Designing environmental uncertainty information for experts and non-experts: does data presentation affect users' decisions and interpretations?

Article

Accepted Version

Mulder, K. J., Lickiss, M., Black, A., Charlton-Perez, A. J. ORCID: https://orcid.org/0000-0001-8179-6220, McCloy, R. ORCID: https://orcid.org/0000-0003-2333-9640 and Young, J. S. (2020) Designing environmental uncertainty information for experts and non-experts: does data presentation affect users' decisions and interpretations? Meteorological Applications, 27 (1). e1821. ISSN 1350-4827 doi: https://doi.org/10.1002/met.1821 Available at https://centaur.reading.ac.uk/85132/

To link to this article DOI: http://dx.doi.org/10.1002/met.1821

Publisher: Wiley

Designing Environmental Uncertainty Information for Experts and Non-Experts: Does Data Presentation Affect Users' Decisions and Interpretations?

Kelsey J. Mulder

*Department of Meteorology, University of Reading, United Kingdom**

Matthew Lickiss, Alison Black

Department of Typography and Graphic Communication, University of Reading

Andrew Charlton-Perez

Department of Meteorology, University of Reading

Rachel McCloy

Department of Psychology, University of Reading

Joe Young

Department of Atmospheric Sciences, University of Utah, USA

*Author Current Affiliation: Hiscox, 1 Great St Helen's, London EC3A 6HX, United Kingdom.

E-mail: kelsey.mulder@hiscox.com

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/met.1821

Abstract

Uncertainty information in natural hazard forecasts is increasingly being explicitly communicated. This study was designed to determine whether different ways of communicating uncertainty graphically affects decisions and interpretations of forecasts and whether expertise was a factor in decisions and interpretations from forecasts explicitly showing uncertainty. In a hypothetical decision-making task regarding ice thickness and shipping, 138 experts and non-experts received ice-thickness forecasts in four different presentations expressing uncertainty: worded probability, spaghetti plot, fan plot, and box plot. These forecasts contained no measures of central tendency. There was no consistent difference in decision or best-guess forecast (deterministic ice thickness forecast based on the forecast representation) between the different forecast representations. However, participants interpreted different amounts of uncertainty across the different forecast representations. Experts made significantly more economically rational decisions than non-experts, interpreted lower best-guess forecasts, and inferred significantly more uncertainty than non-experts. These results suggest that care be taken in choosing how uncertainty is represented in forecasts, especially between expert and non-expert audiences.

Keywords: Communication, Uncertainty, Decision Making

1. Introduction

Because of the devastating effects natural hazards can have on lives and property, it is natural to try to forecast these events to lessen negative impacts. However, natural hazard forecasts are uncertain by nature due to imperfect observations and models (Galesic *et al.*, 2016) as well as the intrinsic chaos of the natural system. Rather than giving a deterministic (i.e. single-value) forecast, some natural hazard forecasts are now given with measures of uncertainty. The measures of uncertainty include range forecasts (e.g., high temperature of 68–74°F) and likelihood of occurrence (30% chance of rain), among others, and are designed to communicate the confidence of the forecasts to the users.

The addition of uncertainty to natural hazards forecasts has been shown to help users interpret complex information in a way that is useful to decisions. For example, Nadav-Greenberg and Joslyn (2009) provided temperature forecasts and asked participants to decide whether or not to protect roads against ice. Participants' decisions were more economically rational when the forecasts included uncertainty information. These participants also actively chose to use uncertainty information: when given a deterministic forecast and the choice to use uncertainty information, 85% of participants chose to use uncertainty information (Nadav-Greenberg and Joslyn, 2009). Other decision-based studies have found that lay people can successfully interpret uncertainty information for use in decision-making in situations such as interpreting a 5-day temperature forecast (Roulston and Kaplan, 2009), deciding whether to protect fruit from frost or releasing water from a reservoir (Morss *et al.*, 2010), or using river flood defences (Ramos *et al.*, 2013).

Although uncertainty information has been shown to help decision-making, there are many different ways of expressing uncertainty information, which may lead to different interpretation of the information. For example, Henrich *et al.* (2015) tested interpretation of earthquake risk in terms of frequencies and time frames (1600 dead in 500 years, 10% chance of 1600 dead in 50 years, 2 deaths per year) and sample frame (1.9 deaths per 100,000, 19 deaths per million). They found that relating the risk to a time frame and stating a probability associated with it (e.g., 10% chance of 1600 dead in 50 years) was found as riskier than other ways of framing the risk, even though the risk was the same across presentation modes. Additionally, the phrase "there is a 30% chance of rain tomorrow" leads to multiple different interpretations, such as 30% of the day it will rain, 30% of the area will experience rain, or on 30% of days like this one, it rained (Gigerenzer *et al.*, 2005; Abraham *et al.*, 2015). However, worded communication is not the only way forecasts are disseminated. Because natural hazards vary over time, time series are a straightforward and therefore common way of communicating forecasts. Therefore, communication of uncertainty in graphical form is also important. Similar to expressing uncertainty verbally, graphically displaying uncertainty can be done in many different ways, also leading to different interpretations. For example, in a study testing three map-based tornado probability forecasts (deterministic polygon, five-probability gradient, and ten-probability spectrum) with university students in the United States, the deterministic polygon promoted the highest overall fear and protective action (Ash *et al.*, 2014). However, there was an increased response in the highest risk zones for the gradient and spectrum forecasts, showing participants were more likely to take protective action and were more fearful if they were located in a high-risk area if given these forecasts (Ash *et al.*, 2014). When testing between different formats of box plots to predict voter turnout, snowfall, and payout from a fund, Correll and Gleicher (2014) found that participants could make adequate decisions using the box plots, but that their responses changed if subtle graphical changes were made to the box plot, such as adding a point to the graph and asking the likelihood of an outcome relative to that point. Similarly, when interpreting temperature data between different styles of line graphs (e.g., random lines, line with bands around, line with error bars), participants tend to infer normal distributions around points along the line graphs, but the distribution of uncertainty inferred around each point changed based on the way uncertainty was presented (Tak *et al.*, 2015).

However, different forecast representations do not always lead to different decisions. Nadav-Greenberg and Joslyn (2009) found that decisions on de-icing roads given temperature forecasts were no different when comparing between visual box plots and worded frequency. Similarly, when comparing visualisations communicating wildfire risks, participants' decisions changed very little when they were given plenty of time to consider their decisions (Cheong *et al.*, 2016). However, under time pressure, they made more rational decisions by using a representation with a colour-filled multi-boundary map and the least rational decisions using worded probability and single boundary maps (Cheong *et al.*, 2016).

There are three aspects of previous research that warrant further exploration. First, most of the research listed above compared decisions given different forecast representations for forecasts familiar to the user (e.g., temperature, Nadav-Greenberg and Joslyn, 2009). However, familiarity with forecasts can introduce personal bias into decisions and inferences (e.g., Tversky and Kahneman, 1974). Therefore, further research should be conducted on natural hazard decisions unfamiliar to users to reduce bias. Second, there is a discrepancy in previous research on whether there are significant differences in interpretation of forecast information if it is given in different formats. Third, many studies are conducted on members of the public (e.g., Morss *et al.*, 2010), students (e.g., Savelli and Joslyn, 2013), or expert groups (e.g., Cox *et al.*, 2013), but rarely comparing the responses between expert and public groups (e.g., Doyle *et al.*, 2014). Other studies used expert and stakeholder groups for a very specific hazard (e.g., Thompson *et al.*, 2015; Mulder *et al.*, 2017). In real-time decisions, both experts and non-experts may be using the same information. It is therefore important to compare how experts and non-experts interpret forecast information from different types of graphical forecast representations.

The purpose of this research is to conduct a decision-based survey on an unfamiliar hazard to determine how decisions and interpretations of forecast information differ between different forecast representation and between expert and non-expert groups. Specifically, this paper seeks to answer the following questions:

1. How does the presentation of uncertainty in forecast representations explicitly showing uncertainty affect participants' interpretation of, and decisions based on, this forecast?
2. Is there a difference in interpretation of, or decisions based on, forecasts explicitly showing uncertainty between experts and non-experts?

We hypothesize that the interpretations and decisions will change given different forecast representations because they may cause the user to focus on different aspects of the forecast (e.g., a fan plot may focus the user on extreme values more than a box plot). We also hypothesize that experts will perform better in decision tasks and interpret the forecasts differently than non-experts because of their experience with this type of information.

2. Methods

2.1. Survey Instrument

In the survey, which was implemented both on paper and online, participants were told they were the manager of a vegetable company which ships its vegetables across an often-frozen, fictional strait. Based on 72-hour ice-thickness forecasts, participants were asked to choose which ship to use: a small ship (costing a fictional £1,000) which could crush through ice up to 1-meter thick or a large ship (costing a fictional £5,000) which could crush through ice up to 5-meters thick. If the ice was too thick for the small ship, there would be an additional £8,000 charge due to vegetable spoilage. No real monetary incentives were involved in the survey and no results of actual ice thickness were given during participation.

Ice thickness was chosen for the decision task because these are not standard forecasts and were unlikely to have been used by our participants in the past, meaning the participants were not biased by past experience of these forecasts (e.g., Tversky and Kahneman, 1974). Participants did not receive any training on how to interpret the forecasts.

The decision question was classified as correct or incorrect based on economic rationality. Based on the costs and losses given to participants, the forecast probability at which it was economically rational (according to expected value theory, Bernoulli, 1954) to choose the large ship was 56%. Participants were also asked their single-value (or best-guess) forecast for ice thickness 72 hours from now, given the forecast. They were given a short line to input their open-ended response. Participants were not limited to any degree of precision and some responded to 0.001 m. The purpose of this question was to allow participants to interpret a central tendency of the forecast to help test their interpretations. Participants also indicated the maximum and minimum ice thickness they expected in 72 hours. This question was designed to bound participants' forecasts and show how much uncertainty they were interpreting into the forecast.

In the survey, there were five forecast representations: deterministic line (not analysed here), worded probability, spaghetti plots, fan plots, and box plots (Figure 1). The worded probability was used as the control condition in the survey. Although there are many more ways to communicate uncertainty information, including different wordings (such as frequency versus proportions, e.g., Henrich *et al.*, 2015), these representations were chosen based on representations commonly used to communicate uncertainty in hazard information over time. Each forecast representation was created to represent one of three probabilities of ice thickness exceeding 1 meter: 30%, 50%, and 70%. These probabilities were defined by the spaghetti plot, where the number of forecasts exceeding 1 meter corresponded to the probability (e.g., three of ten forecasts exceeding 1 meter represented a 30% probability of exceeding 1 meter). The fan and box plots were derived from the 10th, 25th, 75th, and 90th percentiles of the spaghetti plot. Participants would have been able to deduce that the probability at which it was economically rational to choose the large ship was >56% from the spaghetti plot, but would have to make an educated guess on the other forecast representations.

In this study, participants from an online survey as well as participants from a supporting paper survey were used. The paper survey had a within-participant design, with each participant receiving 24 forecast representations (3 probabilities x 8 forecast representations: deterministic line, worded probability, box with and without a median line, fan with and without a median line, and spaghetti with and without a median line). The online survey was the same as the paper survey except that it was a between-participant design where participants were randomly assigned to either receive uncertainty forecasts with or without median lines. All online survey participants received the worded probability and deterministic forecasts, resulting in 15 forecast representations (3 probabilities x 5 forecast representations: deterministic line, worded probability, box with or without a median line, fan with or without a median line, and spaghetti with or without a median line). The forecast representations at different probabilities were presented in random order for both the paper and online surveys to reduce ordering bias. Participants were asked to answer all questions for each forecast representation separately. Note that participants in both the paper and online surveys were able to take as much time as they needed to complete the survey questions.

This article focuses on the effects of forecast representation, which was within-subject for both the paper and online survey samples. To control for anchoring, where the framing of the information presented may affect a participant's answer (Tversky and Kahneman, 1974), only forecasts without median lines were used. Responses from the paper survey and online sample given forecasts without median lines were combined and used in the present study. Although paper survey participants were given forecasts both with and without median lines, the combination of the randomisation of the forecast representations and time limit in which the participants were asked to complete the survey (around 30 minutes) suggest that effect of receiving both forecasts was minimal on their responses. The omission of the online sample with median lines and inclusion of the paper survey results in 138 total participants.

Data were cleaned such that illogical responses (e.g., maximum ice thickness smaller than minimum) and outliers were removed. Outliers were defined as best-guess forecasts greater than 3 m (the y-axis of the forecast representation only showed up to 3 m), maximum ice thickness greater than 6 m (twice the size of the y-axis of the forecast representation), and minimum ice thickness greater than 2 m (twice the ice thickness threshold the participants were trying to predict). Between 0 and 6 (0–4.8%) survey responses were removed per question for illogical responses and 0–6 (0–4.8%) per question for outliers.

2.2. Participants

Participants were surveyed in spring 2016. Targeted recruitment ensured that multiple user groups participated in the survey: the public, stakeholders who make weather and climate-based decisions in their job, forecasters, and academics. Public participants were recruited in person at University of Reading open, public events and as they entered the University of Reading library. Online, public participants were recruited through social media. Stakeholders and forecasters were targeted through snowball sampling, with the survey link being sent in personalized emails to partners from the Natural Environment Research Council (NERC) Probability, Uncertainty & Risk in the Environment (PURE) programme. Academics were recruited in person at University of Reading atmospheric science workshops and research group meetings.

Participants were classified as either experts (participants who typically create forecast representations, i.e. academics and forecasters, n=79), or non-experts (participants who typically use representations, i.e. public and stakeholders, n=59). These classifications were based both upon which survey link they used (all participants received the same survey; different URLs were used based on targeted user-group) and job description given in response to demographic questions in the survey.

The total 138 participants (49 from the paper survey, 89 from the online survey) comprised 39 members of the public, 20 stakeholders, 19 forecasters, and 60 academics. These participants were aged 18-67, mean 36. There were 52 females, 82 males, and 4 who did not report their gender.

3. Results

3.1. Ship decision

For the following section, the number of economically rational decisions (e.g., choosing the large ship when probability of exceeding 1 m was greater than 56%) for each forecast representation and experts and non-experts were tallied for a total of three possible economically rational decisions. For example, each participant received the worded probability representation for a 30%, 50%, and 70% probability forecast, giving three opportunities for making an economically rational decision (Figure 2).

The interacting effect between forecast representation and expertise on the number of economically rational decisions was not significant ($ANOVA_1$, all ANOVA results can be found in Table 1). There was a significant main effect of expertise on the number of economically rational decisions made ($ANOVA_2$). Experts had significantly more economically rational decisions than non-experts (experts: mean=1.81, sd=0.74; non-experts: mean=1.67, sd=0.65, Figure 2). The main effect of forecast representation on the number of economically rational decisions was not significant ($ANOVA_3$).

3.2. Best-guess ice thickness

It should be noted that none of the forecast representations tested here included a measure of central tendency to control for anchoring.

Participants' best-guess forecasts differed based on forecast representation, probability, and expertise. The combined interacting effect between the three independent variables of forecast representation, probability, and expertise was significant ($ANOVA_4$).

There was a significant interacting effect of forecast representation and probability on best-guess forecasts ($ANOVA_5$). As probability increased, the forecast representation with the highest best-guess forecast changed (Figure 3). The forecast representation with the highest best-guess forecast was the fan plot at 30% (mean=0.97, sd=0.33), spaghetti plot at 50% (mean=1.15, sd=0.34), and the worded probability at 70% (mean=1.48, sd=0.36). This supports the finding above that forecast representation did not affect decisions. In other words, there was no consistent difference in perceived central tendency across forecast representations.

There were no significant interacting effects between expertise and forecast representation ($ANOVA_6$) or expertise and probability ($ANOVA_7$). In other words, the best-guess forecast did not change by combinations of expertise and forecast representation or expertise and probability of exceeding 1-meter ice thickness.

The main effects of probability and forecast representation were significant ($ANOVA_8$ and $ANOVA_9$, respectively). As probability increased, the best-guess forecast increased with 70% probability forecasts (mean=1.39, sd=0.31) significantly greater than 50% and 30% forecasts ($t\text{-test}_1$ and $t\text{-test}_2$, respectively; all t-test results can be found in Table 2).

The forecast representation with the highest best-guess forecast was the fan plot (mean=1.17, sd=0.35), significantly larger than the worded probability, spaghetti, and box plot ($t\text{-test}_3$, $t\text{-test}_4$, and $t\text{-test}_5$, respectively). The best-guess forecasts for the worded probability, spaghetti, and box plots were not significantly different from each other (worded probability vs. spaghetti: $t\text{-test}_6$; worded probability vs. box: $t\text{-test}_7$; spaghetti vs. box: $t\text{-test}_8$).

The main effect of expertise on best-guess forecasts was significant ($ANOVA_{10}$). Experts forecast a significantly smaller best-guess forecast (mean=1.06, sd=0.30) than non-experts (mean=1.24, sd=0.42). This suggests that experts and non-experts interpret different central values for the forecasts, which would in turn affect their decisions based upon these forecasts.

3.3. Interpreting uncertainty in ice thickness

Participants were asked the maximum and minimum possible ice thickness for each given forecast. This information was converted to a range of uncertainty (maximum-minimum) among expert and non-experts and for each forecast representation and probability to determine whether their perception of the amount of uncertainty varied.

The combined interacting effect between the three independent variables of forecast representation, probability, and expertise was significant ($ANOVA_{11}$). There was a significant interacting effect between forecast representation and probability on the amount of uncertainty inferred ($ANOVA_{12}$). At all probability levels, participants inferred the most uncertainty into the spaghetti plots Figure 4, (30%: mean=1.60, sd=0.57; 50%: mean=1.89, sd=0.59; 70%: mean=2.23, sd=0.64) followed by the box plots (30%: mean=1.48, sd=0.58; 50%: mean=1.76, sd=0.65; 70%: mean=2.04, sd=0.68). At 30% and 70%, participants inferred more uncertainty into the fan plot (30%: mean=1.38, sd=0.50; 70%: mean=1.91, sd=0.55) than the worded probability (30%: mean=1.18, sd=0.86; 70%: mean=1.79, sd=1.12). However, at 50%, the amount of inferred uncertainty between the fan plot and worded probability was equal (fan: mean=1.68, sd=0.57; box: mean=1.68, sd=1.14).

The interacting effects between expertise and forecast representation and expertise and probability were not significant ($ANOVA_{13}$ and $ANOVA_{14}$, respectively). In other words, the amount of uncertainty inferred did not vary significantly with different combinations of expertise and forecast representation or probability and forecast representation.

The main effects of forecast representation and probability were significant in predicting the amount of uncertainty inferred ($ANOVA_{15}$ and $ANOVA_{16}$, respectively).

The amount of uncertainty inferred into the forecasts increased with increasing probability, which corresponded to what was shown in the forecasts. The amount of uncertainty inferred for the 70% forecasts (mean=2.00, sd=0.77) was significantly greater than that for the 30% and 50% forecasts ($t\text{-test}_9$ and $t\text{-test}_{10}$, respectively).

The amount of uncertainty inferred from the spaghetti plot was the largest when combining all probabilities (mean 1.91, sd=0.65, Figure 4). The amount of uncertainty inferred from the spaghetti plot was significantly greater than that for the box, fan, and worded probability ($t\text{-test}_{11}$, $t\text{-test}_{12}$, and $t\text{-test}_{13}$, respectively). The amount of uncertainty inferred for the box plot was second highest, controlling for probability and expertise, significantly greater than the fan and worded probability representations ($t\text{-test}_{14}$ and $t\text{-test}_{15}$, respectively).

There was a significant main effect of expertise on the amount of uncertainty participants interpreted from the forecasts ($ANOVA_{17}$). Experts (mean=1.84, sd=0.75) inferred significantly more uncertainty than non-experts (mean=1.57, sd=0.75, Figure 4).

4. Discussion

Because previous research has shown that uncertainty information improves natural hazard-related decisions, it is important to understand how different representations of uncertainty in forecasts influence decisions and interpretations of that information. The current research focused on how different forecast representations showing uncertainty affected participants' interpretations from and decisions based on the forecast and if expertise affected interpretations and decisions for an unfamiliar type of forecast. Because the forecast was unfamiliar to the participants, the amount of personal bias in the forecasts was reduced. We believe these results to be transferrable to natural hazards and forecasts familiar to recipients of that information, however this is an area of possible future research.

The forecast representations were specifically designed to show the amount of uncertainty in different ways based on the same information. Even though the way in which uncertainty was represented differed, there were no consistent differences between participants' decisions or best-guess forecasts across forecast representations, contrary to our first hypothesis. One explanation for this finding could be that participants were given as much time as they needed to complete the survey, making these findings similar to those of Cheong *et al.* (2016). These differences were more salient across different probabilities of the forecast exceeding the ice thickness threshold of 1-meter. These results suggest that the way in which uncertainty is represented does not affect users' interpretation of central tendency. Past research in natural hazard decision making has disagreed on whether changing forecast representation changes decisions. The results from this research are similar to previous research, which suggests that changing the forecast representation did not change decisions (e.g., Nadav-Greenberg and Joslyn, 2009; Cheong et al., 2016). These findings are contrary to Savelli and Joslyn (2013), who suggested that visualisations may be detrimental to decisions when compared with verbal methods of communication. However, our research only tested one form of verbal forecast. Further verbal formats could be tested in future.

The only dependent variable which changed significantly with forecast representation was the amount of uncertainty participants inferred from the forecast. Participants inferred the most uncertainty in the spaghetti plot followed by the box plot, then the fan plot. Participants inferred the least amount of uncertainty for the worded probability representation at 30% and

70%. The amount of uncertainty inferred was equal between the fan plot and worded probability at 50%. This result suggests that, similar to previous literature with more familiar natural-hazard forecasts (e.g., tornado, temperature, hurricane, Ash *et al.*, 2014; Tak *et al.*, 2015; Ruginski *et al.*, 2015), interpretations of forecast information can be different when given different forecast representations for an unfamiliar hazard. Interpreting uncertainty differently depending on differing forecast representations could be problematic in real-life decision-making situations; for example, increasing coastal protection as an adaption to sea-level rise. When uncertainty is not fully understood or considered, these adaptations may be more or less than what is necessary, either costing too much money or not providing enough protection. A limitation of this study is that only a single threshold was studied whereas in many real-life natural hazard scenarios, multiple thresholds exist. An area of future research would be to introduce multiple thresholds.

The reason participants interpreted different amounts of uncertainty based on forecast representation is likely due to the design of the plots. The spaghetti plot explicitly showed single model outcomes whereas the fan and box plots smoothed that information into intervals (Figure 1). Similarly, when testing box plots, Correll and Gleicher (2014) found that the public interpreted points within the boxes to be more likely than those within the whiskers, a similar smoothing effect.

An additional difference between the representations was that the spaghetti plot showed absolute minimum and maximum values whereas the fan and box plots showed the 10th and 90th percentiles of values, as indicated in the keys on the plots. It is possible that participants did not adequately understand the fan and box plots or read the figure key and subsequently reported the same amount of uncertainty as they saw on the plot without adding some form of buffer to account for the 0-10 and 90-100 percentiles. Thompson *et al.* (2015) found that putting different information in the key did impact users' interpretation of volcanic hazard maps and graphs, contrary to what was hypothesised herein. Further research is necessary to determine how the interpretation of the key affected the interpretation of uncertainty for these forecast representations.

Participants consistently reported less uncertainty for the fan plot than the box plot, even though they were showing the same range of percentiles. This suggests that users of the information may interpret different amounts of uncertainty than is shown for different graphical forecast representations explicitly showing uncertainty. This is similar to findings comparing verbal modes of communicating uncertainty: different representations of the same uncertainty can be interpreted as more or less risky (e.g., Henrich *et al.*, 2015).

Participants also consistently inferred less uncertainty for the worded probability forecast representation than other forecast representations. The worded probability was deliberately vague in this experiment, used as the control condition. It is unsurprising that the amount of uncertainty inferred for the worded probability was different than the other forecast representations because the worded probability forecast gave no indication of an actual predicted ice thickness.

The second research question addressed in this paper was the effect of expertise on decisions and interpretations from the forecasts. Confirming our hypothesis, expertise significantly affected the value of the best-guess forecasts, which subsequently affected the decisions the participants made. Experts made more economically rational decisions than non-experts. In additions, experts inferred significantly more uncertainty than non-experts. These findings suggest that care should be taken in choosing which forecast representations are shown to non-expert audiences because they may interpret different amounts of uncertainty than expert audiences. These findings are contrary to Doyle *et al.* (2014), who found no differences in interpretations of probabilistic volcanic eruption forecasts between experts and non-experts. Differences in interpretations and decision-making between experts and non-experts could be due to the amount of experience each group has in using and making decisions with these forecast representations. However, previous research has been contradictory in what "experience" means. For example, both Roulston and Kaplan (2009) and Tak *et al.* (2015) tested numeracy as a predictor for making economically rational decisions. Tak *et al.* (2015) found numeracy affected decision making whereas Roulston and Kaplan (2009) found numeracy did not have an effect.

In real-world situations, experts and non-experts are likely receiving the same information and may interpret that information differently. The results from this paper suggested that there was no one forecast representation that encouraged more economically rational decision making, even when controlling for expertise. In other words, there was no single forecast representation that enhanced an expert or non-expert's decision. Therefore, we cannot recommend a particular forecast representation for expert or non-expert groups. It is possible that providing multiple forecast representations would help a user's understanding of a best-guess from the forecast. However, it is important to keep in mind that different forecast representations affect the amount of uncertainty inferred from the forecast (e.g., more uncertainty was inferred from the spaghetti and box plots than the fan plot and worded probability), so there is no one-size-fits-all solution.

5. Conclusions

Uncertainty information is increasingly being used in communication of natural-hazard related information. This research sought to determine if the way in which uncertainty information was presented affected users' interpretations of and decisions made using that information. The results suggest that changing the way uncertainty is shown in forecasts does affect the amount of uncertainty inferred from the forecast but does not change the decisions or central tendencies inferred from users. Results also suggest that experts and non-experts interpreted a different amount of uncertainty, different central tendency, and made different decisions based on the forecasts. These results suggest that deciding how uncertainty is represented should be based upon the type of audience receiving the information.

Acknowledgments

We thank Gabriel Kpaka for assistance collecting public survey data; Elizabeth Stephens and Natalie Harvey for input into the ice thickness scenario; Members of the Tropical Hour group, Atmospheric Blocking Workshop, Atmospheric Circulation in Regional Climate Change Workshop participating in the survey; Carol Loftus for permission to attend University of Reading public events to survey members of the public, and all survey participants. This research is funded by the Natural Environment Research Council (NERC) under the Probability, Uncertainty and Risk in the Environment (PURE) Programme (NE/J017221/1). Data created during the research reported in this article are openly available from the University of Reading Research Data Archive at http://dx.doi.org/10.17864/1947.107.

References

Abraham, S., R. Bartlett, M. Standage, A. Black, A. Charlton-Perez & R. McCloy, 2015: Do location-specific forecasts pose a new challenge for communicating uncertainty? *Met. App., 22*, 554–562, doi:10.1002/met.1487.

Ash, K. D., R. L. Schumann III & G. C. Bowser, 2014: Tornado warning trade-offs: Evaluating choices for visually communicating risk. *Weather, Climate, and Society, 6,* 104–118, doi:10.1175/WCAS–D–13–00 021.1.

Bernoulli, D., 1954: Exposition of a new theory on the measurement of risk. *Econometrica, 22,* 23–36.

Cheong, L., S. Bleisch, A. Kealy, K. Tolhurst, T. Wilkening & M. Duckham, 2016: Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty. *International Journal of Geographical Information Science, 7,* 1377–1404, doi:10.1080/13658 816.2015.1131829.

Correll, M. & M. Gleicher, 2014: Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions of Visualization and Computer Graphics, 20,* 2142–2151.

Cox, J., D. House & M. Lindell, 2013: Visualizing uncertainty in predicted hurricane tracks. *International Journal for Uncertainty Quantification, 3,* 143–156.

Doyle, E. E. H., J. McClure, D. M. Johnston & D. Paton, 2014: Communicating likelihoods and probabilities in forecasts of volcanic eruptions. *Journal of Volcanology and Geothermal Research, 272,* 1–15, doi: 10.1016/j.jvolgeores.2013.12.006.

Galesic, M., A. Kause & W. Gaissmaier, 2016: A sampling framework for uncertainty in individual environmental decisions. *Topics in Cognitive Science, 8,* 242–258, doi:10.1111/tops.12 172.

Gigerenzer, G., R. Hertwig, E. van den Broek, B. Fasolo & K. V. Katsikopoulos, 2005: "A 30% chance of rain tomorrow" How does the public understand probabilistic weather forecasts. *Risk Analysis, 25,* 623–629, doi:10.1111/j.1539–6924.2005.00 608.x.

Henrich, L., J. McClure & M. Crozier, 2015: Effects of risk framing on earthquake risk perception: Life-time frequencies enhance recognition of the risk. *International Journal of Disaster Risk Reduction*, *13*, 145–150, doi: 10.1016/j.ijdrr.2015.05.003.

Morss, R. E., J. K. Lazo & J. L. Demuth, 2010: Examining the use of weather forecasts in decision scenarios: Results from a US survey with implications for uncertainty communication. *Met. App.*, *17*, 149–162, doi:10.1002/met.196.

Mulder, K. J., M. Lickiss, N. Harvey, A. Black, A. Charlton-Perez, H. Dacre & R. McCloy, 2017: Visualizing volcanic ash forecasts: Scientist and stakeholder decisions using different graphical representations and conflicting forecasts. *Wea. Climate. Soc., 9,* 333–348, Doi: 10.1175/WCAS–D–16–0062.1.

Nadav-Greenberg, L. & S. L. Joslyn, 2009: Uncertainty forecasts improve decision making among nonexperts. *J. Cog. Eng. and Dec. Making, 3,* 209–227, doi: 0.1518/155534 309X474460.

Ramos, M. H., van Andel, S. J. & F. Pappenberger, 2013: Do probabilistic forecasts lead to better decisions? *Hydrology and Earth System Sciences, 17*, 2219-2232, doi: 10.5194/hess-17-2219-2013.

Roulston, M. S. & T. R. Kaplan, 2009: A laboratory-based study of understanding of uncertainty in 5-day site-specific temperature forecasts. *Met. App.*, *16*, 237–244, doi:10.1002/met.113.

Ruginski, I. T. & Coauthors, 2015: Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition and Computation, 16*, 154–172.

Savelli, S. & S. Joslyn, 2013: The advantages of predictive interval forecasts for non-expert users and the impact of visualizations. *Applied* Cognit. Psychol., 27, 527–541, doi: 10.1002/acp.2932.

Tak, S., A. Toet & van Erp J., 2015: Public understanding of visual representations of uncertainty in temperature forecasts. *J. Cog. Eng. and Dec. Making, 9*, 241–262, doi:10.1177/1555343415591 275.

Thompson, M. A., J. M. Lindsay & J. C. Gaillard, 2015: The influence of probabilistic volcanic hazard map properties on hazard communication. *J. Applied Volcan., 4*, 1–24, doi:10.1186/s13 617–015–0023–0.

Tversky, A. & D. Kahneman, 1974: Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Captions

FIGURE 1: The four forecast representations used in this analysis: (a) percent probability, (b) spaghetti plot, (c) fan plot, and (d) box plot. All forecasts represent the same information: three of ten forecasts show ice greater than 1-meter thick. The same plots were produced for 50% and 70% chance of ice greater than 1-meter thick (not shown). The dotted line in each representation shows 1-meter ice thickness, the threshold based on which participants had to make decisions.

FIGURE 2: Number of economically rational decisions per forecast representation (maximum 3). Responses of experts (academics and forecasters) and non-experts (stakeholders and public) are separated. Results are shown for the worded probability, spaghetti, fan, and box plots. The top and bottom of the whiskers represent the 90th and 10th percentiles, respectively. The top and bottom of the box represent the 75th and 25th percentiles, respectively. The bar in the box represents the median. The star represents the mean. Circles on either side of the whiskers are outliers.

FIGURE 3: Participants' best-guess forecast in meters for the (a) 30%, (b) 50%, and (c) 70% forecast. Responses of experts (academics and forecasters) and non-experts (stakeholders and public) are separated. Results are shown for the worded probability, spaghetti, fan, and box plots. The top and bottom of the whiskers represent the 90th and 10th percentiles, respectively. The top and bottom of the box represent the 75th and 25th percentiles, respectively. The bar in the box represents the median. The star represents the mean. Circles on either side of the whiskers are outliers.

FIGURE 4: Same as in Figure 3, but for range (maximum-minimum) of ice thickness, in meters. The horizontal dashed line is the range of ice thickness shown on the graphic representations (except worded probability).

TABLE 1: Summary of ANOVA tests.

TABLE 2: Summary of t-tests, all of which were Bonferroni-corrected t-tests.

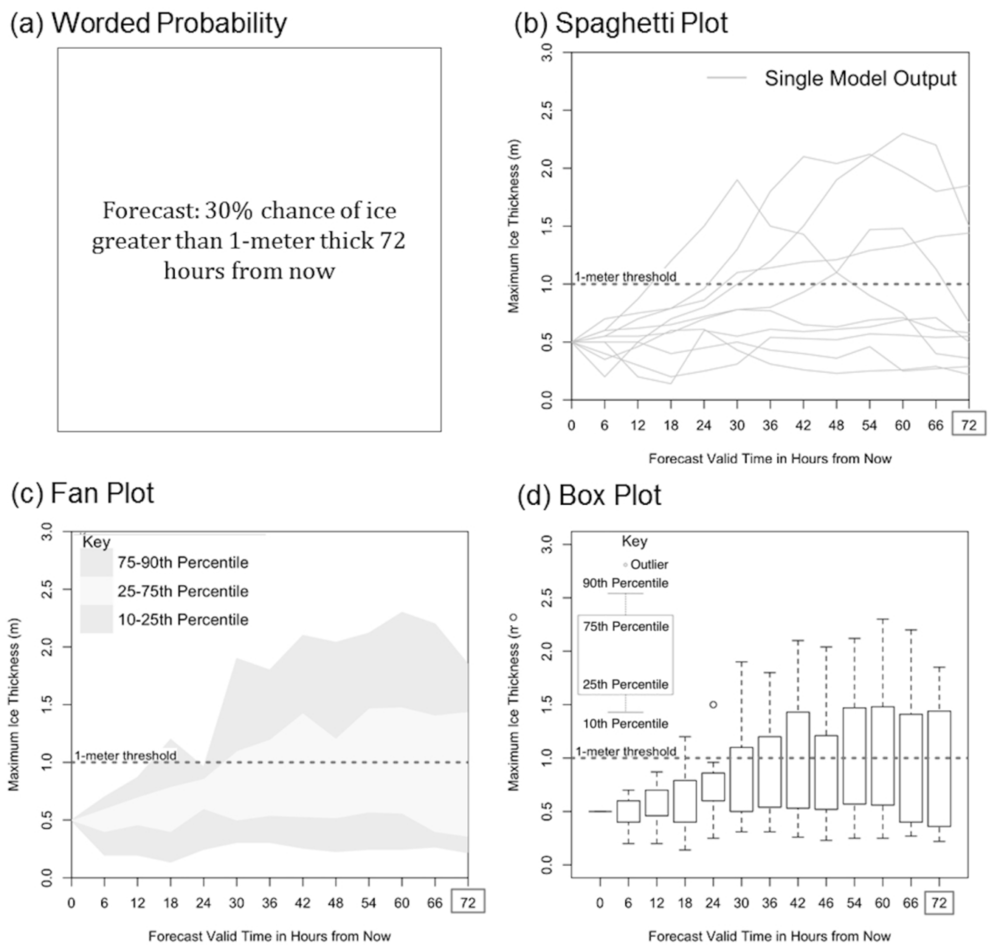

Fig. 1: The four forecast representations used in this analysis: (a) percent probability, (b) spaghetti plot, (c) fan plot, and (d) box plot. All forecasts represent the same information: three of ten forecasts show ice greater than 1-meter thick. The same plots were produced for 50% and 70% chance of ice greater than 1-meter thick (not shown). The dotted line in each representation shows 1-meter ice thickness, the threshold based on which participants had to make decisions.

254x240mm (150 x 150 DPI)

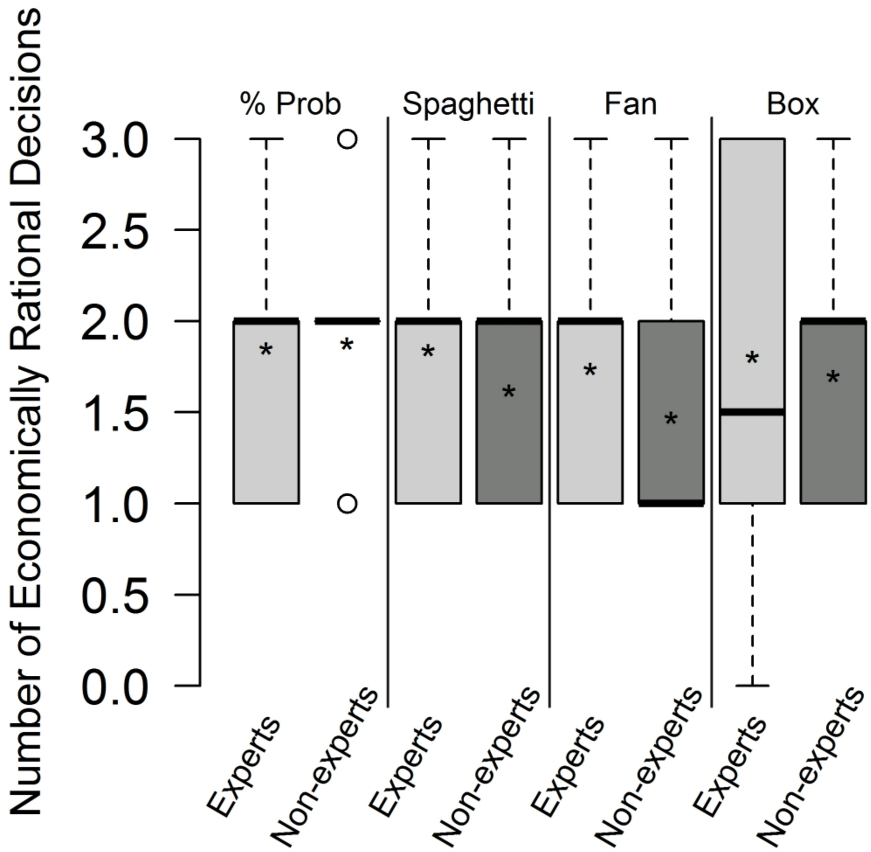

Fig. 2 Number of economically rational decisions per forecast representation (maximum 3). Responses of experts (academics and forecasters) and non-experts (stakeholders and public) are separated. Results are shown for the worded probability, spaghetti, fan, and box plots. The top and bottom of the whiskers represent the 90th and 10th percentiles, respectively. The top and bottom of the box represent the 75th and 25th percentiles, respectively. The bar in the box represents the median. The star represents the mean. Circles on either side of the whiskers are outliers.

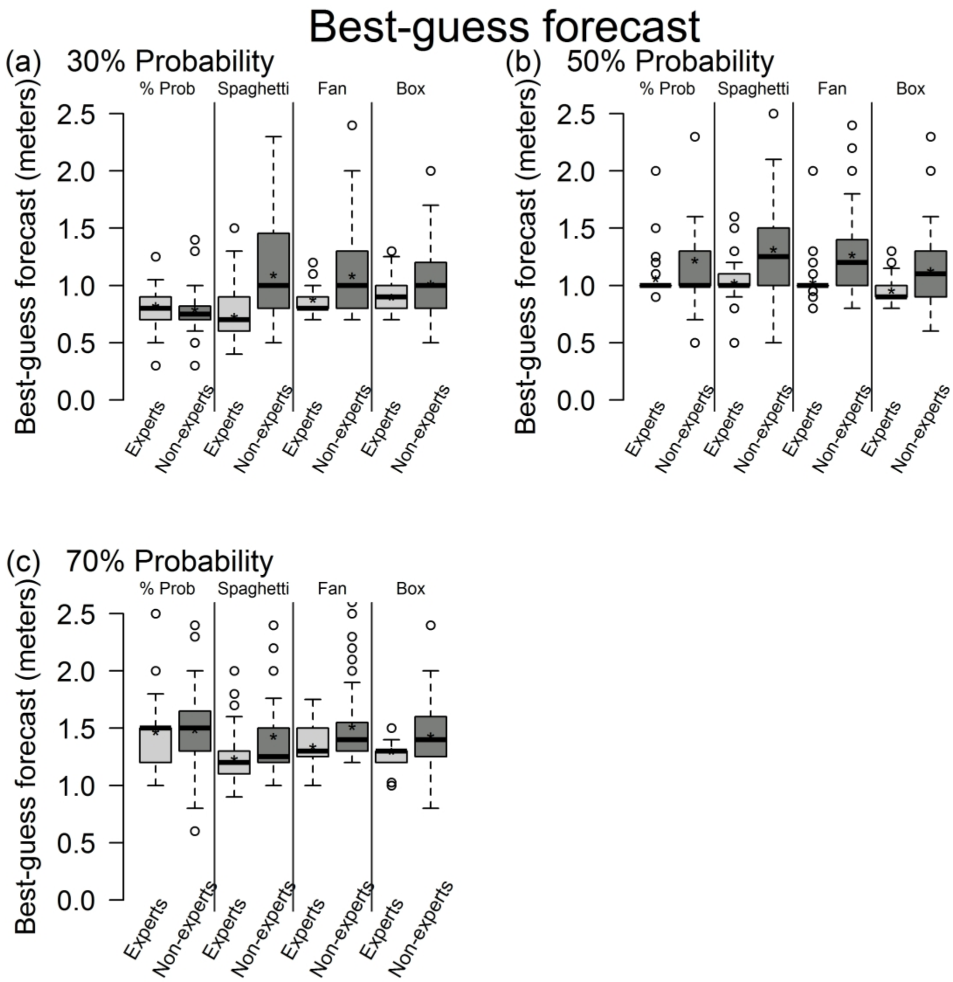

Fig. 3: Participants' best-guess forecast in meters for the (a) 30%, (b) 50%, and (c) 70% forecast. Responses of experts (academics and forecasters) and non-experts (stakeholders and public) are separated. Results are shown for the worded probability, spaghetti, fan, and box plots. The top and bottom of the whiskers represent the 90th and 10th percentiles, respectively. The top and bottom of the box represent the 75th and 25th percentiles, respectively. The bar in the box represents the median. The star represents the mean. Circles on either side of the whiskers are outliers.

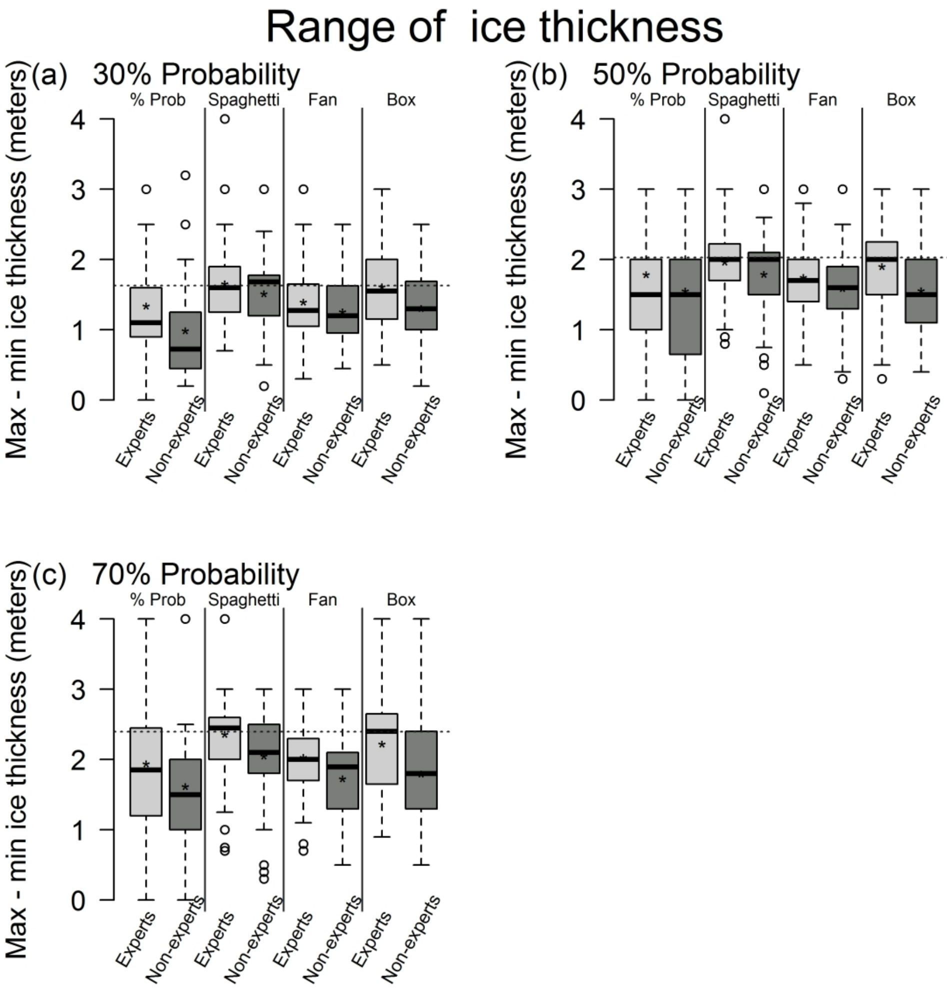

Figure 4: Same as in Figure 3, but for range (maximum–minimum) of ice thickness, in meters. The horizontal dashed line is the range of ice thickness shown on the graphic representations (except worded probability).

TABLE 1: Summary of ANOVA tests.

	F	p	η^2	df
$ANOVA_1$	0.61	0.612	0.21	3
$ANOVA_2$	5.92	0.025	0.69	1
$ANOVA_3$	0.29	0.835	0.10	3
$ANOVA_4$	4.59	0.003	0.08	3
$ANOVA_5$	7.18	<0.001	0.13	3
$ANOVA_6$	0.14	0.938	<0.01	3
$ANOVA_7$	1.24	0.266	<0.01	1
$ANOVA_8$	67.33	<0.001	0.19	1
$ANOVA_9$	10.44	<0.001	0.40	3
$ANOVA_{10}$	33.00	<0.001	0.20	1
$ANOVA_{11}$	4.59	0.003	0.08	3
$ANOVA_{12}$	7.18	<0.001	0.13	3
$ANOVA_{13}$	0.14	0.938	<0.01	3
$ANOVA_{14}$	1.24	0.266	<0.01	1
$ANOVA_{15}$	10.44	<0.001	0.18	3
$ANOVA_{16}$	67.33	<0.001	0.40	1
$ANOVA_{17}$	33.00	<0.001	0.19	1

TABLE 2: Summary of t-tests, all of which were Bonferroni-corrected t-tests

	t	p	mean	sd	df
t-test$_1$	21.66	<0.001	1.11	0.32	489
t-test$_2$	34.52	<0.001	0.91	0.31	483
t-test$_3$	2.41	0.05	1.14	0.44	323
t-test$_4$	4.24	<0.001	1.12	0.39	382
t-test$_5$	4.48	<0.001	1.11	0.28	378
t-test$_6$	0.09	1.00	-	-	326
t-test$_7$	0.66	1.00	-	-	325
t-test$_8$	0.04	1.00	-	-	383
t-test$_9$	24.38	<0.001	1.41	0.65	480
t-test$_{10}$	10.61	<0.001	1.76	0.75	470
t-test$_{11}$	6.43	<0.001	1.76	0.67	371
t-test$_{12}$	13.18	<0.001	1.64	0.59	369
t-test$_{13}$	5.57	<0.001	1.65	1.08	295
t-test$_{14}$	5.67	<0.001	-	-	377
t-test$_{15}$	2.73	0.02	-	-	304