Resolution of probabilistic weather forecasts with application in disease management

Hughes, G; McRoberts, N; Burnett, FJ

Published in:
Phytopathology

DOI:
10.1094/PHYTO-07-16-0256-R

First published: 22/12/2016

Document Version
Peer reviewed version

Analytical and Theoretical Plant Pathology

Resolution of probabilistic weather forecasts with application in disease management

G. Hughes, N. McRoberts, and F. J. Burnett

First and third authors: Crop and Soil Systems Research Group, SRUC, The King's Buildings, West Mains Road, Edinburgh EH9 3JG, UK; second author: Plant Pathology Department, University of California, Davis, CA 95616-8751, USA.

Corresponding author: G. Hughes; E-mail address: gareth.hughes@sruc.ac.uk

ABSTRACT

Hughes, G., McRoberts, N., and Burnett, F. J. 20XX. Resolution of probabilistic weather forecasts with application in disease management. Phytopathology XXX:XXX-XXX.

Predictive systems in disease management often incorporate weather data among the disease risk factors, and sometimes this comes in the form of forecast weather data rather than observed weather data. In such cases, it is useful to have an evaluation of the operational weather forecast, in addition to the evaluation of the disease forecasts provided by the predictive system. Typically, weather forecasts and disease forecasts are evaluated using different methodologies. However, the information theoretic quantity *expected mutual information* provides a basis for evaluating both kinds of forecast. Expected mutual information is an appropriate metric for the average performance of a predictive system over a set of forecasts. Both relative entropy (a divergence, measuring information gain) and specific information (an entropy difference, measuring change in uncertainty) provide a basis for the assessment of individual forecasts.

Additional keywords: disease forecast, weather forecast, forecast evaluation, information theory, expected mutual information, forecast skill.

Weather factors play an important part in the reproduction and dispersal of many plant pathogens. It is therefore not surprising that meteorological data are often included in analyses of risk that aim to support decision making via predictive systems in disease management. Most often, actual weather data are deployed in this context, but sometimes forecast weather data are used (Bourke 1970; Gent et al. 2013; Olatinwo and Hoogenboom 2014). In cases where forecast data are used, it is likely that an evaluation of weather forecast accuracy will be required (e.g., Vincelli and Lorbeer 1988b). A difficulty that then arises is that methods for the evaluation of weather forecasts (often called *forecast verification* in meteorology; e.g., Casati et al. 2008) are quite different from methods usually used for the evaluation of disease forecasts.

It is not difficult to see why different methods are used for the evaluation of disease forecasts and weather forecasts. Predictive systems that support decision making in crop disease management are developed on the basis of yield and/or disease data obtained from untreated crops. This is because the ultimate objective of devising a scheme for predicting whether or not there is a need for preventative treatment cannot be achieved by analysis of data from crops where the treatment has already been applied. Then, in application, the use of a predictive system precludes knowing with certainty whether or not the eventual level of disease in a crop treated at the economic threshold would – without the treatment – actually have exceeded the economic injury level. That is to say, where the predictions may lead to interventions that change the observed outcomes, comparison of predictions and observations is not a useful basis for evaluation. Thus, evaluation of a predictive system with application in crop disease management takes place during the development of the system, when achievable and acceptable average error rates for operational crop protection decisions are set (e.g., Yuen et al. 1996).

Unlike the predictive systems developed and applied in disease management, general weather forecasts exist independent of any particular decision process(es) in which they may fulfill a role. It may be possible to act so as to mitigate the effect of the weather on the basis of such a forecast, but the objective of a predictive system that underlies weather forecasting is obviously not to enable the prevention of undesirable weather (Project Cirrus notwithstanding). Even defining whether, for example, rain or no rain is the undesirable state may prove to be beyond widespread consensus; whereas in the case of disease management, it is clear that exceeding the level of disease denominated as the economic threshold is the undesirable state, compared with not exceeding the threshold. For weather forecast evaluation, then, it is possible to record the actual weather that occurs following a given forecast. Thus, weather forecast evaluation can take place while a predictive system is operational, on the basis of data sets accumulated over a period of time comprising both the forecasts made and the corresponding observations.

So, given that different methods are usually used for the evaluation of disease forecasts and of weather forecasts, the purpose of the analysis outlined here is to provide a common currency for the evaluation of both types of forecast. Specifically we investigate the application of the information theoretic quantity *expected mutual information* in this context. The article is set out as follows. The main information theoretic concepts to be applied here in the evaluation of disease forecast data and weather forecast data are outlined (more detailed background is available in, for example, Hughes (2012) and Hughes and McRoberts

(2014)). Expected mutual information is calculated for an example disease forecast data set and for an example weather forecast data set with potential application in disease forecasting. Expected mutual information characterizes forecast properties on the basis of average performance, but we also briefly consider the evaluation of individual forecasts. To conclude, there is a general discussion.

THEORY AND ANALYSIS

Data. Example data for the analyses presented here are taken from epidemiological studies of the *Botrytis squamosa* – onion pathosystem (Vincelli and Lorbeer 1988a;b and see also Vincelli and Lorbeer 1989). Our efforts here are directed neither towards an investigation of this particular pathosystem nor an evaluation of the work described in the cited papers. Rather, the choice of example data has been made on the basis of the correspondence between the particular disease management problem investigated by Vincelli and Lorbeer (1988a;b, 1989), involving both weather forecasts and disease forecasts, and the generic problem with which we are concerned here. In addition, the exemplary clarity of data presentation in Vincelli and Lorbeer (1988a;b) means that we may pursue the generic problem using their published data as the basis of an example. Note that the approach described is not restricted to analysis of binary predictors. Neither of the example data sets used here takes the format of a 2×2 table. For the disease forecast data (Table 1), there are three categories of observation and two categories of prediction. For the weather forecast data (Table 2), there are two categories of observation and seven categories of prediction.

Disease forecasts. First we consider an information theoretic basis for evaluation of disease forecast data. Here, we use the 4yr data set from Table 4 of Vincelli and Lorbeer (1988a) relating to the evaluation of an inoculum production index for forecasting sporulation by *B. squamosa*. The data are normalized and presented here in a prediction-realization table (Table 1) (calculations here and throughout are shown correct to 3 d.p.). We adopt the notation of

Topp et al. (2013) for Table 1 and subsequent analyses of these data. The observed categories are denoted o_j (j=1,2,3) for 'low', 'medium' and 'high' spore episodes, respectively. The bottom row of the table contains the distribution of observations Pr(O). The forecast categories are denoted f_i (i=1,2) for 'low' and 'high' sporulation forecasts, respectively. The right-hand margin of the table contains the distribution of forecasts Pr(F). The body of the table contains the joint probabilities $\Pr(o_j \cap f_i)$.

Consider the observed spore episode categories o_1 ('low'), o_2 ('medium') and o_3 ('high'), with corresponding probabilities $\Pr(o_1)$, $\Pr(o_2)$ and $\Pr(o_3)$, $\sum_j \Pr(o_j) = 1,\ 0 \le \Pr(o_j) \le 1,\ j = 1,2,3$ (Table 1). We calculate expected information content or *entropy*, denoted $H(O)$, as:

$$H(O) = -\sum_j \Pr(o_j) \cdot \ln(\Pr(o_j)) \tag{1}$$

Note that $H(O) \ge 0$ and that we take $\Pr(o_j)\cdot\ln(\Pr(o_j)) = 0$ if $\Pr(o_j) = 0$, since $\lim_{x \to 0} x \cdot \ln(x) = 0$.

Natural logarithms are used throughout, so information quantities are calculated in units of *nits* (MacDonald 1952). If any $\Pr(o_j) = 1$, $H(O) = 0$. $H(O)$ has its maximum value when all the $\Pr(o_j)$ have the same value. We can think of entropy as characterizing the extent of our uncertainty prior to receipt of a (notional) perfect forecast or, alternatively, how much information such a forecast would deliver. We can also calculate the conditional entropy (conditional, that is, on the sporulation forecast), denoted $H(O|F)$, as:

$$H(O|F) = -\sum_i \Pr(f_i) \cdot \sum_j \Pr(o_j|f_i) \cdot \ln(\Pr(o_j|f_i)) \tag{2}$$

where $H(O|F) \ge 0$ and (referring to Table 1) $\Pr(o_j|f_i) = \Pr(o_j \cap f_i) / \sum_j \Pr(o_j \cap f_i)$. The sporulation forecasts, as characterized in Table 1, are imperfect.

Then expected mutual information between the observation and forecast distributions, denoted $I_M(O,F)$, is calculated as $H(O) - H(O|F)$:

$$I_M(O,F) = \sum_j \Pr(o_j) \cdot \ln\left[\frac{1}{\Pr(o_j)}\right] - \sum_i \Pr(f_i) \cdot \sum_j \Pr(o_j|f_i) \cdot \ln\left[\frac{1}{\Pr(o_j|f_i)}\right] \tag{3}$$

with $I_M(O,F) \geq 0$, and equality only if O and F are independent. We may interpret expected mutual information as the average reduction in uncertainty about the observations O resulting from use of a predictive model to provide forecasts F. Suppose that we have a perfect forecaster such that F and O are identical, then use of the forecaster would account for all the uncertainty in O and $H(O|F) = H(O|O)$, $I_M(O,F) = H(O) - H(O|O) = H(O)$. This tells us that the maximum expected mutual information $I_M(O,F)$ between O and F, that would characterize a perfect forecaster, is the entropy $H(O)$. For a predictor based on a 2×2 prediction-realization table, the relationship between $I_M(O,F)$ and its components $H(O)$ and $H(O|F)$ can be depicted in a simple information graph (see Fig. 1 in Hughes et al. 2015). While this does not apply to the predictor in question here (see Table 1), the graph still provides an heuristic conceptualization of $H(O)$, $H(O|F)$ and $I_M(O,F)$.

Now, referring to the data in Table 1, from equation 1, $H(O) = 0.973$ nits; from equation 2, $H(O|F) = 0.887$ nits; and from equation 3, $I_M(O,F) = 0.086$ nits.

Weather forecasts. Now we turn to an information theoretic basis for evaluation of weather forecast data. Here, we use the combined data set from Table 2 of Vincelli and Lorbeer (1988b). These data, presented in Table 2, combine rainfall observations from three forecast intervals; 0-12, 12-24 and 24-36 hours after forecast. We base our notation for Table 2 and subsequent analyses of these data on that of Hughes and Topp (2015). The table illustrates an evaluation data set for a probabilistic rainfall forecaster. Here the data are presented in 7 forecast categories. Compared with the presentation of Vincelli and Lorbeer (1988b), we have combined the categories for rainfall probability forecasts 0.8, 0.9 and 1.0 into a single category centered on 0.9. This matches the combined category for 0-0.2 rainfall probability forecasts in the original presentation, and avoids the use of a probability forecast of 1.0. For further discussion relating to probability forecasts of zero and one, see Roulston and Smith (2002) and Hughes and Topp (2015). The forecast category index is denoted k, the rainfall probability forecast in category k is denoted p_k, (the no-rainfall probability forecast is the complement), the number of rainfall observations in category k is denoted o_k, and the number of observations in category k is denoted n_k. Then we have $\bar{o}|k = o_k/n_k$ for the average frequency of rainfall observations in forecast category k, $\sum_k o_k$ for the total number of rainfall observations, $\sum_k n_k$ (denoted N) for the total number of observations, and $\bar{o} = \sum_k o_k/N$ for the overall average frequency of rainfall observations. In meteorological forecast evaluation, the average frequency of a weather event $(\bar{o})$ is referred to as the *climatological probability* (Joliffe and Stephenson 2012) (often abridged to *climatology*). The individual rainfall/no rainfall forecast data can be reconstructed from the data in Table 2. For example, in forecast category $k = 4$, $n_k = 31$ individual forecasts ("rainfall probability (p_k) = 0.5") were made. Of these 31 forecasts, $o_k = 14$ were followed by rainfall, and 31−14 = 17 were followed by no rainfall. Note in passing that comparison of o_k/n_k to p_k (see Table 2) is the basis of an evaluation of the reliability of a forecast, as mentioned briefly below.

Probability forecasts – as widely used in meteorology – provide a forecast probability p that a weather event will subsequently occur. Vincelli and Lorbeer (1988b) describe a forecaster with two outcome categories: rainfall (as defined in the study) either occurs within the forecast interval or it does not. For a useful forecaster, (qualitatively) we expect more observed rainfall events when the forecast probability for rainfall is closer to 1 and fewer when the forecast probability is closer to 0. Quantitative methods for the evaluation of forecasters based on comparison of forecast probabilities and corresponding observed frequencies are called *scoring rules*. It is convenient here to think of a scoring rule as a means of attaching a penalty score to a forecast; the better the forecast, the smaller the penalty. In practice, we are usually interested in the evaluation of a forecaster based on the average penalty score for a data set comprising a sequence of forecasts and the corresponding observations; a better forecaster achieves a smaller average score. Vincelli and Lorbeer's (1988b) evaluation of their forecaster is based on the Brier Score (*BS*) (Brier 1950). Here, we use the Divergence Score (*DS*) (Weijs et al. 2010), because it allows an information theoretic interpretation of the forecast evaluation. The two scores are closely related – both are examples of Bregman divergences (see, e.g., Hughes and Topp 2015).

The divergence score is based on the Kullback-Leibler divergence (D_{KL}) (Kullback and Leibler 1951). For an individual forecast made in one of two possible outcome categories (for forecast interval t), the divergence score is the Kullback-Leibler divergence between the (Bernoulli) distributions of observations (o_t, $1-o_t$) and of forecasts (p_t, $1-p_t$):

$$D_{KL}\left(o_t \| p_t\right) = o_t \cdot \ln\left(\frac{o_t}{p_t}\right) + \left(1-o_t\right) \cdot \ln\left(\frac{1-o_t}{1-p_t}\right) \tag{4}$$

and then the overall divergence score is the average over N individual forecasts:

$$DS = \frac{1}{N} \cdot \sum_{t=1}^{N} D_{KL}\left(o_t \| p_t\right) \tag{5}$$

Note that $D_{KL}\left(\bullet \| \bullet\right) \geq 0$ and that the divergence is not necessarily symmetric with respect to the arguments. As above, we take 0·ln(0) = 0. Then, from equation 4, $D_{KL}\left(o_t \| p_t\right) = 196.666$ nits; and from equation 5 (with $N = 527$), $DS = 0.373$ nits.

The overall divergence score *DS* has a decomposition into three components: *uncertainty* (*UNC*), *reliability* (*REL*), and *resolution* (*RES*) such that:

$$DS = UNC + REL - RES \tag{6}$$

(Weijs et al. 2010). *UNC* is the entropy $H\left(\bar{o}\right)$:

$$UNC = -\left[\bar{o} \cdot \ln(\bar{o}) + (1-\bar{o}) \cdot \ln(1-\bar{o})\right] \tag{7}$$

The reliability component (*REL*) is the weighted average divergence between the distribution of rainfall observations $(\bar{o}|k)$ and that of the corresponding probability forecasts (p_k) for the specified forecast categories:

$$REL = \frac{1}{N} \cdot \sum_k n_k \cdot D_{KL}\left(\bar{o}|k \| p_k\right) \tag{8}$$

with:

$$D_{KL}\left(o|k \| p_k\right) = \bar{o}|k \cdot \ln\left(\frac{\bar{o}|k}{p_k}\right) + \left(1-\bar{o}|k\right) \cdot \ln\left(\frac{1-\bar{o}|k}{1-p_k}\right) \tag{9}$$

If the average frequencies of rainfall observations for forecast categories are close to the corresponding forecast probabilities, *REL* is close to zero (such probability forecasts are referred to as *well-calibrated* by meteorologists). Note in passing that the evaluation of disease forecasts as described above does not involve calculation of a reliability component. This is because the probability forecasts in that case are the Bayesian posterior probabilities calculated on the basis of the prediction-realization table, which has the effect of ensuring perfect calibration. The resolution component (*RES*) is the weighted average divergence between the distribution of rainfall observations $(\bar{o}|k)$ and that of the forecasts based only on the climatological probability $\bar{o}$. Then:

$$RES = \frac{1}{N} \cdot \sum_k n_k \cdot D_{KL}\left(\bar{o}|k \| \bar{o}\right) \tag{10}$$

with:

$$D_{KL}\left(o|k \| \bar{o}\right) = \bar{o}|k \cdot \ln\left(\frac{\bar{o}|k}{\bar{o}}\right) + \left(1-\bar{o}|k\right) \cdot \ln\left(\frac{1-\bar{o}|k}{1-\bar{o}}\right) \tag{11}$$

and *RES* is thus a measure of the degree to which the probability forecast categories sort the observations into distinct groups (Wilks 2011). As defined by equation 10, *RES* is the expected mutual information between observation and forecast distributions (Weijs et al. 2010). We can therefore use equation 10 to calculate the equivalent quantity in the context of weather forecast evaluation to that calculated by equation 3 in the context of disease forecast evaluation.

Referring to the data in Table 2 we calculate $N = 527$, $\bar{o} = 0.190$. Then from equation 7, $UNC = 0.486$ nits; from equation 8, $REL = 0.045$ nits; and from equation 10, $RES = 0.158$ nits. We note that $DS = 0.486 + 0.045 - 0.158 = 0.373$ (in nits), in accordance with both equation 6 and the calculation based on equations 4 and 5.

Forecast skill. In meteorology, a *skill score* uses an appropriate set of reference forecasts to calculate a normalized version of a score (e.g., a Brier score or a divergence score). Skill scores are often interpreted as percentage improvement over the reference forecasts achieved by use of the particular probability forecasts in question (Wilks 2011). In the case of the divergence score, we have the divergence skill score (*DSS*):

$$DSS = 1 - \frac{UNC + REL - RES}{UNC} = \frac{RES - REL}{UNC} \quad (12)$$

(Weijs et al. 2010). Thus equation 12 calculates a skill score ranged between 0 and 1. A skill score of 0 (or 0%) arises if forecasts are made only on the basis of the climatological probability $\bar{o}$ (in which case $REL = 0$ and $RES = 0$); while a skill score of 1 (or 100%) is characteristic of a perfect forecaster (in which case $REL = 0$ and $RES = UNC$) (Weijs et al. 2010). Using the calculated values of the components of *DS* based on the data in Table 2 (see above), $DSS = 0.232$ (23.2%) for the example (combined) data set (0-36 hours after forecast). Using external reference data, Vincelli and Lorbeer (1988b) calculated Brier skill scores of

30.9%, 27.8% and 27.6% respectively for the rainfall observations from the three separate forecast intervals (0-12, 12-24 and 24-36 hours after forecast).

Weijs et al. (2010) point out that for the special case of binary forecasts, the *ranked mutual information skill score* (*RMIS*) of Ahrens and Walser (2008) can be written as the expected mutual information between forecasts and observations (*RES*) divided by the entropy of the observations (*UNC*). Compared with *DSS* (equation 12), the calculation of *RMIS* does not take *REL* into account (in effect assuming perfect reliability, *REL* = 0).

If we now return to disease forecasts and consider the normalized version of the expected mutual information between the observation and forecast distributions, we have:

$$\textit{normalized}\ I_M(O,F)=\frac{H(O)-H(O|F)}{H(O)}=\frac{I_M(O,F)}{H(O)} \tag{13}$$

This ratio is characteristic of a normed (uncorrected) association measure as discussed by Särndal (1974). The interpretation of any association measure of this type is that of relative reduction in uncertainty about the observations as a result of receiving the forecasts. It takes a value of 0 (or 0%) when the forecasts and observations are independent and 1 (or 100%) when the forecasts perfectly predict the observations. Using the calculated values based on the data in Table 1, $\textit{normalized}\ I_M(O,F)=0.088$ (8.8%) for the example data set (not dissimilar to values for disease forecasts calculated in Hughes (2012)). In passing, note that calculation of this association measure, and of the divergence skill score above, do not depend on the choice of logarithmic base (unlike the information quantities on which they are based).

Individual forecasts. At this stage, we have illustrated the use of expected mutual information in evaluating the average performance of both disease forecasts and weather forecasts. Note that in the case of disease forecasts, expected mutual information was calculated as the difference between two entropies (equation 3). However, in the case of weather forecasts, expected mutual information was calculated as the (average) divergence between two distributions (equation 10). Although the two calculations both lead to expected mutual information (e.g., DelSole and Tippett 2007), they do so in ways that take different perspectives on the perception of individual forecasts.

For the sake of brevity, we will consider just the weather forecast data set here, but the same principles are applicable to the analysis of disease forecasts. The outcome of the calculation of the difference between the entropy for a forecast based on the climatological probability $\overline{o}$ and the entropy for a probability forecast in category k is $H(\overline{o})-H(\overline{o}|k)$ (where $H(\overline{o})$ is the entropy as in equation 7 (*UNC*) and the $H(\overline{o}|k)$ are the conditional entropies calculated as in Table 2; both in nits). This quantity is referred to as *specific information*. The outcome of the calculation of the divergence between the observation distribution following a probability forecast in category k and the observation distribution for a forecast based on the climatological probability $\overline{o}$ is $D_{KL}(\overline{o}|k\|\overline{o})$ in nits (equation 11). This quantity is referred to as *relative entropy*. Expected specific information and expected relative entropy are both expected mutual information. Here, we take a diagrammatic approach to comparing specific information and relative entropy.

Figure 1 shows the binary (Shannon) entropy curve, with a horizontal dashed line intersecting the curve at $\overline{o}=0.190$ (the climatological probability), the corresponding entropy being $H(\overline{o})=$ 0.486 nits (equation 7). Then, at each $\overline{o}|k$, the vertical distance between this horizontal line and $H(\overline{o}|k)$ (on the entropy curve) is the corresponding specific information $H(\overline{o})-H(\overline{o}|k)$ in nits. Specific information characterizes the change in uncertainty that results from the observation $\overline{o}|k$ as compared with the forecast based on the climatological probability $\overline{o}$. Specific information may be positive $\left(H(\overline{o}) > H(\overline{o}|k)\right)$, in which case uncertainty has decreased, or negative $\left(H(\overline{o}) < H(\overline{o}|k)\right)$, in which case uncertainty has increased. Thus, in Figure 1, uncertainty increases when $0.19 < \overline{o}|k < 0.81$, specifically here for forecast index categories k = 3, 4, 5, 6 and 7 (see also Table 1). Note that although some particular $\overline{o}|k$ may result in increased uncertainty, uncertainty taken on average over all $\overline{o}|k$ cannot increase, because expected specific information is expected mutual information $H(\overline{o}) - (n_k/N) \cdot H(\overline{o}|k) \geq 0$.

Figure 2 shows the binary (Shannon) entropy curve, with a tangent to the curve drawn at $\overline{o} = 0.190$ (the climatological probability). Then, at each $\overline{o}|k$, the vertical distance between this tangent and the entropy curve is the corresponding relative entropy $D_{KL}\left(\overline{o}|k \| \overline{o}\right)$ in nits. This diagrammatic interpretation is possible because relative entropy is an example of a Bregman divergence (Gneiting and Raftery 2007). Although Bregman divergences are usually defined as properties of convex functions, it is easier here for the purpose of comparison with Figure 1 if we think of the Bregman divergence for a concave function (Zhang 2004) in order to keep the concave entropy curve (Shannon 1948, see Fig. 7) as the basis of the calculation. Calculated values of $D_{KL}\left(\overline{o}|k \| \overline{o}\right)$ in nits are given in Table 2. Expected relative entropy is expected mutual information (*RES*) (equation 10).

DISCUSSION

The assessment of forecast quality is a current issue in relation to decision making in agriculture (Kusunose and Mahmood 2016). We have seen that expected mutual information – either as an information quantity or in its normalized form – has application in the evaluation of the average performance of both disease forecasts and weather forecasts. Relative entropy and specific information both provide a basis for evaluating individual forecasts. The calculations for expected mutual information, relative entropy and specific information have no additional data requirements over and above a typical data set collected for forecast evaluation. To demonstrate this, our illustration of the methodology uses previously published data from disease forecast and weather forecast evaluation studies of a pathosystem (Vincelli and Lorbeer 1988a;b).

While expected mutual information provides a principled basis for forecast evaluation on the basis of average performance, both relative entropy (a divergence, measuring information gain) and specific information (an entropy difference, measuring change in uncertainty) provide a basis for the evaluation of individual forecasts. The expected value of both relative entropy and of specific information is equal to expected mutual information, but in the context of evaluation of individual forecasts they have different analytical properties, as discussed by DelSole (2004) and by DelSole and Tippett (2007). In the past, this has caused some confusion, notably in the clinical literature, as discussed by Hughes and McRoberts (2015). The scenario-based examples reviewed in Hughes and McRoberts (2015) lack the generality of analytical comparisons, but serve to illustrate the source of the confusion – that an information gain does not necessarily result in a reduction in uncertainty.

Relative entropy arises naturally in the context of the decomposition of the divergence score (Weijs et al. 2010; Hughes and Topp 2015) and is ≥ 0, and in practice >0 if we assume that for no operational forecaster will forecasts and corresponding observations be independent. Thus in practice the result of using such a forecaster is always a gain in information when comparing the post-forecast probability of an event with the pre-forecast probability (as illustrated by the analysis depicted in Figure 2). In the context of crop protection decision making, however, that same comparison of the post-forecast probability of an event with the pre-forecast probability may result in negative specific information, in which case uncertainty has increased following the forecast (as illustrated by some parts of the analysis depicted in Figure 1). This is, it should be emphasized, a generic problem of forecasting that is laid bare by the application of information theory; a problem that may have implications for the uptake and continued use of predictive systems in disease management if, for example, developers of forecasters generally prioritize information gain when evaluating individual forecasts, while operational users of those forecasts generally prioritize the need for uncertainty reduction. We should not expect to resolve this issue in favor of one or the other metric. As DelSole (2004) writes, "*no universal principle exists to settle this question*." Rather, evaluation of a forecaster requires careful consideration of both aspects of performance, and the perspectives of both users and developers.

Neither relative entropy nor specific information formed part of the original exposition of information theory as presented by Shannon (1948), who was interested in the average performance of information channels. Shannon did discuss expected mutual information (as *information transfer rate*). This metric has immediate application both in the evaluation of disease forecasts and of weather forecasts and provides a common currency for use in evaluation where both kinds of forecast are combined in crop protection.

ACKNOWLEDGMENT

SRUC receives grant-in-aid from the Scottish Government.

LITERATURE CITED

Ahrens, B., and Walser, A. 2008. Information-based skill scores for probabilistic forecasts. Mon. Weather Rev. 136:352-363.

Bourke, P. M. A. 1970. Use of weather information in the prediction of plant disease epiphytotics. Annu. Rev. Phytopathol. 8:345-370.

Brier, G. W. 1950. Verification of forecasts expressed in terms of probability. Mon. Weather Rev. 78:1-3.

Casati, B., Wilson, L. J., Stephenson, D. B., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E. E., Browne B. G., and Mason S. 2008. Forecast verification: current status and future directions. Meteorol. Appl. 15:3-18.

DelSole, T. 2004. Predictability and information theory. Part I: Measures of predictability. J. Atmos. Sci. 61:2425-2440.

DelSole, T., and Tippett, M. K. 2007. Predictability: recent insights from information theory. Rev. Geophys. 45:RG4002, doi:10.1029/2006RG000202.

Gent, D. H., Mahaffee, W. F., McRoberts, N., and Pfender, W. F. 2013. The use and role of predictive systems in disease management. Annu. Rev. Phytopathol. 51:267-289.

Gneiting, T., and Raftery, A. E. 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 102:359-378.

Hughes, G. 2012. Applications of Information Theory to Epidemiology. The American Phytopathological Society, St. Paul, MN.

Hughes, G., and McRoberts, N. 2014. The structure of diagnostic information. Australas. Plant Pathol. 43:267-286.

Hughes, G., McRoberts, N., and Burnett, F. J. 2015. Information graphs for binary predictors. Phytopathology 105:9-17.

Hughes, G., and Topp, C. F. E. 2015. Probabilistic forecasts: scoring rules and their decomposition and diagrammatic representation via Bregman divergences. Entropy 17:5450-5471.

Joliffe, I. T., and Stephenson, D. B. (eds.) 2012. Forecast Verification: a Practitioner's Guide in Atmospheric Science (2nd ed.). John Wiley & Sons, Ltd., Chichester, UK.

Kullback, S., and Leibler, R. A. 1951. On information and sufficiency. Ann. Math. Statist. 22:79-86.

Kusunose, Y., and Mahmood, R. 2016. Imperfect forecasts and decision making in agriculture. Agr. Syst. 146:103-110.

MacDonald, D. K. C. 1952. Information theory and its application to taxonomy. J. Appl. Phys. 23:529-531.

Madden, L. V. 2006. Botanical epidemiology: some key advances and its continuing role in disease management. Eur. J. Plant Pathol. 115:3-23.

Olatinwo, R., and Hoogenboom, G. 2014. Weather-based pest forecasting for efficient crop protection. Pages 59-78 in: Integrated Pest Management: Current Concepts and Ecological Perspective. D. P. Abrol, ed. Academic Press, San Diego, CA.

Roulston, M. S., and Smith, L. A. 2002. Evaluating probabilistic forecasts using information theory. Mon. Weather Rev. 130:1653-1660.

Särndal, C. E. 1974. A comparative study of association measures. Psychometrika 39:165-187.

Shannon, C. E. 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27:379-423 & 632-656.

Topp, C. F. E., Wang, W., Cloy, J. M., Rees, R. M., and Hughes, G. 2013. Information properties of boundary line models for N_2O emissions from agricultural soils. Entropy 15:972-987.

Vincelli, P. C., and Lorbeer, J. W. 1988a. Forecasting spore episodes of *Botrytis squamosa* in commercial onion fields in New York. Phytopathology 78:966-970.

Vincelli, P. C., and Lorbeer, J. W. 1988b. Relationship of precipitation probability to infection potential of *Botrytis squamosa* on onion. Phytopathology 78:1078-1982.

Vincelli, P. C., and Lorbeer, J. W. 1989. BLIGHT-ALERT: a weather-based predictive system for timing fungicide applications on onion before infection periods of *Botrytis squamosa*. Phytopathology 79:493-498.

Weijs, S. V., van Nooijen, R., and van de Giesen, N. 2010. Kullback-Leibler divergence as a forecast skill score with classic reliability-resolution-uncertainty decomposition. Mon. Weather Rev. 138:3387-3399.

Wilks, D. S. 2011. Statistical Methods in the Atmospheric Sciences, 3rd ed. Academic Press, Oxford, UK.

Yuen, J., Twengström, E., and Sigvald, R. 1996. Calibration and verification of risk algorithms using logistic regression. Eur. J. Plant Pathol. 102:847-854.

Zhang, T. 2004. Statistical behavior and consistency of classification methods based on convex risk minimization. Ann. Statist. 32:56-134.

Table 1. Prediction-realization table based on 204 forecasts and corresponding observations of *B. squamosa* spore episodes (Vincelli and Lorbeer 1988a). The data are normalized and presented here in the notation of Topp et al. (2013).

	Observed category, o_j			
Forecast category, f_i	Low (j=1)	Medium (j=2)	High (j=3)	Row sums
Low (i=1)	0.324	0.049	0.025	0.397
High (i=2)	0.255	0.157	0.191	0.603
Column sums	0.578	0.206	0.216	1.000

Table 2. Probabilistic weather forecast evaluation data set based on 527 forecasts of rainfall and corresponding observations[a] (Vincelli and Lorbeer 1988b).

k [b]	1	2	3	4	5	6	7
p_k	0.1	0.3	0.4	0.5	0.6	0.7	0.9
o_k	7	15	11	14	17	15	21
n_k	271	94	50	31	30	22	29
$\overline{o}\vert k$	0.026	0.160	0.220	0.452	0.567	0.682	0.724
$H(\overline{o}\vert k)$	0.120	0.439	0.527	0.688	0.684	0.625	0.589
$D_{KL}(\overline{o}\vert k\Vert\overline{o})$	0.128	0.003	0.003	0.178	0.349	0.575	0.673

[a] rainfall ≥ 1.3mm recorded in the forecast interval (Vincelli and Lorbeer 1988b).

[b] notation: k, forecast category index; p_k, rainfall probability forecast in category k (no-rainfall probability forecast is the complement); o_k, number of rainfall observations in category k; n_k number of observations in category k; $\overline{o}|k$ is the average frequency of rainfall observations in category k; $H(\overline{o}|k)$ is the entropy $-\left[\overline{o}|k \cdot \ln(\overline{o}|k) + (1-\overline{o}|k) \cdot \ln(1-\overline{o}|k)\right]$ (in nits) and $D_{KL}(\overline{o}|k\|\overline{o})$ is the divergence as specified by equation 11 (in nits).

Fig. 1. Specific information. The solid line represents the Shannon entropy curve. The point on the curve at $\bar{o}$ = 0.190 (the climatological probability) is marked ■; the corresponding entropy is $H(\bar{o})$ = 0.486 nits and a horizontal straight line (long-dashed) is drawn through this point. Each observed value of $\bar{o}|k$ is marked ● on both the horizontal line representing the entropy $H(\bar{o})$ (where the points are labelled with the corresponding forecast category index *k*) and the curve representing the conditional entropy $H(\bar{o}|k)$ (see Table 2). The entropy differences $H(\bar{o}) - H(\bar{o}|k)$ (in nits) are indicated by vertical (short-dashed) lines. Uncertainty has decreased following the forecast if $H(\bar{o}|k) < H(\bar{o})$; uncertainty has increased following the forecast if $H(\bar{o}|k) > H(\bar{o})$.

Fig. 2. Relative entropy. The solid line represents the Shannon entropy curve. The point on the curve at $\bar{o}$ = 0.190 (the climatological probability) is marked ■; the corresponding entropy is $H(\bar{o})$ = 0.486 nits and the tangent to the curve (long-dashed) is drawn at this point. Each observed value of $\bar{o}|k$ is marked ● on both the tangent (where the points are labelled with the corresponding forecast category index *k*) and the curve; and the divergences $D_{KL}(\bar{o}|k \| \bar{o})$ are indicated by vertical (short-dashed) lines representing information gain in nits (see Table 2). In general $D_{KL}(\bar{o}|k \| \bar{o}) \geq 0$, and here all the calculated divergences are >0.

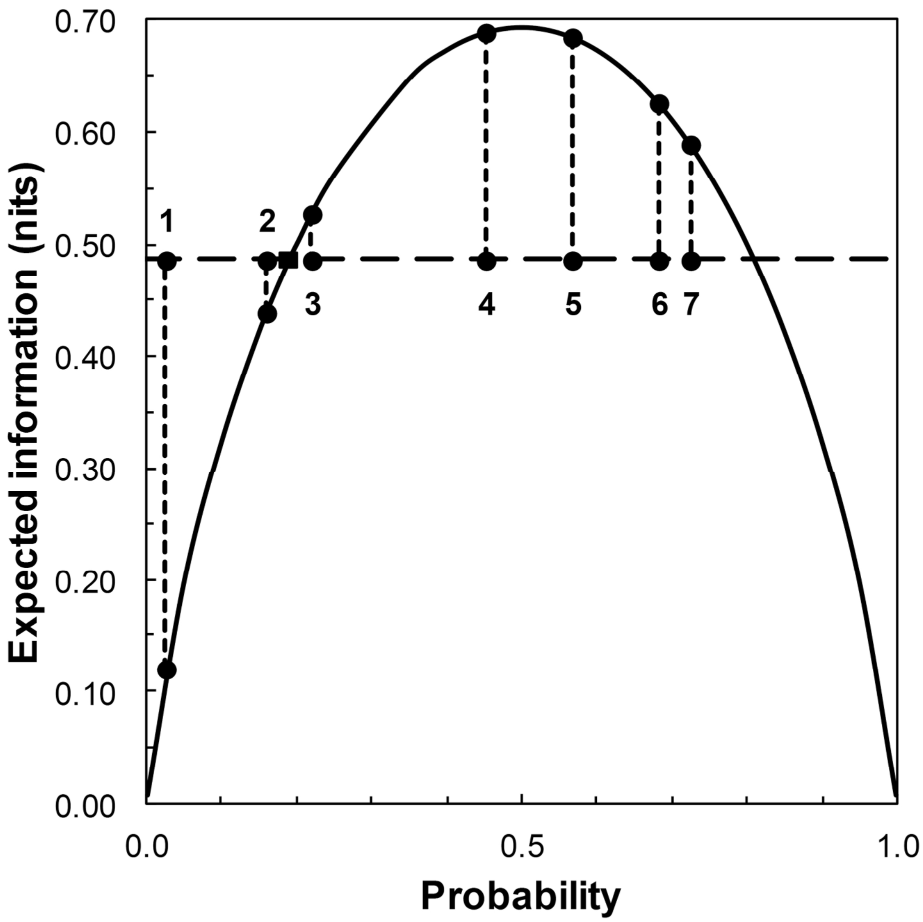

Figure 1
104x103mm (600 x 600 DPI)

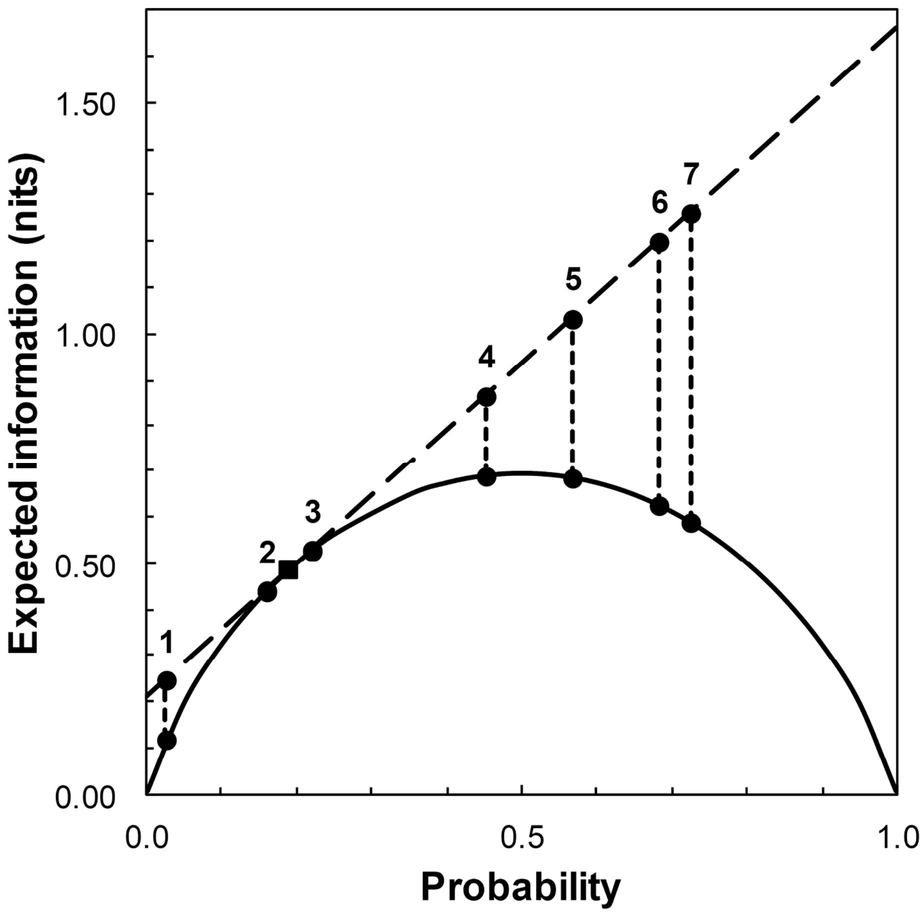

Figure 2
104x103mm (600 x 600 DPI)