

Article

Information Properties of Boundary Line Models for N₂O Emissions from Agricultural Soils

Cairistiona F.E. Topp^{1,*}, Weijin Wang², Joanna M. Cloy³, Robert M. Rees³ and Gareth Hughes^{1,*}

¹ Crop and Soil Systems, SRUC, West Mains Road, Edinburgh, EH9 3JG, UK

² Division of Science Delivery, Department of Science, IT, Innovation and the Arts, GPO Box 5078, Brisbane, QLD 4001, Australia; E-Mail: Weijin.Wang@qld.gov.au (W.W.)

³ Carbon Management Centre, SRUC, West Mains Road, Edinburgh EH9 3JG, UK; E-Mails: Joanna.Cloy@sruc.ac.uk (J.M.C); Bob.Rees@sruc.ac.uk (R.M.R.)

* Authors to whom correspondence should be addressed; E-Mails: Kairsty.Topp@sruc.ac.uk (C.F.E.T.); Gareth.Hughes@sruc.ac.uk (G.H.).

Received: 15 January 2013; in revised form: 25 February 2013 / Accepted: 27 February 2013 /

Published: 5 March 2013

Abstract: Boundary line models for N₂O emissions from agricultural soils provide a means of estimating emissions within defined ranges. Boundary line models partition a two-dimensional region of parameter space into sub-regions by means of thresholds based on relationships between N₂O emissions and explanatory variables, typically using soil data available from laboratory or field studies. Such models are intermediate in complexity between the use of IPCC emission factors and complex process-based models. Model calibration involves characterizing the extent to which observed data are correctly forecast. Writing the numerical results from graphical two-threshold boundary line models as 3×3 prediction-realization tables facilitates calculation of expected mutual information, a measure of the amount of information about the observations contained in the forecasts. Whereas mutual information characterizes the performance of a forecaster averaged over all forecast categories, specific information and relative entropy both characterize aspects of the amount of information contained in particular forecasts. We calculate and interpret these information quantities for experimental N₂O emissions data.

Keywords: nitrous oxide; emissions; boundary line model; agricultural soils; expected mutual information; normalized mutual information; specific information; relative entropy

1. Introduction

Over the last few decades, the atmospheric concentration of nitrous oxide (N₂O) has been increasing. Agricultural soils are a significant source of N₂O emissions [1]. Under the Kyoto Protocol, over 170 nations agreed to develop national inventories of anthropogenic emissions. These are calculated using the IPCC emission factors [2], which assume that the annual N₂O emissions from agricultural soils are proportional to the N applied, and which are mostly based on emission factors developed by Bouwman [3]. This is the simplest approach to calculating emissions.

However, the rate of N₂O emissions is affected by soil characteristics and climate [4–8], soil management [1,4,9] and the crops planted [4,10,11]. Thus, in order to incorporate these factors into estimates of N₂O emissions, more complex process-based models have been developed. These include DAYCENT [12] and DNDC [13,14]. Although it is recognized that they have the ability to test different management and mitigation options, these models require large amounts of input data and need to be calibrated for different agricultural systems.

The use of empirical models provides an approach to estimating N₂O emissions intermediate in complexity between the IPCC emission factor approach and process-based models [15]. The idea is to use soil and climate data to estimate the corresponding level of N₂O emissions [15]. Conen *et al.* [15] and Wang and Dalal [16] have used a boundary line model approach, based on water filled pore space, soil temperature and soil mineral N, to determine levels of N₂O emissions from agricultural soils. Calibration for different agricultural systems is still required, but acquisition of the relevant input data is simpler than for process-based models.

In writing for this special issue of *Entropy on Applications of Information Theory in the Geosciences*, our emphasis is on practical applications. That is to say, we are interested in applications that relate to the analysis of models and observed data, and we are writing for colleagues in the experimental geosciences who, it is to be hoped, are attracted to this special issue by the prospect of finding, then assimilating and utilizing such applications. In particular, we will use an information theoretic analysis to address the properties of boundary line model forecasts of N₂O emissions from agricultural soils.

2. Models and Data

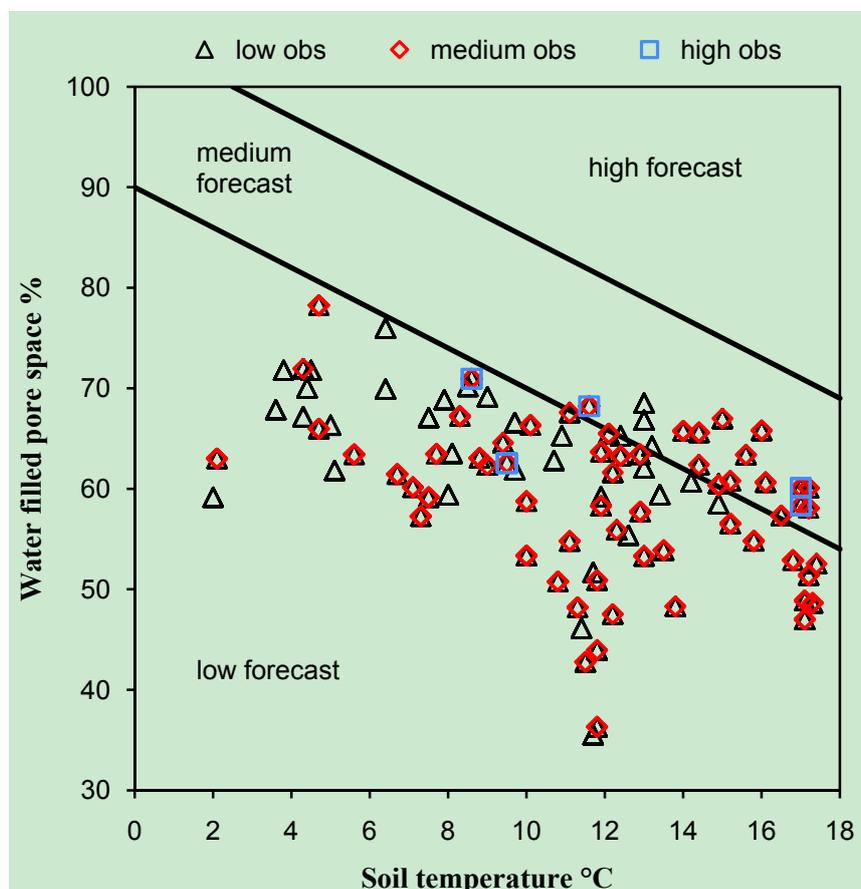
2.1. Boundary Line Models

The response variable of interest here is N₂O emission (or ‘flux’, g N₂O-N ha⁻¹ day⁻¹). For the kind of boundary line model under consideration here, a two-dimensional region of parameter space is delimited by appropriate ranges (considering the observed data) of two continuous explanatory variables, soil water-filled pore space (WFPS, %) and soil temperature (T, °C). The parameter space is partitioned into sub-regions by means of thresholds based on relationships between N₂O flux and the explanatory variables; two such thresholds partition the parameter space into three sub-regions of forecast N₂O flux, denoted ‘low’, ‘medium’ and ‘high’.

For an introductory example, Figure 1 shows observed data from an intensive study of N₂O emissions from sandy loam grassland soils at a site in Dumfries (SW Scotland) between March 2011 and March 2012, in which inorganic N fertilizer (ammonium nitrate and urea) treatments at a range of

application rates (0, 80, 160, 240, 320, 400 kg/ha N) and additional 320 kg/ha N plus the nitrification inhibitor DCD were used. In Figure 1, the observed data are superimposed on the partitioned parameter space using the boundary lines for forecast N₂O emissions as calculated in Conen *et al.* [15]. In this case, although the majority of the observed low emissions (<10 g N₂O-N ha⁻¹ day⁻¹) were correctly forecast, only a minority of observed medium emissions (10-100 g N₂O-N ha⁻¹ day⁻¹) and no observed high emissions (>100 g N₂O-N ha⁻¹ day⁻¹) were correctly forecast. Here, these data serve only to make the point that it is important to be able to characterize the extent to which observed levels of N₂O flux are correctly or incorrectly forecast by a boundary line model. Although these data are not analyzed further, we will discuss in Section 5 how other data sets present similar issues in terms of estimating N₂O emissions from agricultural soils within defined ranges.

Figure 1. The parameter space delimited by observed ranges of water filled pore space (WFPS, %) and soil temperature (T, °C) in 2011-2012 at a grassland site in Dumfries, SW Scotland, receiving inorganic fertilizer, is the basis for a boundary line model. Observed N₂O emissions were categorized as ‘low’ (<10 g N₂O-N ha⁻¹ day⁻¹), ‘medium’ (10-100 g N₂O-N ha⁻¹ day⁻¹) or ‘high’ (>100 g N₂O-N ha⁻¹ day⁻¹), as in Conen *et al.* [15]. There were 715 ‘low’ observations, 322 ‘medium’ observations and 19 ‘high’ observations (*N* = 1056), resulting in many overlapping data points on the graph. The boundary lines between forecast emission categories are $WFPS(\%) + 2 \cdot T(^{\circ}C) = 90$ (low-medium) and $WFPS(\%) + 2 \cdot T(^{\circ}C) = 105$ (medium-high), as described in Conen *et al.* [15].



At this stage, before discussing the details of any of the information quantities that may be used to analyze forecasts of N₂O emissions, we note that the boundary line approach (as described in [15,16]) is intrinsically suitable for such analysis. For example, Tribus and McIrvine [17] note: “In modern information theory, probabilities are treated as a numerical encoding of a state of knowledge. One’s knowledge about a particular question can be represented by the assignment of a certain probability (denoted p) to the various conceivable answers to the question.” Thus the boundary line approach, as described, starts by characterizing the conceivable answers to the question of the magnitude of N₂O flux (‘low’, ‘medium’ and ‘high’). Note that this specification restricts our attention to discrete distributions of probabilities.

2.2. Data

The data analyzed in this study come from an assessment of the boundary line approach for forecasting N₂O emission ranges from Australian agricultural soils [16]. Observed emissions were categorized as ‘low’ (<16 g N₂O-N ha⁻¹ day⁻¹), ‘medium’ (16-160 g N₂O-N ha⁻¹ day⁻¹) or ‘high’ (>160 g N₂O-N ha⁻¹ day⁻¹). Boundary lines were calculated separately for pasture and sugarcane soils (Table 1) and for cereal cropping soils (Table 2, see also Figure 2 in [16]). For a boundary line plot in which two thresholds partition the parameter space into three sub-regions of forecast N₂O flux, denoted ‘low’, ‘medium’ and ‘high’, we can present the data in a 3×3 *prediction-realization table* in which the columns correspond to the observations, the rows to the forecasts. Theil [18] uses this terminology to refer both to the cross-tabulated frequencies of observations and forecasts, and to the estimated probabilities obtained by normalization of the frequencies. Here we present the normalized version of the data (Tables 1 and 2, based on Figure 2 in [16]).

We adopt the following notation for Table 1 and subsequently. The observed categories are denoted o_j ($j=1,2,3$) for ‘low’, ‘medium’ and ‘high’ N₂O flux categories, respectively. The bottom row of the table contains the distribution $\Pr(O)$. The forecast categories are denoted f_i ($i=1,2,3$) for ‘low’, ‘medium’ and ‘high’ N₂O flux categories, respectively. The right-hand margin of the table contains the distribution $\Pr(F)$. The body of the table contains the joint probabilities $\Pr(o_j \cap f_i)$. All values (here and throughout) are rounded to 4 d.p.

Table 1. Prediction-realization table based on $N=271$ observations of N₂O flux for pasture and sugarcane soils from Figure 2 of [16]. The boundary lines between forecast emission categories were $\text{WFPS}(\%)+0.71 \cdot T(^{\circ}\text{C}) = 63$ (low-medium) and $\text{WFPS}(\%)+0.71 \cdot T(^{\circ}\text{C}) = 75$ (medium-high).

Forecast category, f_i	Observed category, o_j			Row sums
	1. Low	2. Medium	3. High	
1. Low	0.2841	0.0554	0.0074	0.3469
2. Medium	0.1181	0.1513	0.0959	0.3653
3. High	0.0480	0.0812	0.1587	0.2878
Column sums	0.4502	0.2878	0.2620	1.0000

3. Analysis of Information Properties

Shannon's two papers, collected together as [19], represent the beginning of modern information theory; Cover and Thomas's text [20] provides a comprehensive overview of half a century of progress. In the present context, where we are more concerned with applications of information theory relating to the analysis of models and data than with the mathematical theory of communications, readers may also find useful background in Attneave [21], Theil [18] or Hughes [22].

3.1. Information Content

We write p ($=\Pr(E)$, $0 \leq \Pr(E) \leq 1$) to denote a generic probability of an event E . Then $h(p) = \log(1/p) = -\log(p)$ is the *information content* of a message that tells us, without error, that E has occurred (thus, in the present context, such a message constitutes a perfect forecast). If p is small, the information content of this message is large, and vice versa. If $p=0$ (the event is impossible) and you tell us that E has definitely occurred, the information content of this message is indefinitely large; if $p=1$ (the event is certain) and you tell us that E has definitely occurred, the information content of this message is zero. The base of the logarithm serves to define the units of information. In the mathematical theory of communications, logarithms base 2 are often used and the unit of information is the *bit*. Here we will use natural logarithms for calculations, and refer to the unit of information as the *nit* [23]. We will write \log in places where the reference is generic.

Now consider the observed N_2O flux categories o_1 ('low'), o_2 ('medium') and o_3 ('high'), with corresponding probabilities $\Pr(o_1)$, $\Pr(o_2)$ and $\Pr(o_3)$, $\sum_j \Pr(o_j) = 1$, $\Pr(o_j) \geq 0$, $j = 1, 2, 3$. We cannot calculate the information content $h(\Pr(o_j))$ until the message is received, because the message ' o_j occurred' may refer to any one of o_1 , o_2 or o_3 . We can, however, calculate *expected information content* before the message is received. This quantity, often referred to as the *entropy*, is the weighted average of the information contents of the possible messages. Since the message ' o_j occurred' is received with probability $\Pr(o_j)$, the expected information content, denoted $H(O)$, is:

$$H(O) = -\sum_j \Pr(o_j) \log(\Pr(o_j)) \quad (1)$$

We note that $H(O) \geq 0$ and take $\Pr(o_j) \log(\Pr(o_j)) = 0$ if $\Pr(o_j) = 0$, since $\lim_{x \rightarrow 0} x \log(x) = 0$. If any $\Pr(o_j) = 1$, $H(O) = 0$. This is reasonable since we expect nothing from a forecast if we are already certain of the actual outcome. $H(O)$ has its maximum value when all the $\Pr(o_j)$ have the same value. This is also reasonable, since a message that tells us what actually happened will have a larger information content when all outcomes are equally probable than when some outcomes are more probable than others. Providing an everyday-language metaphor by means of which to characterize entropy is no easy task. Tribus and McIrvine [17] give a brief account of Shannon's own difficulty in this respect. In the present context, entropy can be thought of as characterizing either information or uncertainty, depending on our point of view. At the outset, we know that just one of a number of events will occur, and the corresponding probabilities of the events. Entropy quantifies how much information we will obtain, on average, from a message that tells us what actually happened.

Alternatively, entropy characterizes the extent of our uncertainty prior to receipt of the message that tells us what happened.

Now, similarly, we can calculate the entropy of the distribution of forecast N₂O flux probabilities:

$$H(F) = -\sum_i \Pr(f_i) \log(\Pr(f_i)) \quad (2)$$

and the entropy of the joint probability distribution:

$$H(O, F) = -\sum_i \sum_j \Pr(o_j \cap f_i) \log(\Pr(o_j \cap f_i)) \quad (3)$$

Working in natural logarithms, we calculate from Table 1 [using Equations (1), (2) and (3), respectively] the expected information contents in nits as follows: $H(O) = 1.0687$, $H(F) = 1.0936$ and $H(O, F) = 1.9585$.

3.2. Expected Mutual Information

In the unfortunate situation that the forecasts were independent of the observations, we would have $H(O, F) = H(O) + H(F)$. When the joint entropy is smaller than the sum of the individual entropies, this indicates association between forecasts and observations. Then $H(O, F) = H(O) + H(F) - I_M(O, F)$, where the *expected mutual information*, denoted $I_M(O, F)$, is a measure of the association. To calculate $I_M(O, F)$ directly:

$$I_M(O, F) = \sum_i \sum_j \Pr(o_j \cap f_i) \log \left[\frac{\Pr(o_j \cap f_i)}{\Pr(o_j) \Pr(f_i)} \right] \quad (4)$$

and $I_M(O, F) \geq 0$, with equality only if O and F are independent. Working in natural logarithms, we calculate from Table 1 [using Equation (4)] the expected mutual information in nits: $I_M(O, F) = 0.2038$, and note also that $I_M(O, F) = H(O) + H(F) - H(O, F)$.

3.2.1. The G^2 -test

The correspondence between expected mutual information as estimated above and the χ^2 statistic was noted at least as far back as Attneave [21]. Essentially, $G^2 = 2 \cdot N \cdot I_M(O, F) \approx \chi^2$ (in which N is the total number of observations) (see, e.g., [24]).

3.2.2. Conditional Entropy

Returning to our prediction-realization table (Table 1), with probabilities $\Pr(o_j)$ and $\Pr(f_i)$ in the margins and $\Pr(o_j \cap f_i)$ in the body of the table, we note $\Pr(o_j \cap f_i) = \Pr(f_i|o_j)\Pr(o_j) = \Pr(o_j|f_i)\Pr(f_i)$ (Bayes' theorem). Recalling Equation (4), we can now write:

$$I_M(O, F) = \sum_i \sum_j \Pr(o_j \cap f_i) \log \left[\frac{\Pr(o_j|f_i)}{\Pr(o_j)} \right] \quad (5)$$

and then after some rearrangement:

$$I_M(O, F) = \sum_j \Pr(o_j) \log \left[\frac{1}{\Pr(o_j)} \right] - \sum_i \Pr(f_i) \sum_j \Pr(o_j | f_i) \log \left[\frac{1}{\Pr(o_j | f_i)} \right] \quad (6)$$

in which we recognize the first term on the right-hand side as the entropy $H(O)$. The second term is the conditional entropy $H(O|F)$. Thus we note $I_M(O, F) = H(O) - H(O|F)$. Numerically, we can calculate $\Pr(o_j | f_i) = \Pr(o_j \cap f_i) / \sum_j \Pr(o_j \cap f_i)$. For completeness, we can also write:

$$I_M(O, F) = \sum_i \Pr(f_i) \log \left[\frac{1}{\Pr(f_i)} \right] - \sum_j \Pr(o_j) \sum_i \Pr(f_i | o_j) \log \left[\frac{1}{\Pr(f_i | o_j)} \right] \quad (7)$$

in which we recognize the first term on the right-hand side as the entropy $H(F)$. The second term is the conditional entropy $H(F|O)$. Thus we note $I_M(O, F) = H(F) - H(F|O)$. Working in natural logarithms, we calculate from Table 1 the conditional entropies in nits: $H(O|F) = 0.8649$ and $H(F|O) = 0.8898$, and note also that $H(O, F) = H(O) + H(F|O) = H(F) + H(O|F)$.

We can interpret the expected mutual information $I_M(O, F) = H(O) - H(O|F)$ in terms of the average reduction in uncertainty about O resulting from use of a forecaster (*i.e.*, a predictive model) F . Suppose that we have a forecaster such that F and O are identical, so that use of the forecaster accounts for all the uncertainty in O . Then $H(O|F) = H(O|O)$ and $I_M(O, F) = H(O) - H(O|O) = H(O)$. This tells us that the maximum of the expected mutual information $I_M(O, F)$ between O and F , that would characterize a perfect forecaster, is the entropy $H(O)$. Also, we have $H(O) - H(O|F) = I_M(O, F) \geq 0$, so we must have $H(O|F) \leq H(O)$ with equality only if F and O are independent. Reassuringly, this tells us that *on average*, as long as F and O are not independent, use of a forecaster F will decrease uncertainty in O .

3.2.3. Normalized Mutual Information

We have seen that expected mutual information $I_M(O, F)$ varies between 0 (indicating that F and O are independent) and $H(O)$ (indicating that F is a perfect forecaster of O). Sometimes (for example, when making comparisons between analyses) it is useful to calculate a normalized version of expected mutual information. Some care is required here, as different normalizations have been documented in the literature. Here, following Attneave [21] (see also Forbes [25]), we adopt:

$$\text{normalized } I_M(O, F) = \frac{H(O) - H(O|F)}{H(O)} \quad (8)$$

as a measure of association that lies between 0 (indicating that F and O are independent) and 1 (indicating that F is a perfect forecaster of O). We can interpret the normalized version of $I_M(O, F)$ as a measure of the proportion of entropy in O explained by covariate F . From Table 1 [using Equation (8)], working in natural logarithms, we calculate normalized $I_M(O, F) = 0.1907$.

3.3. Specific Information

We have seen that on average, as long as F and O are not independent, use of a forecaster F will decrease uncertainty in O . However, although uncertainty is reduced on average, certain specific

forecasts may increase uncertainty. In the current example, we have $H(O) = 1.0687$, $H(O|F) = 0.8649$ and $I_M(O,F) = 0.1907$. For a specific forecast f_i , we have:

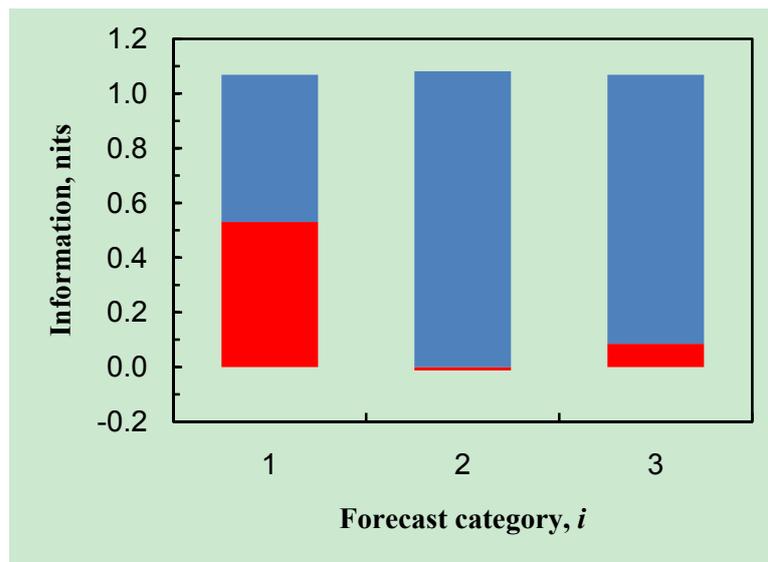
$$H(O|f_i) = -\sum_j \Pr(o_j|f_i) \log(\Pr(o_j|f_i)) \tag{9}$$

and from Table 1 [using Equation (9)], working in natural logarithms, we calculate $H(O|f_1) = 0.5382$, $H(O|f_2) = 1.0813$, and $H(O|f_3) = 0.9839$ nits. *Specific information*, denoted $I_S(f_i)$, is then:

$$I_S(f_i) = H(O) - H(O|f_i) \tag{10}$$

(DelSole [26] refers to this quantity as ‘predictive information’). This quantity may be positive (when $H(O) > H(O|f_i)$, in which case uncertainty has decreased), or negative (when $H(O) < H(O|f_i)$, in which case uncertainty has increased). For the present example, based on Table 1, the results are illustrated in Figure 2.

Figure 2. For each forecast category i , the bar comprises a red component $I_S(f_i)$, and a blue component $H(O|f_i)$ which together sum to $H(O)$ in each case. The weighted average of red components is equal to $I_M(O,F)$ (the $\Pr(f_i)$ provide the appropriate weights).



Expanding Equation (10), we have:

$$I_S(f_i) = \sum_j \Pr(o_j) \log\left[\frac{1}{\Pr(o_j)}\right] - \sum_j \Pr(o_j|f_i) \log\left[\frac{1}{\Pr(o_j|f_i)}\right] \tag{11}$$

and thus we can see that Equation (6) calculates the expected value of $I_S(f_i)$; that is to say, expected mutual information is expected specific information over all forecast categories:

$$I_M(O,F) = \sum_i \Pr(f_i) I_S(f_i) \tag{12}$$

3.4. Relative Entropy

Recall Section 3.1, where we defined the information content of a message that tells us, without error, that an event has occurred. In practice, not all the messages we receive will tell us without error that the event in question has occurred (see, e.g., Figure 1 above). So now we can think of a message as serving to transform a set of prior probabilities into a corresponding set of posterior probabilities. In this case, we can generalize as follows:

$$\begin{aligned}
 \text{information content of message} &= \log \left[\frac{\text{probability of event after message received}}{\text{probability of event before message received}} \right] \\
 &= \log[\Pr(E|message)] - \log[\Pr(E)] \\
 &= h[\Pr(E)] - h[\Pr(E|message)]
 \end{aligned}
 \tag{13}$$

which includes the previous example of the message that tells us without error what has occurred (i.e., $\Pr(E|message)=1$) as a special case. We now generalize Equation (1) in order to calculate expected information content for such a message.

Recall the observed N₂O flux categories o_1 ('low'), o_2 ('medium') and o_3 ('high'), with corresponding probabilities $\Pr(o_1)$, $\Pr(o_2)$ and $\Pr(o_3)$, which we will call the prior (i.e., pre-forecast) probabilities. A message f_i is received which serves to transform these prior probabilities into the posterior probabilities $\Pr(o_j|f_i)$, with $\sum_j \Pr(o_j|f_i) = 1$, $\Pr(o_j|f_i) \geq 0$, $j = 1,2,3$. The information content of this message as viewed from the perspective of a particular o_j is [from Equation (13)]:

$$\text{information content of message } f_i = \log \left[\frac{\Pr(o_j|f_i)}{\Pr(o_j)} \right]
 \tag{14}$$

The expected information content of the message f_i is $I(f_i)$, is the weighted average of the information contents, the weights being the posterior probabilities $\Pr(o_j|f_i)$:

$$I(f_i) = \sum_j \Pr(o_j|f_i) \log \left[\frac{\Pr(o_j|f_i)}{\Pr(o_j)} \right]
 \tag{15}$$

referred to here as the *relative entropy* (also widely known as the Kullback-Leibler divergence). The quantity $I(f_i) \geq 0$, and is equal to zero if and only if $\Pr(o_j|f_i) = \Pr(o_j)$ for all j ; thus the expected information content of a message which leaves the prior probabilities unchanged is zero, which is reasonable. For the present example, based on Table 1, the results are illustrated in Figure 3.

Note now that we can re-write Equation (5) as:

$$I_M(O, F) = \sum_i \Pr(f_i) \sum_j \Pr(o_j|f_i) \log \left[\frac{\Pr(o_j|f_i)}{\Pr(o_j)} \right]
 \tag{16}$$

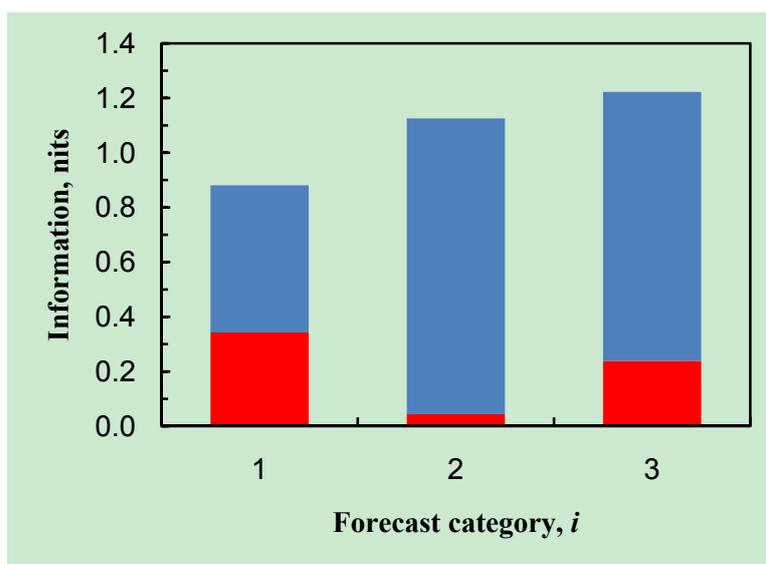
so now, recalling Equation (15), we can write:

$$I_M(O, F) = \sum_i \Pr(f_i) I(f_i)
 \tag{17}$$

and thus we can see that Equation (16) calculates the expected value of $I(f_i)$; that is to say, expected mutual information is expected relative entropy over all forecast categories.

Thus we have two information quantities that characterize a specific forecast f_i , for both of which the expected value is the expected mutual information $I_M(O,F)$. Specific information, $I_S(f_i)$, is based on the difference between two entropies [Equation (10)]. Relative entropy, $I(f_i)$, is based on the difference between two information contents [Equations (13) and (14)].

Figure 3. For each forecast category i , the bar comprises a red component $I(f_i)$, and a blue component $H(O|f_i)$. The weighted average of the sums of the two components is equal to $H(O)$. The weighted average of red components is equal to $I_M(O,F)$. In each case the $\text{Pr}(f_i)$ provide the appropriate weights.



3.5. A Second Data Set

Here we present a second prediction-realization table based on a boundary line plot for cereal cropping soils from Figure 2 of [16] (Table 2) and then summarize the calculations based on both Tables 1 and 2 (Table 3).

Table 2. Prediction-realization table based on $N=247$ observations of N_2O flux for cereal cropping soils from Figure 2 of [16]. The boundary lines between forecast emission categories were $\text{WFPS}(\%)+0.76 \cdot T(^{\circ}\text{C}) = 78$ (low-medium) and $\text{WFPS}(\%)+0.76 \cdot T(^{\circ}\text{C}) = 90$ (medium-high).

Forecast category, f_i	Observed category, o_j			Row sums
	1. Low	2. Medium	3. High	
1. Low	0.7854	0.0324	0.0000	0.8178
2. Medium	0.0972	0.0445	0.0040	0.1457
3. High	0.0121	0.0202	0.0040	0.0364
Column sums	0.8947	0.0972	0.0081	1.0000

Table 3. Summary of information quantities and calculations (working in natural logarithms).

Information quantity	Equation (boldface indicates equation used for calculation)	Value (nits) for pasture and sugarcane soils data (Table 1)	Value (nits) for cereal cropping soils data (Table 2)
$H(O)$	1	1.0687	0.3650
$H(F)$	2	1.0936	0.5659
$H(O,F)$	3	1.9585	0.8430
$I_M(O,F)$	4, 5, 6, 7, 12, 16, 17	0.2038	0.0879
$H(O F)$	Component of 6	0.8649	0.2772
$H(F O)$	Component of 7	0.8898	0.4780
normalized $I_M(O,F)$	8	0.1907	0.2407
$H(O f_1)$	9	0.5382 ^{a,b}	0.1667
$H(O f_2)$	9	1.0813 ^{a,b}	0.7321
$H(O f_3)$	9	0.9839 ^{a,b}	0.9369
$I_S(f_1)$	10, 11	0.5305 ^a	0.1984
$I_S(f_2)$	10, 11	-0.0126 ^a	-0.3671
$I_S(f_3)$	10, 11	0.0848 ^a	-0.5718
$I(f_1)$	15	0.3428 ^b	0.0325
$I(f_2)$	15	0.0442 ^b	0.1882
$I(f_3)$	15	0.2388 ^b	0.9305

^a see Figure 2; ^b see Figure 3

4. Results and Discussion

From Table 3, we see that the entropy $H(O)$ for pasture and sugarcane soils is larger than for cereal cropping soils, indicating greater dispersion in the distribution $\text{Pr}(O)$ for pasture and sugarcane soils. The expected mutual information $I_M(O,F)$ is a measure of association between forecasts and observations. This is larger for pasture and sugarcane soils than for cereal cropping soils, but is difficult to interpret because its maximum value is the entropy $H(O)$. If, instead, we look at the normalized version of $I_M(O,F)$, which lies between 0 and 1, we see that the proportion of entropy in O that is explained by the forecaster F is similar for both data sets (see Table 3).

If we look at the relative entropies $I(f_i)$ for cereal cropping soils (Table 3), the small value for $I(f_1)$ and the large value for $I(f_3)$ are notable. In each case, the largest component of $I(f_i)$ will arise from the information content of a correct forecast, $\ln[\text{Pr}(o_j|f_i)/\text{Pr}(o_j)]$ ($i=j$) [from Equation (14)]. From Table 2, for a correct f_1 ('low') forecast, information content = $\ln[0.7854/(0.8178 \cdot 0.8947)] = 0.0708$ nits. From Table 2, we calculate that an f_1 forecast provides about 96% correct forecasts, but this is set against the fact that almost 90% correct categorizations could be made just on the basis of o_1 without recourse to a forecast. Information content is a measure of the value of a forecast given what we already know. For a correct f_3 ('high') forecast, information content = $\ln[0.0040/(0.0364 \cdot 0.0081)] = 2.6190$ nits. While this is impressively large, it is based on only two o_3 observations, of which one was correctly forecast, so should be regarded with caution. For cereal cropping soils, we note also that the specific information values $I_S(f_2)$ and $I_S(f_3)$ are negative (Table 3). As almost 90% of the observations were in the o_1

category, an f_2 forecast results in $H(O|f_2) > H(O)$ and an f_3 forecast results in $H(O|f_3) > H(O)$ (Table 3); in both cases uncertainty is increased.

For pasture and sugarcane soils, the small value for the relative entropy $I(f_2)$ is notable (Table 3). From Table 1, for a correct f_2 ('medium') forecast, information content = $\ln[0.1513/(0.3653 \cdot 0.2878)] = 0.3639$ nits, not small enough to provide an explanation for the small value of $I(f_2)$ without further investigation. From Table 1, we calculate that an f_2 forecast provides about 41% correct forecasts, smaller than for both an f_1 forecast (about 82%) and an f_3 forecast (about 55%). At the same time, f_2 forecasts make up a larger percentage of the total forecasts (about 36%) than f_1 (about 35%) or f_3 (about 29%). Further, the conditional entropy $H(O|f_2)$ is larger than $H(O|f_1)$ and $H(O|f_3)$; so large, in fact, that an f_2 forecast results in $H(O|f_2) > H(O)$ and $I_S(f_2)$ is negative, indicating increased uncertainty (Table 3). Taken together, these results indicate that the small value for the relative entropy $I(f_2)$ arises because the f_2 forecast category contains relatively large proportions of incorrectly-forecast o_1 and o_3 observations in addition to the correctly-forecast o_2 observations.

On the basis of our analysis, we note that there is a *prima facie* case for considering boundary line models for N_2O flux in which the parameter space is partitioned by a single threshold into two sub-regions. In particular, the advantage of retaining separate medium and high emission categories deserves critical examination. As discussed above:

- for cereal cropping soils, information properties of the three sub-region model largely depend on the prior (*i.e.*, pre-forecast) probabilities $\Pr(o_1)$ (≈ 0.9), $\Pr(o_2)$ (≈ 0.1) and $\Pr(o_3)$ (< 0.01) of the observed N_2O flux categories o_1 ('low'), o_2 ('medium') and o_3 ('high') respectively;
- for pasture and sugarcane soils, information properties of the three sub-region model indicate that observed N_2O flux categories o_1 ('low'), o_2 ('medium') and o_3 ('high') are poorly distinguished in the f_2 forecast category.

Further:

- Conen *et al.* [15] observed that "During most days of the year, emissions tend to be within the 'low' range, increasing to 'medium' or 'high' only after fertilizer applications, depending on soil temperature or WFPS limitations."
- Recalling the data set from Figure 1, we note that as in [15], most emissions were in the 'low' observed range. The proportions of emissions in the 'low' (< 10 g N_2O-N ha $^{-1}$ day $^{-1}$), 'medium' (10-100 g N_2O-N ha $^{-1}$ day $^{-1}$) and 'high' (> 100 g N_2O-N ha $^{-1}$ day $^{-1}$) observed ranges were ≈ 0.68 , ≈ 0.30 and ≈ 0.02 , respectively.

Characterizing the extent to which observed levels of a response variable of interest are correctly or incorrectly forecast is greatly simplified if the parameter space is partitioned by a single threshold. For an example of a boundary line model in which the parameter space is partitioned into two sub-regions, an epidemiological study by De Wolf *et al.* (Figure 2 in [27]) calculates a model with a single threshold, used in forecasting wheat Fusarium head blight epidemics based on within-season weather data. The decision-theoretic and information-theoretic properties of such binary prediction models in epidemiology have been discussed by Madden [28] and Hughes [22] respectively.

5. Conclusions

The boundary line approach provides a simple and practical alternative to more complex process-based models for the estimation of N₂O emissions from soils [15,16]. The boundary line approach categorizes data for observed and forecast emissions; a graphical two-threshold boundary line model can be written as a 3×3 prediction-realization table. Boundary line model data in such a tabular format may be analyzed by information theoretic methods as also applied in, for example, psychology [21], economics [18] and epidemiology [22].

Expected mutual information is a measure of the amount of information about the observations contained in the forecasts, characterizing forecaster performance averaged over all forecast categories. Expected mutual information characterizes the component of entropy $H(O)$ that is associated with forecaster F . A normalized version of expected mutual information, with a range from zero (forecasts are independent of observations) to one (forecasts are perfect), is useful if we want to compare model performance over different data sets. Here, we found that a boundary line approach to forecasting N₂O emission ranges from Australian agricultural soils [16] provided a similar level of performance averaged over low, medium and high forecast categories for cereal cropping soils and for pasture and sugarcane soils.

Whereas mutual information characterizes the average performance of a forecaster, specific information $I_S(f_i)$ and relative entropy $I(f_i)$ both characterize aspects of the amount information contained in particular forecasts. Here is an heuristic interpretation in relation to our analysis of boundary line models. After receiving forecast f_i , we know more than we did before (assuming forecasts are not independent of observations). For N₂O flux categories o_1 ('low'), o_2 ('medium') and o_3 ('high'), we knew the prior probabilities $\Pr(o_1)$, $\Pr(o_2)$ and $\Pr(o_3)$, and now we know the posterior probabilities $\Pr(o_1|f_i)$, $\Pr(o_2|f_i)$ and $\Pr(o_3|f_i)$. Relative entropy is the expected value of the information content of f_i ; $I(f_i)$ cannot be negative. Now, recall that if all the $\Pr(o_j)$ had the same value, this would represent maximum uncertainty about N₂O flux categories O . Generally, larger $H(O)$ represents more uncertainty. So, if after receiving forecast f_i the $\Pr(o_j|f_i)$ are more similar than were the $\Pr(o_j)$, $H(O|f_i)$ will be larger than $H(O)$ and $I_S(f_i)$ will be negative. This represents an increase in uncertainty having received the forecast f_i . When $H(O|f_i)$ is smaller than $H(O)$, $I_S(f_i)$ will be positive; this represents a decrease in uncertainty having received the forecast f_i .

Finally, for completeness, we note two related areas of study not discussed in the present article, but of potential future interest in the context of applications of information theory to the analysis and modeling of N₂O emissions from soils. First, there is another application of boundary line models, not considered here, where a boundary represents the upper or lower limit of a response variable with variation in the value of an explanatory variable (see, *e.g.*, [29,30]). Second, there is a burgeoning interest in information theory as a basis for weather forecast evaluation (*e.g.*, DelSole [26,31]; Weijs *et al.* [32] and Tödter and Ahrens [33] are recent contributions). In future, such work may also contribute to the analysis of models of N₂O emissions from agricultural soils.

Acknowledgments

The research on prediction of N₂O emissions with the boundary line model was funded by the Department of Climate Change and Energy Efficiency (via ex-AGO), Australia. The grassland experimental work was funded by the UK Department for Environment, Food and Rural Affairs (Defra), the Scottish Government, the Department of Agriculture and Rural Development in Northern Ireland and the Welsh Government. SRUC receives grant-in-aid from the Scottish Government.

References

1. Mosier, A.R.; Duxbury, J.M.; Freney, J.R.; Heinemeyer, O.; Minami, K. Assessing and mitigating N₂O emissions from agricultural soils. *Climatic Change* **1998**, *40*, 7–38.
2. IPCC (Intergovernmental Panel on Climate Change). Greenhouse gas emissions from agricultural soils. In *Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 3, Greenhouse Gas Inventory Reference Manual*; Houghton, J.T., Meira Filho, L.G., Lim, B., Treanton, K., Mamaty, I., Bonduki, Y., Griggs, D.J., Callender, B.A., Eds.; IPCC/OECD/IEA: UK Meteorological Office, Bracknell, UK, 1997.
3. Bouwman, A.F. Direct emissions of nitrous oxide from agricultural soils. *Nutr. Cycl. Agroecosys.* **1996**, *46*, 53–70.
4. Bouwman, A.F.; Boumans, L.J.M.; Batjes, N.H. Emissions of N₂O and NO from fertilized fields: Summary of available measurement data. *Global Biogeochem. Cy.* **2002**, *16*, 6.1–6.13.
5. Skiba, U.; Ball, B. The effect of soil texture and soil drainage on emissions of nitric oxide and nitrous oxide. *Soil Use Manage.* **2002**, *18*, 56–60.
6. Flechard, C.R.; Ambus, P.; Skiba, U.; Rees, R.M.; Hensen, A.; van Amstel, A.; van den Pol-van Dasselaar, A.V.; Soussana, J.-F.; Jones, M.; *et al.* Effects of climate and management intensity on nitrous oxide emissions in grassland systems across Europe. *Agr. Ecosyst. Environ.* **2007**, *121*, 135–152.
7. Abdalla, M.; Jones, M.; Williams, M. Simulation of N₂O fluxes from Irish arable soils: effect of climate change and management. *Biol. Fert. Soils* **2010**, *46*, 247–260.
8. Hartmann, A.A.; Niklaus, P.A. Effects of simulated drought and nitrogen fertilizer on plant productivity and nitrous oxide (N₂O) emissions of two pastures. *Plant Soil* **2012**, *361*, 411–426.
9. Rees, R.M. Global nitrous oxide emissions: sources and opportunities for mitigation. In *Understanding Greenhouse Gas Emissions from Agricultural Management*; Guo, L., Gunasekara, A.S., McConnell, L.L., Eds.; ACS Publications: Washington DC, USA, 2012; pp. 257–273.
10. Lesschen, J.P.; Velthof, G.L.; de Vries, W.; Kros, J. Differentiation of nitrous oxide emission factors for agricultural soils. *Environ. Pollut.* **2011**, *159*, 3215–3222.
11. Pappa, V.A.; Rees, R.M.; Walker, R.L.; Baddeley, J.A.; Watson, C.A. Nitrous oxide emissions and nitrate leaching in an arable rotation resulting from the presence of an intercrop. *Agr. Ecosyst. Environ.* **2011**, *141*, 153–161.
12. Del Grosso S.J.; Parton, W.J.; Mosier, A.R.; Ojima, D.S.; Kulmala, A.E.; Phongpan, S. General model for N₂O and N₂ gas emissions from soils due to denitrification. *Global Biogeochem. Cy.* **2000**, *14*, 1045–1060.

13. Li, C.S.; Frolking, S.; Frolking, T.A. A model of nitrous oxide evolution from soil driven by rainfall events: 1. Model structure and sensitivity. *J. Geophys. Res.* **1992**, *97*, 9759–9776.
14. Li, C.S.; Frolking, S.; Frolking, T.A. A model of nitrous oxide evolution from soil driven by rainfall events: 2. Model applications. *J. Geophys. Res.* **1992**, *97*, 9777–9783.
15. Conen, F.; Dobbie, K.E.; Smith, K.A. Predicting N₂O emissions from agricultural land through related soil parameters. *Global Change Biol.* **2000**, *6*, 417–426.
16. Wang, W.; Dalal, R. Assessment of the boundary line approach for predicting N₂O emission ranges from Australian agricultural soils. In Proceedings of the 19th World Congress of Soil Science: Soil Solutions for a Changing World, Brisbane, Australia, 1–6 August 2010; Gilkes, R.J., Prakongkep, N., Eds.; IUSS: Published on DVD, <http://www.iuss.org> (accessed on 1 March 2013).
17. Tribus, M.; McIrvine, E.C. Energy and information. *Sci. Am.* **1971**, *225*, 179–188.
18. Theil, H. *Economics and Information Theory*; North-Holland: Amsterdam, The Netherlands; 1967.
19. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, USA, 1949.
20. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley-Interscience: Hoboken, NJ, USA, 2006.
21. Attneave, F. *Applications of Information Theory to Psychology: A Summary of Basic Concepts, Methods, and Results*; Holt, Rinehart and Winston: New York, NY, USA, 1959.
22. Hughes, G. *Applications of Information Theory to Epidemiology*; APS Press: St Paul, MN, USA, 2012.
23. MacDonald, D.K.C. Information theory and its application to taxonomy. *J. Appl. Phys.* **1952**, *23*, 529–531.
24. Agresti, A. *Categorical Data Analysis*, 3rd ed.; Wiley: Chichester, UK, 2012.
25. Forbes, A.D. Classification-algorithm evaluation: five performance measures based on confusion matrices. *J. Clin. Monitor.* **1995**, *11*, 189–206.
26. DelSole, T. Predictability and information theory. Part I: Measures of predictability. *J. Atmos. Sci.* **2004**, *61*, 2425–2450.
27. De Wolf, E.D.; Madden, L.V.; Lipps, P.E. Risk assessment models for wheat Fusarium head blight epidemics based on within-season weather data. *Phytopathology* **2003**, *93*, 428–435.
28. Madden, L.V. Botanical epidemiology: some key advances and its continuing role in disease management. *Eur. J. Plant Pathol.* **2006**, *115*, 3–23.
29. Schmidt, U.; Thöni, H.; Kaupenjohann, M. Using a boundary line approach to analyze N₂O flux data from agricultural soils. *Nutr. Cycl. Agroecosys.* **2000**, *57*, 119–129.
30. Farquharson, R.; Baldock, J. Concepts in modelling N₂O emissions from land use. *Plant Soil* **2008**, *309*, 147–167.
31. DelSole, T. Predictability and information theory. Part II: Imperfect forecasts. *J. Atmos. Sci.* **2005**, *62*, 3368–3381.
32. Weijs, S.; Van Nooijen, R.; Van de Giesen, N. Kullback-Leibler divergence as a forecast skill score with classic reliability-resolution-uncertainty decomposition. *Mon. Weather Rev.* **2010**, *138*, 3387–3399.

33. Tödter, J.; Ahrens, B. Generalization of the ignorance score: continuous ranked version and its decomposition. *Mon. Weather Rev.* **2012**, *140*, 2005–2017.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).