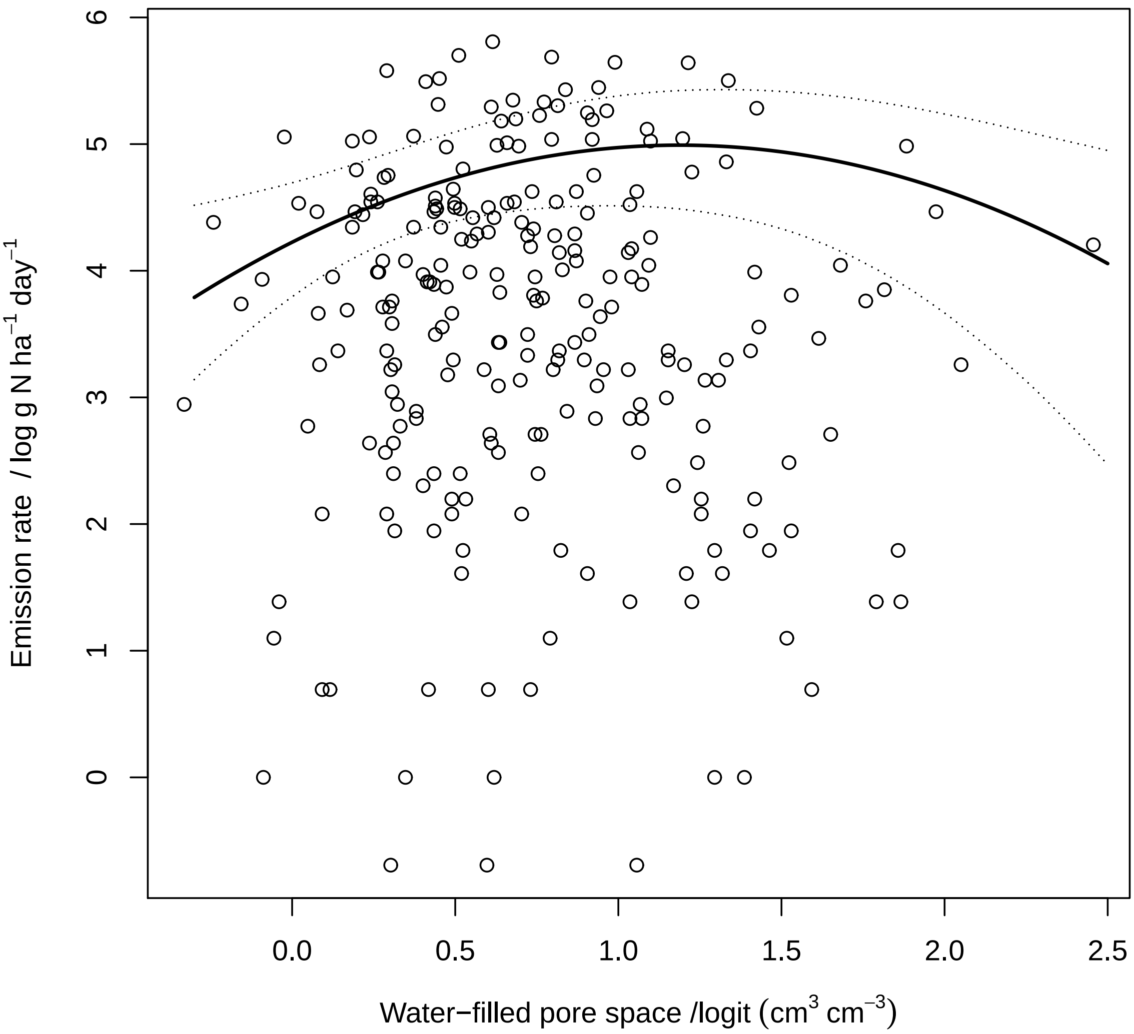
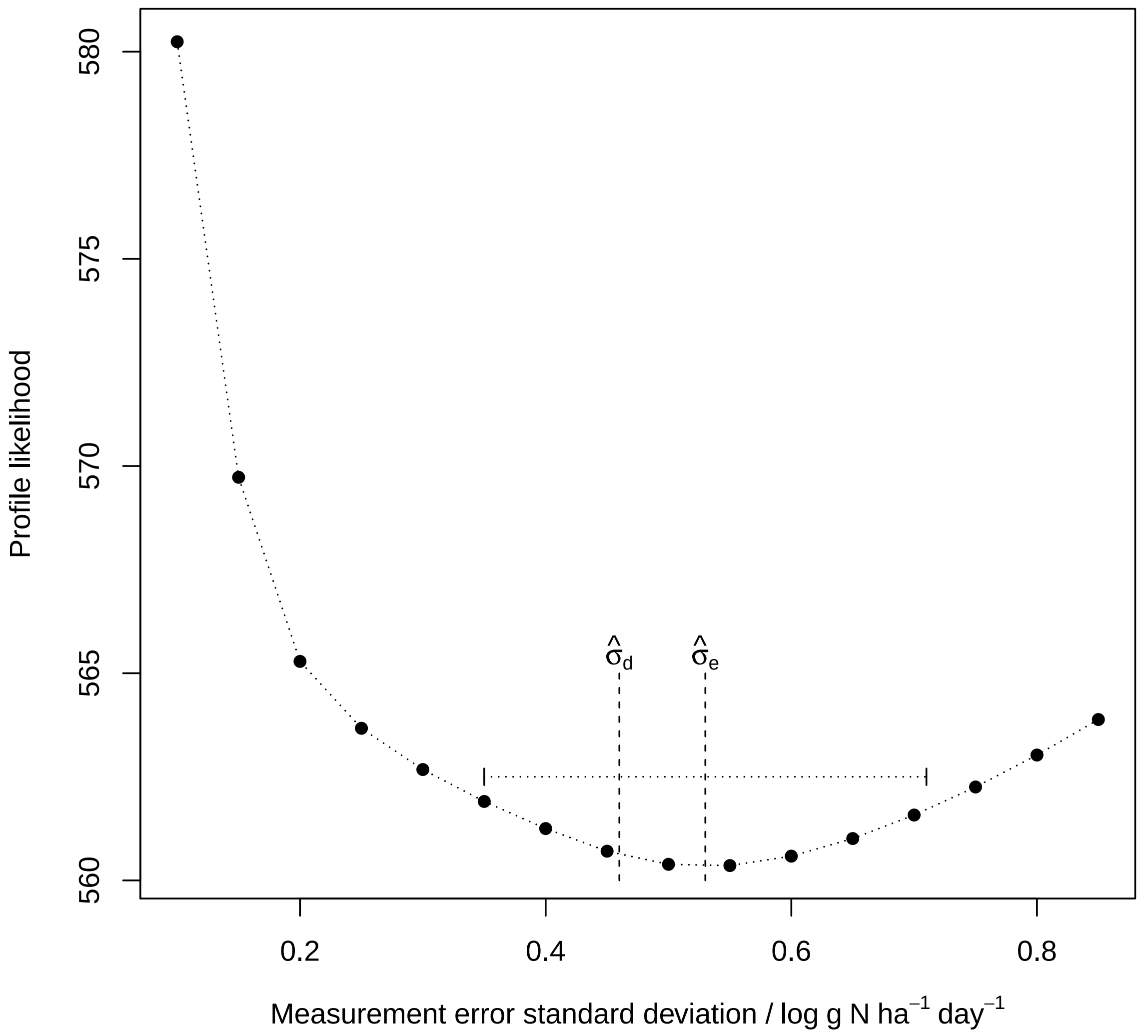

(a)
Emission rate / log g N ha⁻¹ day⁻¹
Water−filled pore space / cm³ cm⁻³
(b)
Emission rate / log g N ha⁻¹ day⁻¹
Water−filled pore space /logit (cm³ cm⁻³)
(c)
Frequency
Emission rate / log g N ha⁻¹ day⁻¹
(d)
Frequency
Water−filled pore space /logit (cm³ cm⁻³)

Figure 1:



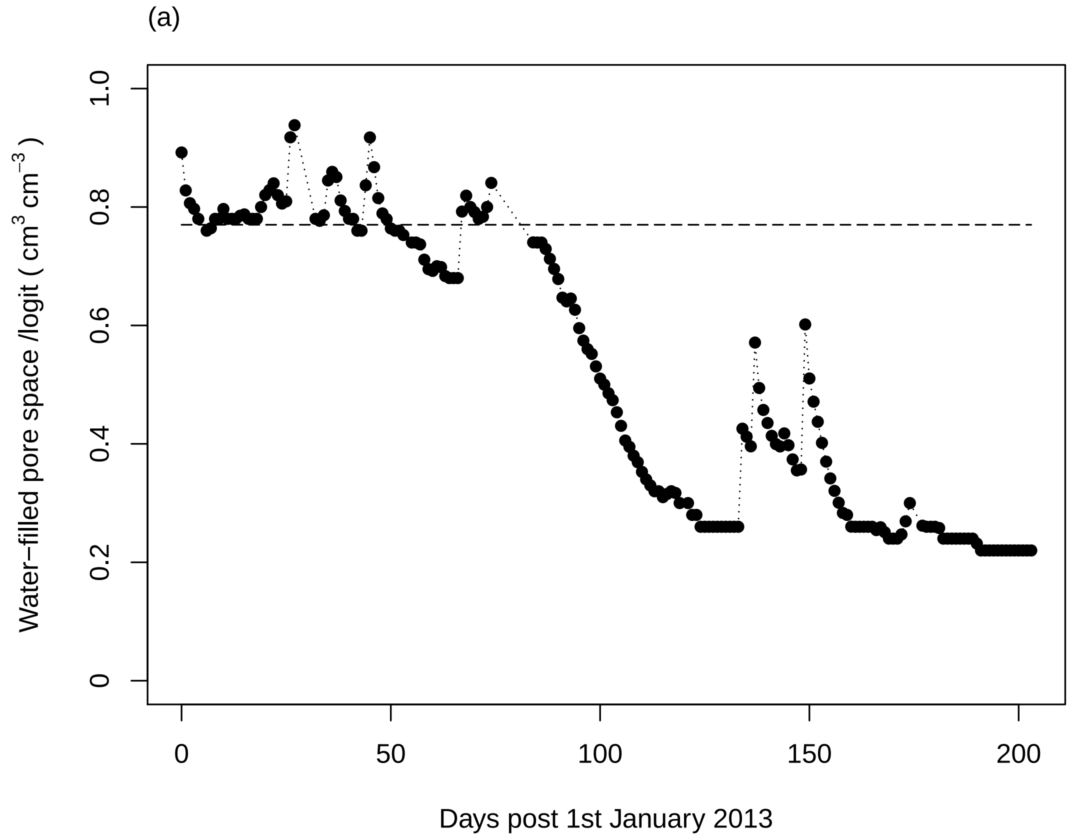
Emission rate / log g N ha^{-1} day^{-1}
6
5
4
3
2
1
0
0.0
0.5
1.0
1.5
2.0
2.5
Water−filled pore space /logit (cm^{3} cm^{-3})

Figure 2:



Profile likelihood
580
575
570
565
560
0.2
0.4
0.6
0.8
Measurement error standard deviation / log g N ha^{-1} day^{-1}
$\hat{\sigma}_d$
$\hat{\sigma}_e$

Figure 3:


(a)
Water−filled pore space /logit (cm3 cm−3)
Days post 1st January 2013
0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200

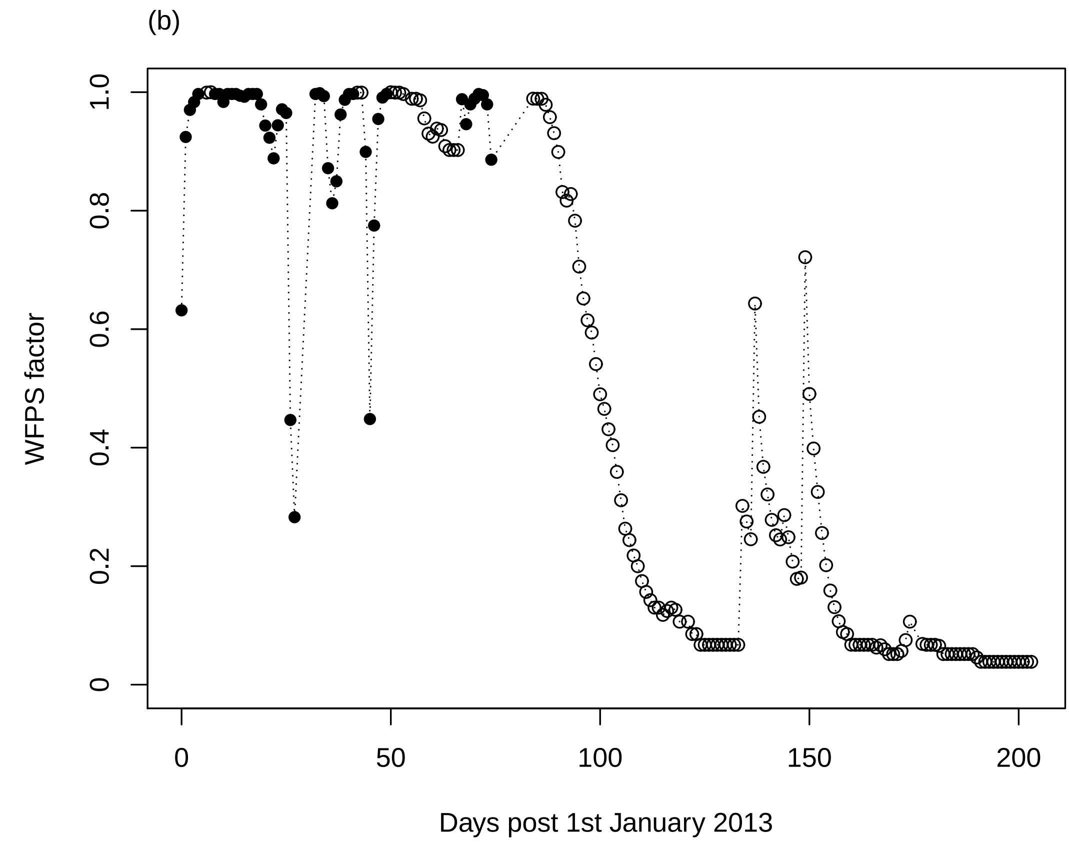

(b)
WFPS factor
Days post 1st January 2013
0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200

Figure 4: