A decision theoretic approach for issuing natural hazard warnings

Theodoros Economou, David B. Stephenson, Jonathan C. Rougier, Robert A. Neal and Kenneth Mylne

Abstract

Warnings for natural hazards help improve societal resilience and are a good example of decision-making under uncertainty. The usefulness of warning systems heavily depends on how well it is defined and thus understood by the stakeholders. However, most operational warning systems are heuristic in that they are not formally or transparently defined. Bayesian decision theory provides a formal framework for issuing warnings under uncertainty and despite having been proposed more than 20 years ago for flood warnings, it has not been fully exploited.

In this study, a simple framework is proposed for hazard warning systems based on Bayesian decision theory. The framework uses probabilistic predictions of the state of nature in conjunction with a loss function to issue optimal warnings with respect to expected loss. The approach is illustrated using 24 hour ahead 4-state warnings of daily severe precipitation totals over the UK and is contrasted to the current heuristic severe weather warning system used operationally by the UK Met Office, for warnings in the 2012/13 and 2013/14 winters. A probability model is proposed to predict precipitation, given forecast information, and loss functions are constructed for two generic stakeholders: a householder and a forecaster.

Weather warnings based on the proposed approach were found to be different than those issued by the Met Office. The results indicated that the existence of a loss function in vital to understanding and assessing the warning rule.

Keywords: Natural hazards; Warnings; Decision theory; ensemble forecasting.

1 Introduction

Early warning systems (EWSs) play a major role in reducing monetary, structural and human loss from natural hazards. The challenge of optimally issuing warnings is complicated—it is a 'wicked' problem (Rittel and Webber, 1973) since the stakes are different for the entity responsible for issuing the warnings and the user receiving them. It is necessary to have shared ownership of the problem, which should be promoted and facilitated by the transparency of the EWS used by the issuer. There is a need for a transparent framework for issuing natural hazard warnings, which encourages the engagement of all the involved stakeholders, especially the user.

An EWS is defined here as a tool that uses a) predictive information of the hazard, b) consequence (loss) information for each warning-outcome combination, and produces a warning according to some well-defined optimality criterion. In other words it is a rule that *transparently* maps predictive and loss information into action. An EWS that is not transparently derived from well-defined inputs is defined here as "heuristic".

Current operational EWSs, such as the UK National Severe Met Office weather Warning Service (NSWWS) (Neal et al., 2014; Met Office, 2015) and the flood warning system of the UK Environment Agency (Environment Agency, 2015), can be characterised as heuristic. The response to (and thus the overall effectiveness of) a warning system depends heavily on users believing that the warning is credible and accurate (Sorensen, 2000). This belief is of course influenced by how well the system is formulated and understood. Agents that issue warnings suffer from the "cry-wolf" syndrome, i.e. fear of loss of belief in the warning system due to false alarms, however it has been argued that this is not true if the basis of the false alarm is well understood (Dow and Cutter, 1998). In other words, there are strong arguments for why an EWS should be as simple, clear and transparent as possible. Such a system will also be amenable to criticism and thus improvement.

This article proposes a simple framework for issuing hazard warnings based on Bayesian decision theory (Lindley, 1985), which offers a coherent and transparent way of optimally issuing warnings. Bayesian decision theory uses probability to quantify uncertainty about the future hazard event and offers a strategy for issuing warnings in a rational way. The framework is illustrated by application to data from the UK Met Office first-guess warnings system (in support of the NSWWS) where predictive information is in the form of ensemble forecasts (multiple predictions of potential future weather from numerical weather models).

Section 2 defines the problem and presents a brief review of recent literature on natural hazard warnings. The Bayesian decision theoretic approach is described in section 3 and then applied to data from and contrasted to the current Met Office first-guess warning system for severe precipitation in Section 4. Section 5 concludes with a brief summary and a discussion.

2 Background

Issuing warnings for events such as severe weather or volcanic eruptions is a prime example of having to make real-time decisions under uncertainty. The uncertainty primarily comes from the fact that the occurrence and intensity of the hazard are unknown and need to be predicted using complex yet imperfect models (e.g. the one proposed here in section 4.3). EWSs therefore rely on predictive information such as numerical model forecasts of severe weather, and observed precursors like earthquake magnitude for predicting tsunamis. We define such observable predictive information as y. We also define the the set of values that the state of nature can take as the state space $\{x\}$ and the set of all possible actions as the action space $\{a\}$. The uncertainty in the prediction of a future x is quantified by the conditional probability $p(x|y)$. Losses from actions are quantified using a loss function $L(a, x)$ which is defined as a map from (a, x) to a real valued number l, i.e. a is taken and when x occurs a loss l is occurred.

In a prediction problem, where the goal is simply to predict the future value of x, the action space and the state space are the same. Relatively simple loss functions $L(a, x)$ can be used for prediction, for instance a 0/1 loss where $l = 0$ only if the prediction comes true. In that case, it can be shown (using the Bayes rule defined in section 3) that the optimal action is to predict x with the highest $p(x|y)$. In a warning problem, where the action is the issuing of a warning, the loss function cannot be so trivial and will likely be different for different stakeholders, for instance the issuer and the householder. Importantly however, the action set in the warning problem can be a lot more useful to stakeholders, since in practice, the action space will be considerably smaller than the state space—for instance a finite set of warning levels compared to an infinite set of severe wind speed values.

Much of the scientific literature in natural hazards addresses the prediction problem, with a plethora of rigorous techniques and models, while the warning problem has received little attention and even less so with respect to decision theory. Sorensen (2000) and Bhattacharya et al. (2012) make this point clear in recent reviews of natural hazard and geohazard EWS, and indicate the need for systems that integrate hazard evaluation and warning dissemination. In the rest of this section we present a review of some operational EWSs for natural hazards that address the warning problem, along with articles that have used decision theoretic approaches for both warning and prediction.

2.1 Review of decision theoretic approaches to natural hazard warning and prediction

Operational natural hazard EWSs at various spatio-temporal scales and administrative regions exist across the globe for severe weather (e.g. UK Met Office severe weather warning system, Neal et al. (2014)), water-related hazards (Alfieri et al., 2012), hurricanes (NOAA, 2015), pacific tsunamis (PTWC, 2015), volcanoes (USGS,

2015) and geohazards (Bhattacharya et al., 2012). A joint European effort for early warning of severe weather is made by national meteorological offices through the website Meteoalarm (Meteoalarm, 2015). All of these systems can be termed heuristic (by our definition), and so a) it is difficult to assess their utility for different users and b) it is unclear whether the rule for issuing warnings is optimal with respect to any loss function (for example, is the rule designed to minimise maximum potential loss?)

Most operational EWSs issue various levels of warning if the predicted probability of occurrence or the predicted magnitude of the hazard is above a certain threshold (e.g. see Iervolino et al. (2006) where an earthquake alarm is triggered if the probability of intense ground motion is high enough). The thresholds are often chosen empirically, e.g. based on localised past damages to infrastructure. However, Martina et al. (2006) used Bayesian decision theory to optimally estimate rainfall thresholds for issuing flood warnings at particular river sections.

Mylne (2002) discusses the evaluation of weather forecasts based on a binary model of user utility. The idea of user actions and associated losses conditional on weather forecasts is introduced, and the expected losses are used to evaluate the forecasts—as opposed to evaluating them solely on forecast skill. This can be considered a first step towards using decision theory for issuing warnings, since actions (warnings) have losses attached to them. The second (missing) step is the strategy for taking optimal action, discussed in the subsequent section. See also Katz and Murphy (1997), Richardson (2012) and references therein for more on value and utility of forecasts.

Krzysztofowicz (1993) seems to be the only published paper that advocates Bayesian decision theory as a way of issuing flood warnings. A flood forecasting system was proposed to estimate the probability of flood occurrence, which was then used in conjunction with a binary utility function of warnings to construct a rule that issues warnings to maximise expected utility. In that respect the essence in Krzysztofowicz (1993) is the one proposed here: warnings are decisions under quantifiable uncertainty and so Bayesian decision rational can be used to issue them optimally.

In Medina-Cetina and Nadim (2008) a Bayesian network is described, which is used to integrate empirical, theoretical and subjective information into a probabilistic joint measure for the hazard. Although not designed as a tool for optimally issuing warnings, the method considers the event of issuing a warning given the available information as a stochastic node in the Bayesian network. This implies that the potential for a decision theoretic approach is there, if one were to extend the Bayesian network to an influence diagram (Smith, 2010) by incorporating decision and utility nodes for the warnings.

3 A Bayesian approach to hazard warning systems

In decision making, a decision needs to be made regardless of whether the data are poor or non-informative. A warning must be issued even if the predictive information relating to the hazard is limited, noting that not issuing a warning is also a decision in itself, with possibly dire consequences. Bayesian decision theory provides a coherent and transparent framework for making optimal decisions, using probability to express the uncertainty about the future and a loss function to quantify the consequences from the various actions (warnings).

3.1 A framework for hazard warnings

We have already defined the state space $\{x\}$, the action space $\{a\}$, the predictive information y, the loss function $L(a, x)$ and the conditional probability $p(x|y)$. A decision rule is a rule that maps y onto a, and Bayesian decision theory provides one such optimal rule (Berger, 1985), namely the Bayes rule, defined as

$$d^* = \arg\min_d \mathbb{E}[L(d(Y), X)]. \tag{1}$$

The Bayes rule theorem implies that problem (1) has a unique solution, namely

$$d^*(y) = \min_a \mathbb{E}[L(a, X)|Y = y] = \min_a \int_X L(a, X)p(X|Y)dX. \tag{2}$$

In other words the optimal decision $d^*(y)$, for given predictive information y, is to take action a that minimises the expected loss. So for a given set of warnings $\{a\}$, the optimal warning is a transparent function of just two things: the loss function $L(a, x)$ and the conditional probability $p(x|y)$.

The Bayesian warning system can be depicted by an influence diagram given in Figure 1. The arrow from state of nature x to predictive information y captures the belief that x and y are indeed related. The state of nature is not connected to the action node as it is unknown at the time that action is taken; only y is known and hence connected to $d^*(y)$. The loss function evaluating the consequence of issuing a warning is a function of x and the decision $d^*(y)$.

To put things in context consider the application in this paper which is the UK Met Office NSWWS first-guess warning system introduced in section 2 where y is an ensemble of m weather forecasts. The warning space is $a = \{\text{Green, Yellow, Amber, Red}\}$ and the state space is a set of severity categories of weather variables $x = \{$very low, low, medium, high$\}$. To formulate this problem using the proposed framework, one would need to first quantify the probability $p(x|y)$ of the state of nature given the information in the forecasts y. This can be done using statistical modelling on historical observations of x and y, as illustrated in section 4.3. Second, there is the challenging task of constructing the loss function, $L(a, x)$, which here would be a 4×4 table shown in Table 1. The values l_{ij} quantify the losses from issuing warning i for weather state j and will be different for different users of the system, e.g. the forecaster (issuer of the warning) and a householder. Eliciting $L(a, x)$ is the most difficult part of the assessment but equally the most important one: an agency responsible for issuing warnings should have no right doing so if it is not able to quantify losses. Section 4 illustrates how the values in Table 1 can be determined for two generic stakeholders.

		Weather intensity x			
		very low	low	medium	high
	Green	l_{11}	l_{12}	l_{13}	l_{14}
Warnings	Yellow	l_{21}	l_{22}	l_{33}	l_{24}
a	Amber	l_{31}	l_{32}	l_{33}	l_{34}
	Red	l_{41}	l_{42}	l_{43}	l_{44}

Table 1: Loss table $L(a, x)$ for weather warnings

With regard to any heuristic decision rule, we can now ask the question of whether this is Bayes rule for any particular loss function. If it is, then that loss function can scrutinised and compared to other alternatives, for example the loss functions proposed in section 4. If not, as is the case with the MOrule, then what is the justification for the decision rule if not decision theory?

Note that a good decision rule can reduce loss and that will depend on how much the losses vary across actions in each state of nature, and also by how much this varies from state to state. In other words, the more the losses vary within and between states of nature, the more useful a decision rule becomes. Having a large action space is a good way to get bigger benefit from a well designed decision rule such as the Bayes rule. Of course, the extend to which losses are reduced also depends on how well y predicts x.

4 Example application to severe weather warnings

This section demonstrates the Bayesian framework for issuing hazard warnings by application to precipitation data that were used in the past by the first-guess NSWWS used operationally by the UK Met Office (see section 2 for a description).

4.1 UK Met Office severe weather warning system

The UK Met Office NSWWS (Neal et al., 2014), provides warnings to civil responder services and the public using a risk-based "traffic light" colour scheme where risk is assessed as a combination of likelihood and impact severity using the matrix illustrated in Figure 2. There are four warning levels (Green, Yellow, Amber, Red) and these are associated with top-level responder advice of "no severe weather", "be aware", "be prepared" and "take action". Warnings are issued subjectively by forecasters using a range of tools to assess the combination of likelihood and impact. Ensemble forecasting systems (EFSs) provide guidance on likelihood, but forecasters also make use of output from a range of forecast models. In an EFS a numerical weather model is run many times under slightly different initial conditions to form an ensemble of predictions as a way of quantifying the uncertainty about the future state of weather. Impact is judged on a range of thresholds based on accumulated experience of aspects of societal vulnerability in different parts of the UK. Forecasters are also aided by an ensemble-based first-guess tool (used in this study) which uses the same likelihood-impact table shown in Figure 2. The tool assesses the likelihood of severe weather categorised as "very low", "low", "medium" and "high" using a range of thresholds which vary geographically according to climate and vulnerability to represent impact. The tool assumes perfect forecasts so that the probability of say a medium intensity event is calculated as the empirical frequency of medium intensity from the ensemble members. The rule (MOrule) is then to choose the highest level warning from the table (see Appendix A.1 for a mathematical definition of the rule), e.g. if there is high likelihood of low impact weather (i.e yellow warning) and a low likelihood of high impact weather (i.e. amber), then an amber warning is issued.

The MOrule is heuristic: no explicit loss function is defined (e.g. what is the consequence of a false alarm?) and hence it is not clear whether it is actually optimal in any way. In addition, ensemble frequencies are used instead of the probability of the hazard given the information from an EFS, i.e. assuming that numerical model output equals reality. In the rest of this section we use historical data, that would have been used by the NSWWS, to construct a Bayesian severe weather EWS as an alternative tool that does not suffer from those issues.

4.2 Data

The available data for this study comprise 12-hourly observations of daily precipitation totals (in mm) for the county of Devon, along with matching forecasts, for the two extended winters of October 2012–March 2013 and October 2013–February 2014. Precipitation is categorised as "very low", "low", "medium" and "high" using threshold values of 18, 25 and 30mm respectively. Table 2 shows an example of the data. The state of nature is labelled $x \in \{1, 2, 3, 4\}$ where increasing numerical value respects increasing intensity. Precipitation forecasts, 24 hours ahead, come from the ensemble forecasting system of the European Centre for Medium-Range Weather Forecasts (ECMWF). This comprises of an $m = 51$ member ensemble implying that there are m forecasts of x. The forecast variable has eight categories defined by precipitation thresholds given in the bottom half of Table 2, and is characterised by the vector $z = (z_1, z_2, \ldots, z_8)$ where z_i is the number of ensemble members in category i.

The probability model introduced in the subsequent section will be estimated over 2012–2013 extended winter period (324 12-hourly values), defined as the "estimation period". Then the model will be used to sequentially predict each 12-hourly time step in the 2013–2014 winter (278 values), defined as the "validation period", updating the model each time accordingly.

Observations	V. low 0-18	Low 18-25	Medium 25-30	High > 30mm	x
27/Oct/13 12:00UTC	0	1	0	0	2
28/Oct/13 00:00UTC	0	0	0	1	4
28/Oct/13 12:00UTC	1	0	0	0	1
...			...		

Forecasts z	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8
Thresholds	0-5	5-10	10-15	15-18	18-20	20-25	25-30	> 30mm
27/Oct/13 12UTC	5	20	16	4	3	3	0	0
28/Oct/13 12UTC	0	0	0	0	2	3	11	35
28/Oct/13 12UTC	51	0	0	0	0	0	0	0
...					...			

Table 2: Example of how observations (state of nature) x and the 8-category ensemble forecasts y are defined.

4.3 A simple probability model for $p(x|y)$

We start by quantifying the marginal probability $p(x)$ as the empirical frequency of each of the four values of x,

$$p(x = j) = \frac{n_j}{n}, \qquad j = 1, 2, 3, 4 \tag{3}$$

where n_j is the number of observed x in category j out of n observations. For the estimation period, $p(x) = (0.88, 0.05, 0.02, 0.05)$.

Before proceeding to consider the form of the predictive information y, we note that the forecasts z contain many zero values e.g., $z = (5, 20, 16, 4, 3, 3, 0, 0)$ (first row of Table 2) with corresponding frequency $(0.09, 0.39, 0.31, 0.08, 0.06, 0.06, 0, 0)$. Interpreting frequency as the forecast probability in each of the eight categories, implies that categories 7 and 8 are impossible. This does not reflect our belief that any category is possible at any time and we therefore apply 'add-one smoothing' (see Murphy (2012) p.79). The forecasts are therefore redefined as $z' = z + \mathbf{1} = (z_1 + 1, \dots, z_8 + 1)$. In the example of the first row of Table 2, the new frequency is $(0.10, 0.36, 0.29, 0.08, 0.07, 0.07, 0.02, 0.02)$.

For the sake of simplicity, we consider a simple univariate value as the predictive information y, that is representative of (forecast) precipitation intensity. We define $y \in \{1, \dots, 8\}$ as the modal label of z. In other words y is such that $z_y = max(z_1, \dots, z_8)$, and in case of tied values, y is chosen as the label closest to the second-most-represented label.

We can now approximate the probability $p(y|x)$ as the empirical frequency of y in each of the four x categories,

$$p(y = k|x = j) = \frac{n_{k,j} + 1}{\sum_{k=1}^{8}(n_{k,j} + 1)} \tag{4}$$

where $n_{k,j}$ is the number of observed y taking the value k when observed x is in category j. Again add-one smoothing is used to reflect our belief that there is non-zero probability of a particular forecast category being dominant.

Using Bayes' theorem, we now have what is needed to calculate $p(x|y)$, i.e.

$$p(x = j|y = k) = \frac{p(y = k|x = j)p(x = j)}{p(y = k)} = \frac{p(y = k|x = j)p(x = j)}{\sum_{j=1}^{4} p(y = k|x = j)p(x = j)} \tag{5}$$

which can be easily computed. Figure 3 shows predictions of $p(x|y)$ for each of the eight values of y, based on data from the estimation period. Note that add-one smoothing ensures that $p(x|y)$ is well-defined for each of the eight labels. The plots suggest that there is more confidence in predicting x for low values of y, reflecting also the fact that the majority of the data are concentrated at low values of x and y. Overall, there seems to be good association between x and y.

4.4 Model comparison

We use "CAL" to name the model in (5), as in 'calibration' of the forecasts borrowing from the nomenclature in ensemble forecasting. For comparison purposes, we also consider $p(x)$ (the empirical frequency of each of the four values of x given in equation (3)) as a model for predicting x and denote that as "CLIM" (as in climatology). At third model is also considered, one that assumes the ensemble forecasting system is a perfect representation of the real world and so relative frequencies of ensemble forecasts suffice. We call this model "ENS". Note that this is how "likelihood" (of weather) is defined for the first-guess NSWWS, although for ENS we use add-one smoothing to avoid probabilities of 0 or 1.

The three models were used to sequentially predict precipitation in the validation period (2013–2014). After each prediction of a 12-hourly time step, models CLIM and CAL were updated accordingly, as would have been done in an operational setting. The Brier score (Broecker, 2012), a commonly used forecast verification score, was used to assess the predictive performance of each model:

$$B_j = \frac{1}{n} \sum_{t=1}^{n} \Big(\theta_j(t) - \mathbb{I}(x(t) = j) \Big)^2 \qquad j = 1, \ldots, 4 \tag{6}$$

This is a 'proper' scoring rule widely used in forecast verification and smaller values imply increased forecast skill. The Brier scores for each precipitation category are shown in Figure 4, indicating that model CAL has overall most skill. Bootstrapping was used to compute 95% confidence intervals of the scores, expressing estimation uncertainty (see Appendix **??** for details).

We also use 'reliability' to asses the predicted probabilities. The probability forecast θ_j, $j = 1, \ldots, 4$ for the binary event $b_j = 1$ if $x = j$ and $b_j = 0$ otherwise, is reliable if $\Pr(b_j = 1|\theta_j) = \theta_j$ (Broecker, 2012). In practice however, even if the forecasting system is reliable, there will be discrepancies between $p_j = \Pr(b_j = 1|\theta_j)$ and θ_j since

p_j has to be estimated. Reliability diagrams plot p_j against θ_j to visually assess how far points lie away from the $p_j = \theta_j$ line (the diagonal). Figure 5 shows reliability diagrams for models ENS and CAL. The consistency bars that have been added along the diagonal (see Appendix A.2 for details) are such that under reliability the points should fall within the bars 95% of the time. The plots indicate that ENS is not an empirically reliable forecasting system (most points are outside the consistency bars) unlike CAL where all points fall within the intervals. Note that to produce the diagrams the forecast probabilities θ_j were binned, and different bins were used for ENS and CAL as the θ_j for each system had different ranges. Note also that by definition, model ENS without add-one smoothing used by the first-guess NSWWS, is not reliable as it can issue $\theta_j = 0$ or 1 on occasions where $b_j = 1$ or 0.

4.5 The loss function

The loss function $L(a, x)$ quantifies the stakeholders' loss for each possible combination of action and state of nature. Here there are four states of precipitation intensity (very low, low medium and high) and four levels of warning: Green (no warning), Yellow, Amber and Red. A loss table $L(a, x)$ is thus a 4×4 loss table as illustrated in Table 1. The values of the loss function should reflect preference among the various warning-event combinations, and here the values are chosen relative to the worst case "green warning for high x" given the value 100, and the best case "green warning for very low x" given the value 0.

We consider two generic stakeholders: an insured householder and the issuer of the warnings to the general public, i.e. the forecaster. The insured householder will suffer loss for having to put up defences in case of a false alarm and also for not doing so in case of a missed event. Although the house is assumed insured, there will be certain discomfort from a medium intensity precipitation event for which no warning was issued. Table 3 shows our loss function for a generic householder. Note the values in the diagonal are not all zero on the basis that one prefers low precipitation values to high ones.

		Weather intensity x			
		very low	low	medium	high
	Green	0	10	80	100
Warnings	Yellow	5	0	40	70
	Amber	30	20	5	15
	Red	40	20	10	5

Table 3: The loss function of a generic householder.

The forecaster issuing the warnings stands to suffer embarrassment from false alarms and missed events. Our loss function for a generic forecaster is given in Table 4. Note the zeros in the diagonal and the loss of 70 if a red warning is issued very low precipitation, compared to 40 for the householder.

, assuming zero losses in the diagonal and loss of 100 for the worst case of "green warning for high x".

		Weather intensity x			
		very low	low	medium	high
Warnings	Green	0	10	70	100
	Yellow	20	0	10	70
	Amber	50	10	0	10
	Red	70	40	20	0

Table 4: The loss function of a generic forecaster.

The loss functions in Tables 3 and 4 do not necessarily reflect the losses for any particular individual, however they do have to be visible thus allowing users to assess it (and even come up with their own loss functions). The system can of course be adapted to any stakeholder that can provide their own loss function. In fact, it would be straightforward to develop an online service where the stakeholder inputs their own loss function, just once, and then receive a warning (e.g. by text message) based on $p(x|y)$ provided say by the UK Met Office.

Lastly, using a generic loss function can be a useful tool for evaluating various approaches to estimating $p(x|y)$, as illustrated in section 4.6.

4.6 The warning rule

Using the loss functions in Tables 3 and 4, and estimates of $p(x|y)$ from model CAL based on the estimation period, the warning rules for the generic householder and forecaster were computed and shown in Table 5. The rules are quite different for the two stakeholders. No red warnings are issued by the forecaster, due the combination of high losses from false alarms (bottom row of Table 4) and high uncertainty in predicting x for high values of y as shown in Figure 3. The householder is more tolerant to false alarms and hence will receive higher warning levels than the forecaster across the range of y.

Modal label y	1	2	3	4	5	6	7	8
Householder	Green	Yellow	Yellow	Yellow	Amber	Red	Red	Red
Forecaster	Green	Green	Green	Yellow	Amber	Amber	Amber	Amber

Table 5: Bayes' warning rule for the generic householder and forecaster

Bayes' warnings for the householder and forecaster were issued in the validation period (winter 2013-14) and Figure 6 depicts them for the last two weeks of October. The height of the bars indicates the value of y for each 12-hourly time step whereas the colour indicates the warning level. The symbols on top of each bar reflect what category of x actually occurred. Warnings issued using the MOrule are also shown in the bottom panel of Figure 6. The plots indicate that the MOrule is even more conservative in terms of issuing high levels of warnings than the generic forecaster proposed here. In fact, the MOrule system only issued one red warning for the whole winter period (2013–2014) implying perhaps that the rule was designed to issue high-level warnings only when the forecast evidence of severe weather is overwhelming.

For each 12-hour time step in the validation period, Figure 7 shows the householder and forecaster accumulated losses that would have been incurred by issuing warnings from a) the Bayesian system with model CAL for $p(x|y)$, b) the the Bayesian system with model ENS for $p(x|y)$ (i.e. raw ensemble frequencies) and c) the Bayesian system with perfect knowledge about the future. Using the probability model CAL for $p(x|y)$ instead of raw ensemble probabilities (ENS) gave smaller losses for both the householder and the forecaster. Note however that the difference in losses is much less pronounced for the forecaster. The losses incurred by having perfect knowledge of the future indicate a lower bound on how much any system can improve by investing in better predicting x.

4.7 Assessment of the system

Krzysztofowicz (1993) suggested quantifying the probability of false alarm and missed event as a way of assessing a Bayesian EWS. Here, we define a missed event as "green warning for high x" and a false alarm as "amber or red warning for very low x". The associated probabilities are $\Pr(d^*(y) = \text{Green or Yellow}|x = 4)$ and $\Pr(d^*(y) = \text{Amber or Red}|x = 1)$ respectively. These probabilities can be derived from the conditional probability distribution $p(y|x)$ (computed in section 4.3, equation (4)), since the events $[d^*(y) = \text{Red}]$ and $[d^*(y) = \text{Green}]$ simply define regions of the y space (see Table 5). Table 6 shows these probabilities for each of the two stakeholders. The forecaster has a high probability of missing an event suggesting that the loss function might need revising.

	False alarm	Missed event
Householder	0.01	0.03
Forecaster	0.01	0.25

Table 6: Probability of a false alarm or a missed event for each stakeholder.

5 Discussion

Bayesian decision theory was used to propose a simple framework offering a transparent way of issuing hazard warnings. The Bayesian EWS uses probabilistic predictions of the hazard in conjunction with a loss function to issue optimal warnings with respect to expected loss. Some methods for constructing and evaluating the probability of the hazard given relevant predictive information were illustrated. In the application to precipitation warnings, the statistical model proposed to calibrate ensemble forecasts was shown to give smaller losses than simply using raw ensemble frequencies.

It was also illustrated that quantifying consequences using a loss function is crucial in understanding and assessing the EWS.

The transparency of the proposed framework implies that it is open to criticism, updating and tailoring, which in turn means that it can accommodate likely changes in hazard forecasts, exposure and vulnerability. Expressing consequences numerically through a loss function, offers the possibility of issuing bespoke warnings to different users with varying risk appetites.

Note that the framework proposed here can be incorporated in a decision support system in which a human agent makes the final decision. This decision will be based on Bayes rule, which the agent may choose to countermand on the basis of complexities that were not accounted for in predicting the probability of the future state of nature or in constructing the loss function.

The Bayesian model presented here to estimate $p(x|y)$ from historical data was kept deliberately simple, in order illustrate the warning system with more clarity, and to show that even with a simple model of $x|y$ one can improve the accuracy of the predictions compared to using wither x or y on their own. More complicated models can of course be used, with the aim of improving the accuracy of $p(x|y)$. This could be achieved by modelling the 8-category forecast variable given x, instead of just the model label. This should maximise the amount of information that can be obtained from the forecasts but it is left for future work, as is allowing for any autocorrelation in the historical time series.

The Met Office first-guess warning system as presented here is a decision support tool. In practice, more that one ensemble forecasting system may be used as well as a deterministic system and the warnings actually issued are finalised by forecasters using subjective judgements and an assessment of societal vulnerability. The current warning level might have an effect on what warning will be issued next and forecaster will act upon their personal subjective beliefs and prior knowledge, adjusting the warning level as appropriate. Some of these particularities can be added to the proposed framework—for instance considering information from other forecasting systems or even forecasts at different lead times as the predictive information y when building the model for $p(x|y)$; or making the the loss functions can depend upon the current warning level. However, not everything in the forecaster's work can be replaced by a mathematical approach but at least the underlying system providing them with a suggested warning to issue should be transparent and defensible.

Acknowledgements

This work was supported by the Natural Environment Research Council [Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learning and Elicitation (CREDIBLE); grant number NE/J017043/1]. We wish to thank Rutger Dankers, Stefan Siegert for their valuable input.

A Appendix

A.1 Likelihood-impact matrix

Here the Met Office rule (Figure 2) is defined mathematically. Suppose the weather variable of interest is w with support $[W_l, W_u]$ and let the four w categories (very low, low, medium and high) be defined by intervals (W_l, u_1), $(u1, u2)$, $(u2, u3)$ and $(u3, W_u)$ respectively. The forecast probabilities $\theta_j = y'_j/m$ are the frequencies of ensemble members in each category $j = 1, \ldots, 4$. For convenience, define the following three probabilities: $p_1 = \Pr(W > u_1) = 1 - \theta_1$, $p_2 = \Pr(W > u_2) = \theta_3 + \theta_4$ and $p_3 = \Pr(W > u_3) = \theta_4$. If we relabel the warnings $d(y)$ (green, yellow, amber, red) into $(1, 2, 3, 4)$, then the Met Office warning rule is

$$d(y) = 1+\mathbb{I}(p_1 > 0.4)||\mathbb{I}(p_2 > 0)||I(p_3 > 0)+\mathbb{I}(p_2 > 0.4)||\mathbb{I}(p_3 > 0.2)+\mathbb{I}(p_3 > 0.6) \quad (7)$$

where $\mathbb{I}(S)$ is 0/1 if S is true/false and the symbol $||$ denotes the logical statement "or".

A.2 Reliability diagram

Consider forecast probabilities $\theta_j(t)$, $t = 1, \ldots, n$, $j = 1, \ldots, 4$, of binary events $b_j(t) = \mathbb{I}(x(t) = j)$. A reliability diagram effectively plots $b_j(t)|\theta_j(t)$ against $\theta_j(t)$. One way to achieve this is to bin $\theta_j(t)$, then calculate $\bar{b}_g$ (the observed frequency of $b_j(t)$ in each bin g) and then plot $\bar{b}_g$ against $\bar{\theta}_g$ (the mean of $\theta_j(t)$ in each g). If the forecasts are reliable then the points on such a plot should lie "near" the 45° line, but not exactly due to sampling variability. 95% consistency bars can be added on the diagonal to asses how much the points would be expected to vary under the assumption of reliability. The R package 'SpecsVerification' (Siegert, 2014) creates such bars by bootstrapping the forecasts $\theta_j(t)$ and then simulating $z_j(t)$ under reliability (i.e. $\Pr(z_j(t) = 1) = \theta_j(t)$). If too many points lie outside the consistency bars then reliability can be rejected. See Broecker (2012) and references therein for more details of reliability diagrams.

References

Alfieri, L., Salamon, P., Pappenberger, F., Wetterhall, F., and Thielen, J. (2012). Operational early warning systems for water-related hazards in Europe. *Environmental Science & Policy*, 21:35–49.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis.* Springer-Verlag, New York, 2nd edition.

Bhattacharya, D., Ghosh, J., and Samadhiya, N. (2012). Review of geohazard warning systems toward development of a popular usage geohazard warning communication system. *Natural Hazards Review*, 13(4):260–271.

Broecker, J. (2012). Probability Forecasts. In Jolliffe, I. and Stephenson, D., editors, *Forecast verification: A practitioner's guide in atmosheric science*. Wiley-Blackwell, 2nd edition.

Dow, K. and Cutter, S. L. (1998). Crying wolf: Repeat responses to hurricane evacuation orders. *Coastal Management*, 26(4):237–252.

Environment Agency (2015). Flood warnings summary. `http://apps.environment-agency.gov.uk/flood/31618.aspx`. Accessed: 04/Feb/2015.

Iervolino, I., Convertito, V., Giogrio, M., Manfredi, G., and Zollo, A. (2006). Real-time risk analysis for hybrid earthquake early warning systems. *Journal of Earthquake Engineering*, 10(6):867–885.

Katz, R. and Murphy, A. (1997). Forecast value: prototype decision making models. In Katz, R. and Murphy, A., editors, *Economic value of weather and climate forecasts*. Cambridge.

Krzysztofowicz, R. (1993). A theory of flood warning systems. *Water Resources Research*, 29(12):3981–3994.

Lindley, D. V. (1985). *Making decisions*. Wiley, 2nd edition.

Martina, M. L. V., Todini, E., and Libralon, A. (2006). A Bayesian decision approach to rainfall thresholds based flood warning. *Hydrology and Earth System Sciences*, 10(3):413–426.

Medina-Cetina, Z. and Nadim, F. (2008). Stochastic design of an early warning system. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 2:223–236.

Met Office (2015). Met Office National Severe Weather Warnings. `http://www.metoffice.gov.uk/public/weather/warnings`. Accessed: 04/Feb/2015.

Meteoalarm (2015). Weather warnings: Europe. Network of European Meteorological Services, Ukkel, Belgium. `http://www.meteoalarm.eu`. Accessed: 04/Feb/2015.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge MA, USA.

Mylne, K. R. (2002). Decision-making from probability forecasts based on forecast value. *Meteorological Applications*, 9(3):307–315.

Neal, R. A., Boyle, P., Grahame, N., Mylne, K., and Sharpe, M. (2014). Ensemble based first guess support towards a risk-based severe weather warning service. *Meteorological Applications*, 21(3):563–577.

NOAA (2015). Hurricane Preparedness - Watches and Warnings. `http://www.nhc.noaa.gov/prepare/wwa.php`. Accessed: 11/Feb/2015.

PTWC (2015). Pacific Tsunami Warning Center. http://ptwc.weather.gov/. Accessed: 11/Feb/2015.

Richardson, D. (2012). Economic value and skill. In Jolliffe, I. and Stephenson, D., editors, *Forecast verification: A practitioner's guide in atmosheric science*. Wiley-Blackwell, 2nd edition.

Rittel, H. W. J. and Webber, M. M. (1973). Dilemmas in a General Theory of Planning. *Policy Sciences*, 4(2):155–169.

Siegert, S. (2014). *SpecsVerification: Forecast verification routines for the SPECS FP7 project*. R package version 0.3-0.

Smith, J. Q. (2010). *Bayesian decision analysis: Principles and Practice*. Cambridge.

Sorensen, J. (2000). Hazard warning systems: review of 20 years of progress. *Natural Hazards Review*, 1(2):119–125.

USGS (2015). U.S. Volcanoes and Current Activity Alerts. http://volcanoes.usgs.gov/. Accessed: 11/Feb/2015.

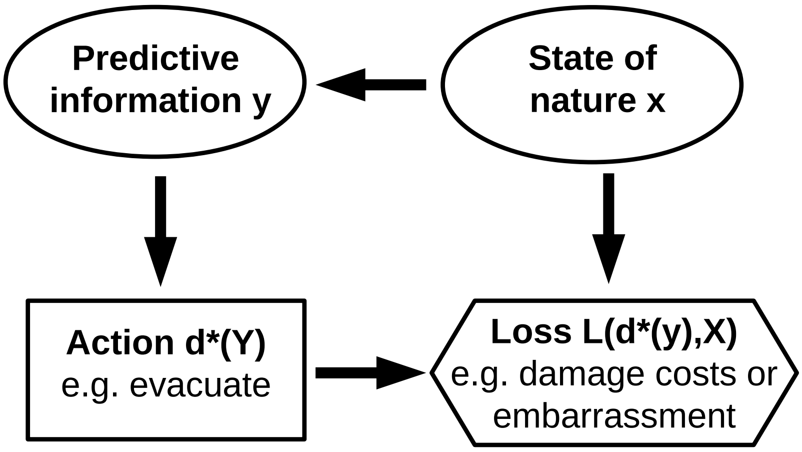

Figure 1: Influence diagram describing the decision problem of issuing hazard warnings. Oval nodes indicates uncertain quantities, rectangular nodes relate to decisions and hexagonal nodes relate to losses.

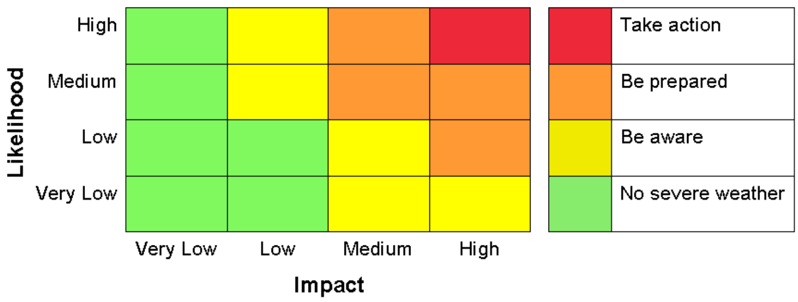

Figure 2: Likelihood-impact matrix. "Impact" refers to the magnitude of the state of nature and "likelihood" refers to the chance of this occurring according to the ensemble forecasting system. The likelihood categories are $< 20\%$ for "Very Low", 20%–40% for "Low", 40%–60% for "Medium" and $> 60\%$ for "High".

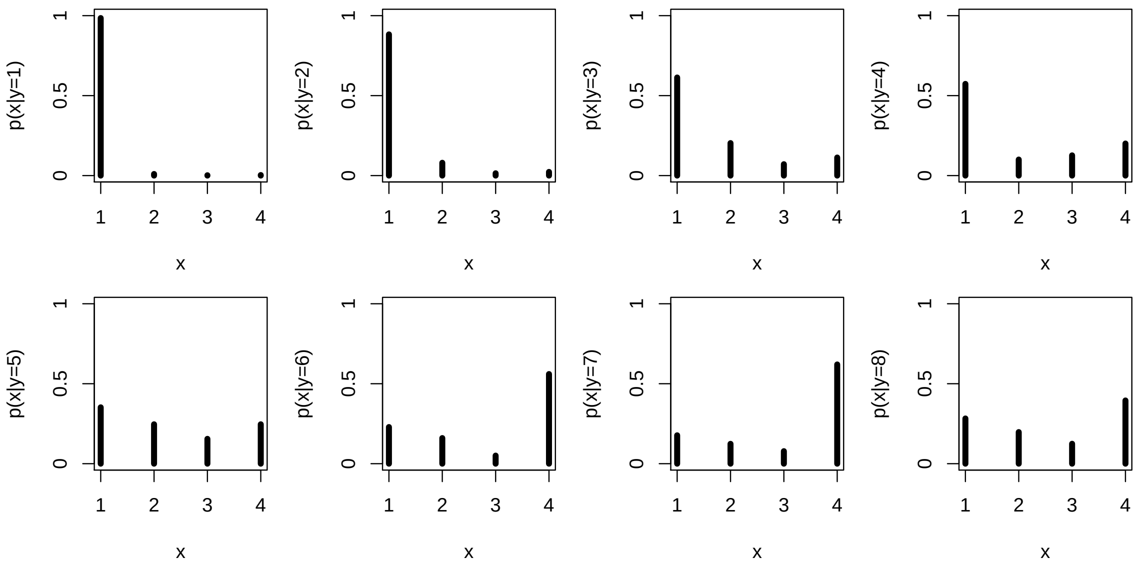

Figure 3: Estimates of the probability $p(x|y)$ for each y category.

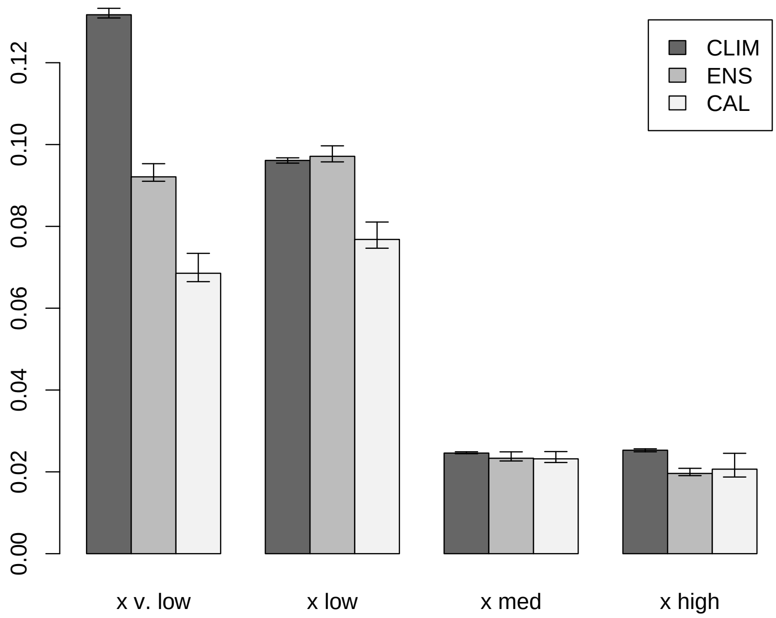

Figure 4: Each set of bars represents the Brier scores for each of the three models (CLIM, ENS and CAL) in each x category. 95% bootstrap intervals, given as "whiskers" at the top of each bar, were calculated by re-sampling the data with replacement 5000 times.

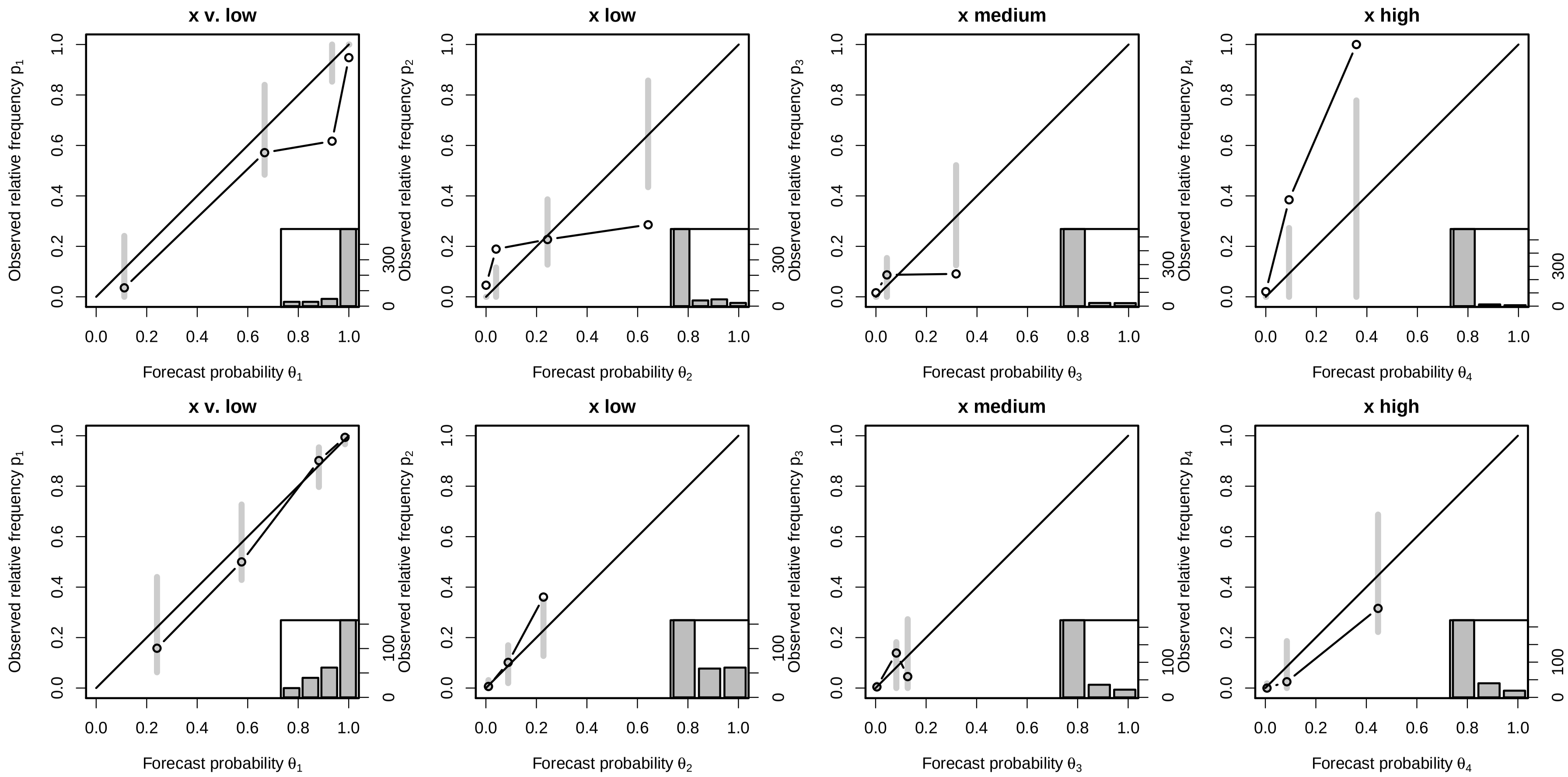

Figure 5: Reliability diagrams for model ENS (top panel) and model CAL (bottom panel). The histograms on the bottom right of each plot indicate the number of points in each bin used to construct the diagrams. The 95% consistency intervals indicate the variability in p_j that would be expected under the assumption of reliability.

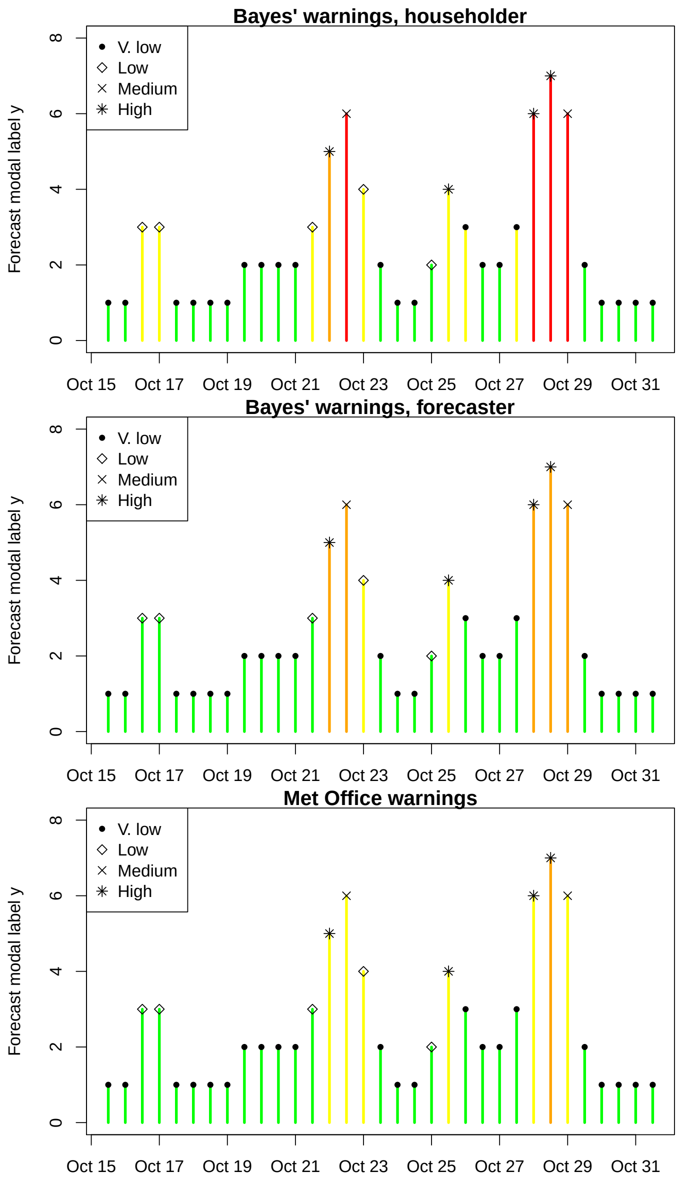

Figure 6: Plots of modal label y against time (12-hourly steps) for the last two weeks of October 2013 (the first month in the validation winter). Top and middle panels show Bayes' warnings for the householder and forecaster respectively. The bottom panel shows warnings based on the Met Office rule. The bars are coloured according to the warnings issued. The actual x values are shown using as symbols on top of each bar.

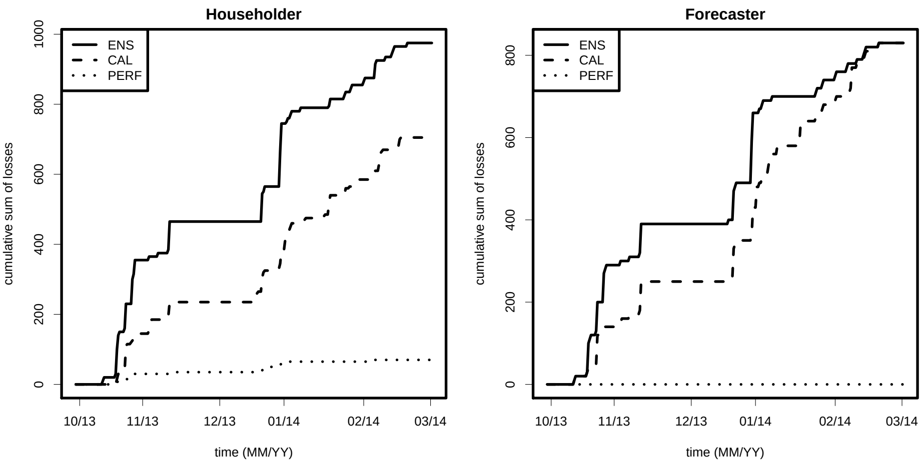

Figure 7: Plots of cumulative losses, for each generic stakeholder, that would have been incurred in the validation period (2013–2014), if warnings were issued using Bayes' rule with a) model ENS probabilities (solid line), b) model CAL probabilities (dashed line), and c) perfect knowledge (dotted line).