Communicating Probability Forecasts – will people understand?

Ken Mylne

Science Fellow in Exploitation of Ensembles

OFFICIAL-SENSITIVE

Document history

Version	Purpose	Date
0.1	Draft	3/4/23
0.2	Revision	5/5/23
1.0	Final	26/6/23

Prepared by

Ken Mylne, Science Fellow, 26/6/23

Reviewed by

Helen Roberts, Socio-Meteorologist, 5/5/23

Authorised for issue by

Ken Mylne, Science Fellow, 26/6/23

Executive Summary

"People don't understand probabilities" – or do they? Weather forecasting science has long been developing ensemble forecasts as a way to improve forecast capability and provide better information to support users' decisions. The science is well proven and, indeed, the Met Office will soon move to an ensemble-only NWP (Numerical Weather Prediction) system. Ensemble forecasts can be used in a number of ways, but fundamentally they provide a probabilistic picture of the weather forecast including a most likely outcome and information on the confidence, uncertainty or risks associated with forecast outcomes. In order to pull through the full benefits of this information it is important to communicate this information effectively to users so that they can make appropriate risk-based decisions. There is a widely-held belief that people will find probabilistic information hard to understand or make use of, which provides a significant obstacle to communicating it.

This challenge for ensemble or probabilistic forecasts has long been recognised and there has been extensive research conducted into effective communication and people's understanding of such forecasts, including several papers led or sponsored by the Met Office. This paper offers a review of that research to help guide future communications of forecasts. **The overwhelming and consistent conclusion found in the literature is that people do understand the probabilistic information and make better decisions when presented with it, provided that the information is presented appropriately.**

Key conclusions include:

- Nearly all of the studies indicate that people make better decisions, have more trust in information, and/or display more understanding of forecast information when forecasters use probability information in place of deterministic statements.
- Providing additional information on uncertainty does not lead to confusion and misinterpretation compared to simple deterministic forecasts.
- The inclusion of a numerical probability (e.g. 30%) alongside a visual or worded description can greatly help with correct interpretation; using both forms helps ensure that both more and less numerate individuals will understand the message.
- Careful choice of language helps to promote understanding e.g. some people may be put off by "30% *probability*" which they consider to be mathematical, but are quite comfortable with "30% *chance*" and interpret it correctly.
- It is important clearly define the events to which probabilities apply, and the way in which forecasters frame messages can influence how audiences interpret risks.

Overall, the literature review provides strong support for communicating probabilistic information to forecast users, including the general public. It does not support the idea that people's understanding should be a barrier to communicating such information. While not every single person will understand or take full advantage of the additional information, most people will benefit and make better decisions as a result. The review also offers a number of suggestions for optimising effective communication.

1. Introduction

Uncertainty is a fundamental part of every weather forecast. The purpose of a forecast is also to help people make decisions – a forecast is useless without a decision, whether it be to change plans or go ahead as planned, to take some protective action against adverse weather or not. Hence, making decisions in the face of uncertainty is also fundamental. One of the big questions or challenges for forecasters is how much to communicate the uncertainty. Will it just confuse people and prevent them making decisions, or will it help them make better, more informed decisions?

These are not new questions, but the increasing use of ensemble and probabilistic forecasting methods in recent years has brought them to the fore. There is extensive scientific evidence that probabilistic forecasts are more skilful - have a greater information content - but widespread concerns remain about whether people can use this to make decisions. Indeed, conversations often open with the assumption that "*People don't understand probabilities",* stated as a known fact. Other concerns frequently raised include *"People just need to make (binary) decisions", "Customers want us to make their decisions for them", "Communicating uncertainty will undermine confidence in our forecasts"*. Concerns such as these have tended to make forecast providers cautious about how, or even whether, to include uncertainty information in their forecasts and services, and have led to repeated calls for more research into user needs and understanding. This is particularly true for *general* forecasts such as public weather service forecasts which are used by a wide range of people for an even wider range of decisions.

As noted, these questions have long been recognised. Alongside the development of ensemble forecasting methods, there has also been a long history of research into the communication of probabilistic forecasts and their use in decision-making. The Met Office has led or commissioned several studies in collaboration with universities. There is also a substantial body of literature from other groups around the world, as this is a challenge facing many forecasting centres. Ripberger *et al* (2022) conducted a systematic literature review on the topic of "Communicating Probability Information in Weather Forecasts", identifying 327 unique studies relevant to the topic. This paper will attempt to summarise the outcomes of all this research to answer some of the above questions and to identify the most effective ways forward to exploit the extra predictive power of ensembles to support good decision-making.

As noted, the concerns listed above may apply particularly to general, public weather forecast services and this will be the focus of this review. There have been good examples of probabilistic forecast products and services being carefully developed to help specific users take planned business decisions (eg Steele *et al, 2021)* but that is not in scope here. Similarly, there is a lot of focus in WMO programmes and the Met Office's WCSSP (Weather and Climate Science for Services Partnership) projects, for example, on using a co-development approach to build a risk-based approach to warning services. Warning services are typically developed in collaboration with decision-makers such as civil contingency planners to consider the entire *Value Chain* (Golding, 2022). These can have a wide public impact, but the key decisions they are designed to support are pre-defined during the co-development process and taken by the civil contingency services. The focus in this review is on whether a wide range of people – the public – can understand and make use of probabilistic forecasts without such advanced pre-planning. Is the assumption that "*People do not understand probabilities"* actually correct, or do professionals do the public a disservice in assuming they do not understand? Do we need further research or is the evidence already there?

This paper will review the extensive literature on the topic of public understanding of probabilistic forecasts with a view to guiding future developments of Met Office PWS forecast services. It is presented as a relatively brief summary of the main, consistent conclusions of research. More details, including examples, are presented in the Annex for those readers wishing to delve into more details. A list of references is also provided, but it should be noted that this list is far from comprehensive and many more references can be found within other reviews (e.g. Ripberger *et al,* 2022).

1.1 Language – Probabilities, Ensembles, Uncertainty…

An immediate challenge in this topic is the use of appropriate language and the avoidance of jargon, including within this paper! There is good evidence that many people may be scared off by words which they consider to be "mathematical" such as *probability;* alternatives such as *chance* or *likelihood* may be less off-putting. Similarly, *ensembles* are a scientific method or system that we use to generate forecasts, but the public user does not need to know about them. Hence, there are good arguments for avoiding use of such words in our public presentations. Some people are concerned that an organization like the Met Office should not talk about *Uncertainty* as it might undermine trust in our message, suggesting that it is better to talk about *Confidence* or *Range* as a more positive message. This is another example of appropriate use of language. However, for the purposes of this paper, we will need to make regular reference to forecasts which in some way describe and communicate uncertainty (or confidence), and which may express the probability (or likelihood or chance) of an event happening, because that is the whole topic of the paper. Usually these forecasts will be based on an ensemble system, but they may have been further processed, for example though a calibration system such as IMPROVER. Conversely, ensemble forecasts may be communicated in many ways. Probabilities provide one well-established way, but ensembles also offer the presentation of alternative scenarios, or "storylines", of what could happen, or allow for a reasonable set of bounds to be put on a forecast value. A range of different "Use Cases" for ensemble forecasts are developed by Walters *et al* (2023). For the purposes of this paper, we shall use the term *Probabilistic* to describe all such forecasts, simply meaning that they include information which expresses some of the uncertainty information. This does NOT mean that we need to, or should, use that term in presenting the forecasts to users – it is just used within this paper to reference the forecasts we are discussing. We will come back to appropriate language to use externally at various points within the paper.

2. Available Literature Reviews

Research on the topic of public understanding of probabilistic forecasts goes back a long way (e.g. Murphy *et al* , 1980) and is of interest and relevance to many forecast providers globally, hence the stimulation of a large literature. Conducting a full review of this literature would be a hugely time-consuming undertaking. To address this challenge, Ripberger *et al* (2022) adopted a systematic approach using electronic search methods, previous literature reviews and citation chains to identify 327 unique studies which they then reviewed for their quality and relevance. From this they provide a summary of key conclusions and also an online tool providing access to their validity assessments of numerous conclusions drawn by these many papers (https://crcm.shinyapps.io/probcom/#section-core-findings). This review goes far beyond what I could have achieved so I draw heavily on their conclusions.

2.1 Ripberger *et al* (2022) – Communicating Probability Information in Weather Forecasts

As Ripberger *et al* used a systematic approach, the relevance of their work to the current review depends strongly on the precise terms of their search. To quote, their "… review focuses on research studies that directly examine the impact of probability information on protective action decision-making, intentions, and behaviours. Most of the studies in the review focus on the "best" or most effective way to communicate probability information. They address questions like: are people more likely to take protective action when probability information is given verbally or numerically?" This is therefore highly relevant and well-aligned to the requirements of this current review.

Jumping straight to the final conclusions of the paper, Ripberger *et al* state **"…the research strongly suggests that 1) average people can make sense of and use probability information if consideration is given to information presentation and 2) assuming appropriate presentation, probability information generally improves decision quality."** They note that a group of studies indicate that most people intuitively infer uncertainty even when given a deterministic forecast and that findings suggest that people think about forecast events in probabilistic terms even when not explicitly told to do so. Their conclusions beg questions about how to present information, so we shall here draw out first some of their more detailed conclusions, and also some of their key recommendations on presentations. The following bullets are essentially direct quotes from the paper, marginally abridged in a few places. *[For clarity, in this section comments in italics and square brackets are my own commentary.]* To indicate the level of support for these statements, figures in brackets indicate the number of references quoted to back up these statements – please refer to the Ripberger *et al* paper for the full lists of these references.

Conclusions and recommendations for effective presentation:

- Some forecasters express a desire to "boil down" complex probability information to a deterministic point forecast for fear of confusing members of the public. Strong evidence in the research literature indicates that these fears are unfounded. Nearly all of the studies we review indicate that people make better decisions, have more trust in information, and/or display more understanding of forecast information when forecasters use probability information in place of deterministic statement. (15 references)

- Experts and nonexperts routinely use verbal probability expressions like “unlikely” or “a good chance” to indicate uncertainty; this practice is particularly common in the weather domain. The first core finding in this area of the review is very simple: there is strong evidence that risk communicators should always include a numeric “translation” for any verbal probability expressions used, and that translation should appear directly in or next to the verbal expression itself (7 references). A verbal expression like “severe thunderstorms are possible this evening” would be more effectively rephrased as “severe thunderstorms are possible (20% chance) this evening”.
- Research strongly indicates that forecasters should avoid vague verbal probability terms (such as “it is possible” or “there is a chance”), as they can be particularly problematic in communication due to variable interpretation (3 references).
- Members of the public demonstrate a basic understanding of probabilistic forecasts; however, uncertainty is best communicated through combined use of numeric and verbal expressions to meet the needs of heterogeneous audiences. Translations are important because less numerate people tend to focus on narrative evidence when evaluating risk communications (the context, their perceptions about the likelihood of comparable events, etc.), while more numerate people tend to focus on the numeric probability of the risk (5 references).
- It is really important to explain the *events* the forecast refers to in an intuitive and clear way. Murphy *et al.* (1980) say “people do not have trouble understanding what ‘30% chance’ means, but . . . they do have trouble understanding exactly what the probability refers to in this kind of forecast”. However, Juanchich and Sirota (2019) argue that previous studies use a cumbersome “correct” answer, and that “X% of simulations predict rain in the forecast area” is a more “fluent” and more easily understood response category; when using this “correct” answer, the vast majority are able to give the correct PoP (*Probability of Precipitation*) interpretation.
- Further to this, findings underscore the need to clarify what forecast periods and confidence intervals mean in the context of a given forecast rather than assuming that will be clear to the audience.
- Evidence on preference between giving numerical likelihood as percentages (e.g. 30% chance) or as frequencies (a 3 in 10 chance) is equivocal, with significant numbers of studies preferring either format. *[This is interesting since experts in public understanding of statistics, such as David Spiegelhalter or Gerd Gigerenzer have often recommended “natural frequencies” as most effective, particularly in communicating risks associated with medical treatments (e.g. Gigerenzer, 2002).*]
- Visualizations often provide an effective way to communicate probability information to people who have a difficult time with words and numbers (4 references). Some probability visualizations are common in the weather domain, but many are not; icon arrays (also known as pictographs), … for example, are common in medicine and epidemiology, but rare in weather risk communication. Icon arrays, graphics that use icons in grids to represent “at risk” populations (usually in proportion to the whole population), are among the most common visualizations in modern risk communication practice and scholarship. Many studies show that they can increase risk comprehension and avoidance actions (16 references). Many scholars theorize that icon arrays are effective because they communicate risk information in ways that show exact percentages while simultaneously conveying “gist” impressions (2 references), which can be especially important because they help people with relatively low numeracy evaluate risks and make informed decisions (4 references). *[When it comes to designing presentations for use in web and app, for example, it is*

worth considering options which have not been used previously in the weather context but which have proved effective in communicating uncertainty to a wide range of people in other fields such as medicine.]

- It is difficult to generalize conclusions between studies on the use of probability information in weather visualizations, as most of them focus on different domains which are difficult to compare, e.g. hurricane tracks or severe thunderstorms. Nevertheless, there are a few common findings in this area of the literature that warrant note. First and foremost, the studies generally agree that including probability information in forecast graphics improves risk comprehension and increases protective action intentions in high-risk areas (7 references).
- Studies generally agree that ensemble/simulation representations promote risk comprehension and awareness of unlikely (but possible) outcomes, but they may distract some people from scenarios that forecasters believe are most likely (3 references).
- Scholars who focus on weather visualizations join with scholars who focus on nonweather visualizations in emphasizing the importance of including text explanations and descriptive labels in probability graphics.

While the conclusions of Ripberger *et al* (2022) are overwhelmingly positive about the benefits of communicating probability information, they also provide several notes of caution about how to go about presenting such information. These are useful to help avoid pitfalls.

- Quoting numerous references, they note that "Efforts to communicate probability information must consider the heterogeneous nature of audiences and the different contexts and biases they entail. For instance, when it comes to both members of the public and experts, individuals sometimes have difficulty accurately interpreting forecasts that include probability information due to a number of different factors, including context, motivated reasoning and numeracy. … The way in which forecasters frame messages can influence how audiences interpret risks."
- The need to clearly define the *events* to which probabilities apply has already been noted above, but is a good example of the need to clearly define the so-called "reference class" – the context of the probability.
- Many studies emphasize the importance of making probabilistic forecasts as straightforward and easy to understand as possible in order to avoid "information overload".

The way in which probabilities are interpreted can be significantly different depending on the context. It is worth bearing in mind when planning routine communications that different users will be using forecasts - and interpreting them – in different contexts depending on their circumstances on the day and the decisions that they are taking. *[On the positive side, this is also one of the strengths of providing probabilistic forecast information rather than a deterministic best-guess, that each individual user can interpret the probability in the context of their decisions, taking account of their individual vulnerabilities and risk appetite.]*

- Several studies find that higher probabilities may lead people to view a forecast as more accurate; for instance, the same forecast would likely be taken to be more accurate if it reported a 70% chance of sun rather than a 30% chance of rain.
- Closely related, the "verbal directionality" of a statement – positive statements about the probability that an event *will* happen versus negative statements about the probability that it *will not* happen can cause people to over- or under-estimate the true likelihood, respectively.
- Under the "trend effect", a moderate risk will cause more concern if it has been upgraded from a low risk than if it has been downgraded from a high risk.

- Similarly, probability can be conflated with severity whereby someone who interprets a "slight chance" of rain showers to mean a 1%–5% chance will likely interpret a "slight chance" of a hurricane to mean something closer to a 10%–15% chance. This is important for forecasters to consider when using verbal probability statements, as it may suggest different interpretations of the same words and phrases, depending on the situation. It further reinforces the benefits noted earlier of combining verbal statements with numerical ones. *[Similarly, in my experience, the interpretation of a numeric probability into verbal terms can depend on the climatological rarity of an event, or its potential severity of impact. For example, a 10% chance of a shower would likely be considered a low probability by most, but a 10% chance of a hurricane may be considered a very high chance – because the climatological base rate of hurricane occurrence is very low (probably much less than 0.1%) and also the potential impact is very severe.]*

2.2 Stephens *et al* (2019)

Another useful summary of earlier research is presented by Stephens *et al* (2019) – a paper we shall come back to – and is reproduced in the Annex (A2.2) for convenience. Key points were:

> Numerous studies have assessed how people interpret a PoP (Probability of Precipitation) forecast, considering whether the PoP reference class is understood. Reference class refers to what this is a probability of, or what is the context of this probability e.g. "10% probability" means that it will rain on 10% of occasions on which such a forecast is given (for a particular area during a particular time period). Some people incorrectly interpret this to mean that it will rain over 10% of the area or for 10% of the time.
>
> The factors which affect understanding are unclear, …but as Morss et al. (2008) concluded, it might be more important that the information can be used in a successful way than understood from a meteorological perspective. It is not clear whether people can make better decisions using PoP than without it.
>
> Evidence suggests that most people surveyed in the US find PoP forecasts important (Lazo et al., 2009; Morss et al., 2008) and that the majority (70 %) of people surveyed prefer or are willing to receive a forecast with uncertainty information (with only 7% preferring a deterministic forecast). Research also suggests that when weather forecasts are presented as deterministic, the vast majority of the US public form their own perceptions of the likely range of weather (Joslyn and Savelli, 2010; Morss et al., 2008). It therefore seems disingenuous to present forecasts in anything but a probabilistic manner; research should be carried out to ensure that weather forecast presentation is optimized to improve understanding.
>
> Morss et al. (2008), testing only non-graphical formats of presentation, found that the majority of people in a survey of the US public prefer a percentage (e.g. 10 %) or non-numerical text over relative frequency (e.g. 1 in 10) or odds (9 to 1 against). For a smaller study of students within the UK, 90% of participants liked the probability format, compared to only 33% for the relative frequency (Peachey et al., 2013).
>
> Joslyn et al. (2009) assess whether specifying the probability of no rain or including visual representations of uncertainty can improve understanding of PoP. They found that including the chance of no rain significantly lowered the number of individuals that

made reference class errors. There was also some improvement when a pie icon was added to the probability, which they suggested might subtly help to represent the chance of no rain.

While much of this research comes from the USA, there seems little reason to think that results should not be broadly applicable to the British public as well.

3. Met Office Led Research

3.1 FEELE

Met Office research in the public understanding of probabilistic forecasts started in 2006 when Mark Roulston joined the Met Office as a consultant in the Ensemble Applications team specifically to research this topic. Addressing this challenge went hand-in-hand with development of the MOGREPS ensemble, first implemented in trial mode in 2005. Roulston joined the Met Office with previous experience in calibrating and interpreting ensemble forecasts (eg. Roulston and Smith, 2003; Roulston and Smith, 2002). In his previous role at Penn State University he had started to develop techniques for assessing people's understanding of uncertainty information in weather forecasts, working with experimental economists (Roulston *et al,* 2006). At the Met Office he set up a new collaboration with Todd Kaplan at the Finance and Experimental Economics Laboratory of Exeter University (FEELE) to continue this work, leading to several publications on the ability of the public to make effective decisions based on various presentations of probabilistic forecasts (Roulston and Kaplan, 2009; Marimo *et al, 2015*; Mu *et al,* 2018).

A full summary of the above papers is provided in the Annex at A3.1, and we here summarise the key conclusions.

In Roulston's first experiment participants were given simple temperature forecasts and asked to act as a road manager and make decisions on whether to apply salt to prevent road icing. Participants were told the cost of salting the roads, and the potential losses if they did not salt and it did freeze. All participants were given the same point temperature forecasts but with different amounts of information on the forecast uncertainty. The control group were given deterministic forecasts with a basic knowledge of typical forecast errors. A second group was given the daily varying level of uncertainty as you might get from ensemble forecasts and the third group was given this plus a probability of the temperature falling below freezing. Results clearly showed that participants given the additional, variable uncertainty information made higher profits, in other words, better decisions. They also reduced the variability of their profits, indicating that they had reduced the level of risk involved in their decisions. However, those participants given the additional information about the probability of freezing did not improve their decisions over those with just the variable uncertainty information. In theory, following Mylne (2002) for example, the probabilities along with the cost-loss information provided, would have been enough to enable participants to make optimal decisions and maximise profits, but few candidates came close to achieving this. Thus, while people did benefit from the uncertainty information, they would have needed further guidance to gain the full benefit of the probabilistic forecasts.

Roulston's second experiment was similar to the first but used new presentations of temperature forecasts. Again, one group of participants were presented only with the single-value temperature forecasts (Group A) while the second group (Group B) were given uncertainty information (see figure 1 in Annex). Both groups of participants were asked the same questions and were rewarded financially when they successfully selected an outcome based on the forecast. In some questions the best option could be selected with either type of forecast, but in selected "swing" questions the extra uncertainty information could help the user to greatly increase their chance of being rewarded. On average participants in group A who received no information on uncertainty picked the most probable option on 68.5% of occasions while those in group B with the uncertainty information picked the most probable on 85.2% of the time, with most of the difference being in the swing questions. This difference between group A and group B was approximately the same when participants were grouped by academic discipline (categorised as Business & Economics, Science & Engineering or Humanities) or by gender. While Roulston and Kaplan (2009) are careful not to interpret the results of these two experiments too generally, they do conclude that "there exist contexts and formats for which the decision-making of non-specialists is improved by providing information about forecast uncertainty. Providing information about forecast uncertainty, therefore, can be a means to improve the value of forecasts to users."

Marimo *et al* (2015) followed a very similar methodology to Roulston and Kaplan (2009) but compared understanding of three forecast temperature presentations: a tabulated deterministic forecast, a tabulated forecast including also upper and lower bounds and a graphical representation of the latter which was being used experimentally on a demonstration area of the Met Office website at the time (figure 2 in Annex). Results and conclusions were very similar to those of Roulston and Kaplan (2009). Participants given the additional uncertainty information on average performed better than those without, regardless of gender or academic department. The significant difference came from the "swing" questions, indicating that the users were able to use the uncertainty information when it matters. For the non-swing questions, those receiving deterministic forecasts did perform marginally better, indicating that in some circumstances the additional information may have led to some confusion for some participants, but overall the benefits greatly outweighed this impact. The experiment also measured the time taken to make choices, and those who received the uncertainty information in graphical form took significantly less time to make decisions than those who got in tabular form. They also performed marginally less well (although not with statistical significance), but this does suggest that the graphical approach may be less costly in time taken on interpretation and understanding.

Marimo *et al* (2009) concluded with the following general recommendations to the Met Office. "The Met Office should provide uncertainty information in temperature forecasts as it improves interpretation and understanding; there is a need to constantly assess and test different presentation formats that can be used to present weather forecasts to different users. The Met Office should engage users throughout the product development process in order to ensure information is communicated effectively and to minimize misinterpretations."

Mu *et al* (2018) focused specifically on the presentation of National Severe Weather Warning (NSWWS) warnings, comparing the decisions made by people seeing (i) the full warning matrix including the "tick in the box" as presented in NSWWS warnings; (ii) the warning colour along with the warning matrix but without a tick in the box or (iii) just the warning colour. It is worth noting that the participants in (ii) and (iii) received essentially the same information, the warning colour, but by seeing the matrix those in (ii) were given a general indication of the complexity of information behind the colour. Consistent with the previous studies, participants that received information on both the likelihood and impact levels not only had higher expected profits but made decisions with higher expected payoffs, given the payoff function information provided, than participants receiving only the warning colour. More detailed analysis revealed that when participants received both warning colour and risk information (impact and likelihood level), they followed both the warning colour and risk information contained in the warning, but followed the colour more. Different participants were also given different payoff functions for the risk-based decisions they were asked to make. Without going into too much detail, participants' behaviour was shown to vary significantly according to what payoff function they received, indicating that people were able to relate the risk information in the warning to particular decisions they were being asked to make in order to minimise losses. In summary Mu *et al* (2018) concluded that giving the impact level and likelihood level helps participants make better decisions.

3.2 The Weather Game

One concern raised about the FEELE work was that the experiments sampled university students and were therefore not representative of the general population. Encouragingly the results were consistent across students of different academic disciplines, but nevertheless they did sample disproportionately more academic sections of the population. One could argue that this is a useful sampling since these are representative of the future decision-makers in society, but it does leave open the question of whether presentation of probabilistic forecasts would work for a wider demographic. In order to address this question, the Met Office initiated a collaboration with Prof. David Spiegelhalter, Winton Professor of the Public Understanding of Statistics at the University of Cambridge, to conduct a study through the use of an online game which could be circulated through social media and attract a much wider participation. Funding was awarded from NERC (Natural Environment Research Council) for a student internship which was awarded to Elisabeth Stephens (a PhD student at the University of Bristol). Stephens developed the *Weather Game* around the idea of putting the player in the shoes of an ice cream seller who has to make decisions on when and where to sell ice cream, based on forecasts of temperature and rainfall. By successfully attracting over 8000 separate players to participate, the Weather Game was able to trial several different probabilistic presentations of the forecasts, with different levels of complexity. More details of the experiments are given in the Annex A3.2 and we report here only the overall conclusions.

Importantly, the game tested not whether people thought they understood the forecasts, but what decisions they made and hence how much money the ice cream seller might make. Results successfully demonstrated:

- Providing additional information on uncertainty did not lead to confusion and misinterpretation compared to simple deterministic forecasts.
- Where clear categorical deterministic forecasts were given, many people inferred a level of uncertainty, in that their confidence was not 100%. This is consistent with previous research (Joslyn and Savelli, 2010; Morss et al., 2008) which also found that when purely deterministic forecasts are presented, the public will make their own estimate of uncertainty.
- For probability of rainfall, there was no clear overall best presentation, but participants answered better where a specific numerical value was given for probability (eg 40%) than where the probability was presented symbolically or with a High/Med/Low rating, even where the exact same information was provided symbolically. Thus, while providing a visual image of the probability is useful, it is better also backed up by the numerical value.
- People given only deterministic forecasts of rainfall (a symbol) significantly over-estimated the likelihood of rainfall, on average.
- For temperature forecasts, the most correct decisions were made by participants who received the most complex presentation of the forecast.

Overall, Stephens *et al* concluded: *We find that participants provided with the probability of precipitation on average scored better than those without it, especially those who were presented with only the "weather symbol" deterministic forecast. This demonstrates that most people provided with information on uncertainty are able to make use of this additional information. Adding a graphical presentation format alongside (a bar) did not appear to help or hinder the interpretation of the probability… In addition to improving decision-making ability, we found that providing this additional information on uncertainty alongside the deterministic forecast did not cause confusion when a decision could be made by using the deterministic information alone. Further, the results agreed with the findings of Joslyn and Savelli (2010), showing that people infer uncertainty in a deterministic weather forecast, and it therefore seems inappropriate for forecasters not to provide quantified information on uncertainty to the public.*

3.3 *Navigator* Research

The Met Office regularly surveys users as to what they want from the public weather forecasts. Common responses are that they want it to be more accurate and also very quick and simple to assimilate. One of the challenges of this sort of survey is that users generally don't know what is possible in terms of forecast improvements or what the options for presentation might be, so they may be inclined to request the same familiar presentations but more accurate. People may be unlikely to consider requesting more information on uncertainty or confidence, for example. One survey which tried to address this challenge was contracted to a company called *Navigator* (Navigator, 2014). In this study the Met Office drew up a number of proof-of-concept presentations which could be presented on the web or a mobile phone, and the survey was based around people's understanding and responses to these concepts. While some of the designs were considered too complex or difficult to understand, this approach did promote a recognition that uncertainty was inherent to the forecasts and positive responses to having more information on this as presented in some of the concepts.

Objectives of this project were defined as:

- Explore reactions to Met Office developed stimulus material displaying probabilistic information for website and smartphones...
  - is the concept of probability introduced and understood?
  - how can probability be communicated most effectively?
- Explore the relevance of probabilistic information
  - understanding of uncertainty associated with the weather forecast
  - what probabilistic information do they want; type of weather, local, regional, national, international, level of probability
  - what role would the information fulfil for them?

Research was undertaken using a first phase involving four focus groups of 7 people each, followed by a second phase with 60 in-depth individual interviews. Each phase spanning a range of age groups, genders and social demographic groups (mainly B, C1 and C2 which ranges from Managerial/Professional (B) to Skilled Manual (C2)) across Met Office users.

All participants work with the ***implicit assumption that forecasts are 'probabilistic'*** to at least some extent and had higher expectations of accuracy for shorter range forecasts. Overall, Navigator found a positive response to the idea of including probability in a forecast, noting that it can help schedule an activity (time of day, choice of day) and prompt preparation ("more or less likely depending on hassle of preparation weighed against possible downside")…with numerous examples of types of decisions given. Some reported a more subtle benefit that they "somehow feel more prepared / empowered, less sense of weather happening to them".

There was a strong recognition of the link with the impact of the weather in people's decision-making, particularly where it linked to safety decisions. This meant that the situations in which people would value information on probabilities depended strongly on their personal sensitivities or circumstances, varying from day to day (eg 10% chance of snow is far more important when driving to the airport than when commuting by train). One challenge is that providing probabilities for everything might lead to information overload, while it might be needed to meet all these different individual requirements. Users liked the idea of being able to 'hover' or 'click' to get further information.

In terms of language, the words *chance* and *risk* were considered useful, with risk best linked to a need for caution. By contrast the words *confidence* and *uncertainty* were less well understood. Using a percentage was felt to be the most succinct and easily grasped way of conveying probability, while 'high', 'medium', 'likely', etc were felt to be too open to subjectivity. "1 in 10" was considered a long- winded way of getting to 10%. Some people indicated that giving very precise probabilities, such as 73%, suggested too much precision. Overall, these conclusions suggest that numeracy understanding was not a major obstacle amongst those surveyed.

Various presentations were tested and some broad conclusions could be summarised as:

- Familiar presentations, similar to those already used on web/app, were popular to quickly assimilate information, but the addition of percent *Chance of* for elements was useful.
- Maps or new types of graphics (eg. plume graphs of temperature or stacked probabilities of precipitation) looked complicated and led to confusion.
- One presentation gave two weather symbols for each time under columns of *most likely* and *chance of* – this was popular and gave most people what they wanted.
- Probabilities for most thresholds not of particular interest, but certain ones stand out for impact (eg 32C for health risk).

Overall conclusions were summarised as:

- Weather forecasts are assumed to be probabilistic to a greater or lesser extent.
- This assumption is brought to bear when interpreting forecasts and deciding how to respond in practical terms.
- **Explicit probabilistic information *is* found helpful**. The simplest and most effective form of communication explored was ‘chance of’ and ‘%’ coupled with a weather symbol.
- However, when adding information to the forecast we should **be wary of making the overall ‘look’ more daunting or adding complexity so that the forecast’s primary role is undermined.**
- Broadly speaking rain, snow, ice and mist / fog are of widest interest, and depending on circumstances probability of 10% or more can be felt useful.

The reader is referred to the slides (Navigator, 2014) for a lot more detail on responses to specific presentation ideas and also more detailed conclusions.

4. Other Relevant papers

While the Met Office has initiated or commissioned a number of studies and experiments discussed above, it is not alone in this endeavour. The question of how to make better use of probabilistic forecasts, and how or whether people can make better decisions from them, is of intense interest to scientists and decision-makers in many countries and related fields. Ramos *et al* (2013) consider a similar experiment to those of the FEELE experiments, but in the field of hydrology and flood management. They conducted a relatively quick and small experiment within a session of the European Geophysical Union (EGU) conference, in which participants were asked to make six decisions in each of two games. The scenario was whether or not to open a control gate in a river which would protect a city from flooding, but instead would flood a basin containing farmers’ fields. The cost of compensation to the farmers was a lot less than that of flooding the city. In Game 1 all participants were given forecasts including a best estimate forecast for the river level plus a range of uncertainty and a probability of exceeding the flood level; in game 2 they were given only the best estimate for most questions, but in the last two were also given the range of uncertainty (but not the probability). The scenarios were the same in both games, as were the outcomes, but the cases came in a different order. All participants played both games. Results showed that people made distinctly different decisions between the two games, and once again they made more optimal decisions and hence more money with the uncertainty information (game 1). Without the extra information participants tended to be more risk averse, wasting money. The paper draws many more conclusions regarding things like the impact of people's previous decision and the resulting outcome. The authors note many caveats to their conclusions, about the idealised nature of the question and the small experiment. They note the fact that participants were mostly scientists or hydrological practitioners which may raise questions about the relevance of the results to the general public but nevertheless overall results are consistent with other studies. It is worth noting that the authors of this study have made their game available to be used in training activities, and this may be worth investigating for future staff training.

The Winton Centre at Cambridge University, chaired by Prof. David Spiegelhalter who also worked with us on the Weather Game, is dedicated to the effective communication of evidence in science. In 2020 they published a short Comment in Nature (Blastland *et al,* 2020) describing five rules for evidence communication. One of their key messages was to unapologetically "Disclose Uncertainties" and that part of telling the whole story is talking about what we don't know. They discuss the topic in the context of covid-19 tests and found that including the uncertainties in test results did not seem to undermine trustworthiness but did improve people's decisions in terms of whether to isolate after receiving a negative test result.

5. Other factors in Understanding

5.1 Clear Expression of Probability Forecasts

A common problem with probability forecasts, such as the Chance of Precipitation on the Met Office web or app, is misinterpretation of what the probability refers to – sometimes called the "reference class". Common misinterpretations are that a 30% chance means that rain will occur in 30% of the area covered by the forecast or for 30% of the time, or even that 30% of meteorologists think that it will rain (Gigerenzer *et al,* 2005). Some people also associate the probability with how much rain they might get – while it is frequently true that a high probability of rain is also associated with large quantities, that is not the correct interpretation of the probability. Gigerenzer *et al*'s conclusion is that it is important to clearly state the "reference class" which the probability refers to, in other words that it is important to clearly state what we are giving the probability of. There is a danger that this leads to cumbersome expression – for example they suggest that a rain forecast might say "There is a 30% probability of rain tomorrow. This percentage does not refer to how long, in what area, or how much it rains. It means that in 3 out of 10 times when meteorologists make this prediction, there will be at least a trace of rain the next day." While we would probably not wish to express every forecast like this, it does suggest some points which might be included in the explanatory information provided with forecasts.

In this context, the explanation can be simplified somewhat where forecasts are provided for specific locations, as in the Met Office web and app. This at least reduces the risk of the 30% of the area interpretation. Giving forecasts every hour also reduces the risk of the 30% of the time interpretation. Having said that, there are potential benefits in also offering probabilities for larger areas or time-spans. For example, heavy rain and lightning associated with thunderstorms may be difficult to pinpoint such that the local probabilities at any one place and time may be very low, but many users may find it more useful to know that the probability is much higher “somewhere within 10km of your location” and/or “at some time during the afternoon”. All these things can be accommodated with careful description of the reference class.

5.2 Previous Experience of Probability Forecasts

One idea is that familiarity with probability forecasts based on long-term exposure to them will improve understanding of them. Gigerenzer *et al* (2005) compared correct interpretation of rainfall probabilities between people in 5 cities where such forecasts had been available for different lengths of time (or not at all). They did find much more correct interpretation in New York where such forecasts had been available for the longest time (since 1965), but only a weak relationship with exposure time in the other 4 cities (all in Europe).

6. Conclusions

Uncertainty has always been fundamental to weather forecasts. As I understand it, the very term “forecast” was coined by Admiral Fitzroy, founder of the Met Office, to differentiate it from a “prediction” by implying a degree of uncertainty. Forecasters have always expressed uncertainty in the phraseology they use, e.g. “…patchy rain at times…” or “…gales may affect some exposed locations…”. Nevertheless, with the advent of automated forecasts driven by NWP models, we have tended to give single best-guess forecasts for most aspects of most forecasts. The development of ensemble forecasts over the past 25 years has opened the possibility of estimating the uncertainty in automated forecasts, and of estimating the probability or likelihood of various outcomes. However, concern over whether people, particularly the wider public, can understand or benefit from having this sort of additional information has led a high degree of caution about providing it. There is a widely-held assumption that “*People do not understand probabilities*” which is frequently stated as the starting point for any discussion around providing such information in public forecasts. Such discussions very frequently go on to conclude that further research is required before any changes can be made.

The purpose of this review was to examine these concerns and assumptions and to see whether they are borne out by research evidence. The first conclusion is that these questions are not new but have been the subject of a huge amount of research over many years, including many studies directly initiated and contributed to by the Met Office. Ripberger *et al* (2022) identified 327 relevant papers and while drawing heavily on their conclusions, I have also included several more. The second conclusion is that all these studies have come to overwhelmingly consistent conclusions, that on average **people do understand probabilistic forecasts** and they **make better decisions when presented with probabilistic information**. The consistency of this result, across a huge number of papers, provides very strong support for moving forward and starting to present much more probabilistic information to the public.

There are, of course, a few caveats to these statements which should be addressed. The experimental results apply “on average” and this does not mean that every single person will understand or benefit from the greater information. Indeed, there were a few cases where researchers identified that a small proportion of people may have been confused, although they noted that these were far outweighed by the benefits. Overall, it is clear that presenting the probabilistic information routinely would allow many people to make better decisions and in so doing would draw out much greater value to society and the tax-payer who funds the forecasts. Furthermore, it would seem perverse to withhold this valuable information from the many who can benefit from it on the basis that a certain number may not, especially when the evidence also suggests that most people will not be confused by the additional information.

The papers reviewed tested many different types of presentation, including tables of numbers, graphical presentations, symbols and words. While the overall conclusions were consistent that the additional information helped, regardless of presentation type, there were some recommendations on which approaches worked best. Considering the general assumption that people don’t understand probabilities, it is perhaps surprising that several studies found that providing a numerical statement of the probability led to the best decision-making. Nevertheless, there is some evidence that the use of the word “probability” may be off-putting or even scary to some people, perhaps because it is associated with mathematics, and that using the word “chance” instead, for example, may be more effective. The use of words to describe level of probability, such as “likely” or “low chance” are often proposed as a more user-friendly approach, especially to those put off by mathematical terminology, but the evidence is that these terms are interpreted very differently by different people and in different contexts, and the use of a number is much more effective in communicating a likelihood.

Related, papers reviewed here in general concluded that the use of natural frequencies (e.g. a 3 in 10 chance) were simply a more long-winded and clumsy way of saying a 30% chance. Interestingly, research in other fields such as medicine have found these approaches extremely effective, and there may still be a case for exploring some of the visual presentations used in other fields.

In the introduction, I discussed briefly the use of appropriate language – the use of “chance” in preference to “probability” is a good example. The so-called directionality of language can be important, so for example a positive message about the probability that an event will happen may convey a different message to a negative one about the probability that it will not happen.

In terms of presentations, information content seemed to be important. Stephens *et al* (2019) found that the most complex presentation of temperature forecasts that she used led to the best decisions, indicating that the benefit of more information outweighed the complexity of interpreting it. On the other hand, the *Navigator* study (Navigator, 2014) did find that map-based presentation and some of the more complex graphical presentations did cause confusion to some users. Several studies found that presenting the same information in more than one way could help to reinforce the message, or perhaps different people gain the same message better from the alternative presentations. A good example would be presenting the chance of something happening both as a graphical representation such as a slider bar and also as a numerical value. It is also extremely important to be clear about the precise meaning of a probability, the so-called reference class; one way to reinforce this is to also give the alternative, for example a 30% probability of rain also means a 70% probability of no-rain.

Finally, a regular conclusion of public satisfaction surveys is that people want to be able to assimilate a weather forecast very quickly, at a glance, on most occasions. There is therefore a good argument for keeping some of the headline presentation similar to what it is now but allowing those who want further detail to “drill down” to access additional layers of uncertainty or probabilistic content. Nevertheless, given the weight of evidence for improved decision-making, in order to gain the maximum benefits to society, this information should be made as accessible as possible, perhaps including an option to go straight to it for those who so choose.

Annex – Further details of the research summarised in the review

A2.2 Stephens *et al* (2019)
A useful summary of earlier research is presented by Stephens *et al* (2019) and is reproduced here for convenience:

> Numerous studies have assessed how people interpret a PoP forecast, considering whether the PoP reference class is understood; e.g. "10% probability" means that it will rain on 10% of occasions on which such a forecast is given for a particular area during a particular time period (Gigerenzer et al., 2005; Handmer and Proudley, 2007; Morss et al., 2008; Murphy et al., 1980). Some people incorrectly interpret this to mean that it will rain over 10% of the area or for 10% of the time. Morss et al. (2008) find a level of understanding of around 19% among the wider US population, compared to other studies finding a good level of understanding in New York (~65 %) (Gigerenzer et al., 2005), and 39% for a small sample of Oregon residents (Murphy et al., 1980). An Australian study found 79% of the public to choose the correct interpretation, although for weather forecasters (some of whom did not issue probability forecasts) there is significant ambiguity, with only 55% choosing the correct interpretation (Handmer and Proudley, 2007).
>
> The factors which affect understanding are unclear, with Gigerenzer et al. (2005) finding considerable variation between different cities (Amsterdam, Athens, Berlin, Milan, New York) that could not be attributed to an individual's length of exposure to probabilistic forecasts. This conclusion is reinforced by the ambiguity among Australian forecasters, which suggests that any confusion is not necessarily caused by lack of experience. But as Morss et al. (2008) concluded, it might be more important that the information can be used in a successful way than understood from a meteorological perspective. Accordingly, Joslyn et al. (2009) and Gigerenzer et al. (2005) find that decision-making was affected by whether the respondents could correctly assess the reference class, but it is not clear whether people can make better decisions using PoP than without it.
>
> Evidence suggests that most people surveyed in the US find PoP forecasts important (Lazo et al., 2009; Morss et al., 2008) and that the majority (70 %) of people surveyed prefer or are willing to receive a forecast with uncertainty information (with only 7% preferring a deterministic forecast). Research also suggests that when weather forecasts are presented as deterministic the vast majority of the US public form their own nondeterministic perceptions of the likely range of weather (Joslyn and Savelli, 2010; Morss et al., 2008). It therefore seems inappropriately disingenuous to present forecasts in anything but a probabilistic manner and, given the trend towards communicating PoP forecasts, research should be carried out to ensure that weather forecast presentation is optimized to improve understanding.

Considering the representation of uncertainty, Stephens *et al* similarly wrote:

> Choosing the format and the level of information content in the uncertainty information is an important decision, as a different or more detailed representation of probability could lead to better understanding or total confusion depending on the individual. Morss et al. (2008), testing only non-graphical formats of presentation, found that the majority of people in a survey of the US public (*n*=1520) prefer a percentage (e.g. 10 %) or non-numerical text over relative frequency (e.g. 1 in 10) or odds. For a smaller study of students within the UK (*n* =90) 90% of participants liked the probability format, compared to only 33% for the relative frequency (Peachey et al., 2013). However, as

noted by Morss et al. (2008), user preference does not necessarily equate with understanding. For complex problems such as communication of health statistics, research suggests that frequency is better understood than probability (e.g. Gigerenzer et al., 2007), but for weather forecasts the converse has been found to be true, even when a reference class (e.g. 9 out of 10 computer models predict that . . .) is included (Joslyn and Nichols, 2009). Joslyn and Nichols (2009) speculate that this response could be caused by the US public's long exposure to the PoP forecast, or because weather situations do not lend themselves well to presentation using the frequency approach because, unlike for health risks, they do not relate to some kind of population (e.g. 4 in 10 people at risk of heart disease).

As well as assessing the decision-making ability using a PoP forecast, it is also important to look at potential methods for improving its communication. Joslyn et al. (2009) assess whether specifying the probability of no rain or including visual representations of uncertainty (a bar and a pie icon) can improve understanding. They found that including the chance of no rain significantly lowered the number of individuals that made reference class errors. There was also some improvement when the pie icon was added to the probability, which they suggested might subtly help to represent the chance of no rain. They conclude that given the wide use of icons in the media more research and testing should be carried out on the potential for visualization as a tool for successful communication.

Tak et al. (2015) considered public understanding of seven different visual representations of uncertainty in temperature forecasts among 140 participants. All of these representations were some form of a line chart/fan chart. Participants were asked to estimate the probability of a temperature being exceeded from different visualizations, using a slider on a continuous scale. They found systematic biases in the data, with an optimistic interpretation of the weather forecast, but were not able to find a clear "best" visualization type.

A3.1 FEELE

We present here a summary of the various papers written by Mark Roulston and others, mostly at the FEELE laboratory at the University of Exeter but starting with Mark's earlier work at Penn State University.

In Roulston's first experiment, conducted with students at Penn State in the USA (Roulston *et al*, 2006), participants were given simple temperature forecasts and asked to act as a road manager and make decisions on whether to apply salt to prevent road icing. Participants were told the cost of salting the roads, and the potential losses if they did not salt and it did freeze. These losses were varied by day of the week so that the optimal decision probability based on cost-loss theory (Mylne, 2002; Richardson, 2001) was not the same every day. All participants were given the same point temperature forecasts but with different amounts of information on the forecast uncertainty. Forecast uncertainty varied from day to day as you would get from ensemble forecasts. The control group were given only the average forecast uncertainty, so they effectively had deterministic forecasts with a basic knowledge of typical forecast errors. A second group was given the daily varying level of uncertainty and the third group was given this plus a probability of the temperature falling below freezing. For many days the best decision was clear whichever forecast version was received but on a few marginal days the extra information was useful. Results clearly showed that participants given the additional, variable uncertainty information made higher profits, in other words, better decisions. They also reduced the variability of their profits, indicating that they had reduced the level of risk involved in their decisions. However, those participants given the additional information about the probability of freezing did not improve their decisions over those with just the variable uncertainty information. In theory, following Mylne (2002) for example, the probabilities along with the cost-loss information provided, should have been enough to have enabled participants to make the optimal decisions and maximise profits, but few candidates came close to achieving this. Thus, while people did benefit from the uncertainty information, they would have needed further guidance to gain the full benefit of the probabilistic forecasts.

Roulston's second experiment, conducted at FEELE, was similar to the first but used new presentations of temperature forecasts which had been selected as the most popular of 5 options in a questionnaire hosted on the Met Office website in June 2006 (figure 1, taken from Roulston and Kaplan, 2009). In the experiment, one group of participants were presented only with the single-value temperature forecast in figure 1a while the second group were shown forecasts in the form of figure 1b including uncertainty information. Two aspects of these presentations were carefully designed to aid understanding by non-specialists: (i) the inclusion of yesterday's temperatures which provides a useful reference to put the numbers in context for people who may not know what 10C feels like; (ii) the captions describing the meaning of the red and orange areas in plain English. Both groups of participants were asked the same questions, and in each case were asked to choose to receive £0.50 if one of two options occurred, as outlined in the figure caption. (The actual outcome was drawn randomly from the statistical distribution behind the forecast uncertainty, so there was an element of luck consistent with the uncertainty information presented to Group B.) As in Roulston's previous experiment, there were a few “swing questions” in which the extra information improved the chance of selecting the most likely option to pay out the £0.50. The experiment was conducted with 153 students divided into two groups receiving presentations A and B. On average participants in group A who received no information on uncertainty picked the most probable option on 68.5% of occasions while those in group B with the uncertainty information picked the most probable on 85.2% of the time, with most of the difference being in the swing questions. This difference between group A and group B was approximately the same when participants were grouped by academic discipline (categorised as Business & Economics, Science & Engineering or Humanities) or by gender. While Roulston and Kaplan (2009) are careful not to interpret the results of these two experiments too generally, they do conclude that “there exist contexts and formats for which the decision-making of non-specialists is improved by providing information about forecast uncertainty. Providing information about forecast uncertainty, therefore, can be a means to improve the value of forecasts to users.”

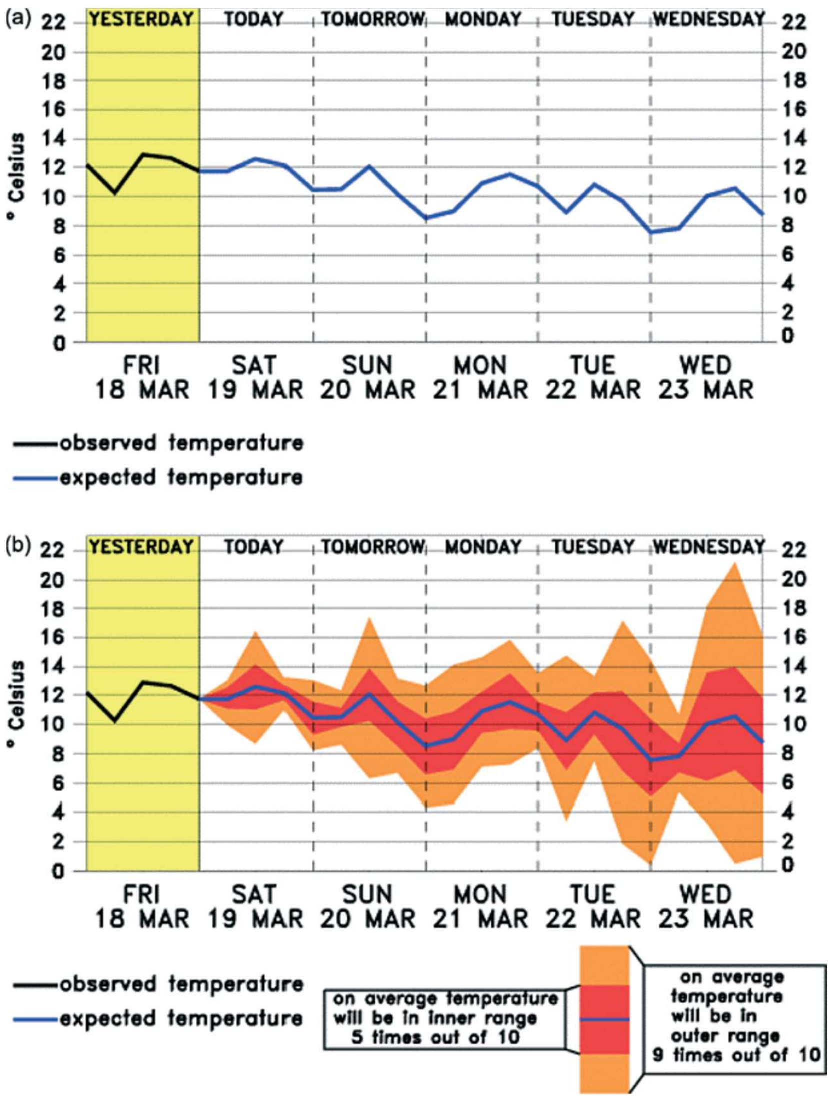

Figure 1. (a) The forecast presented to group A in round 12 of the experiment of Roulston and Kaplan (2009). The options in this round were to receive £0.50 if:

1. The temperature at midday on Monday is below 7 °C or if
2. The temperature at midday on Wednesday is below 6 °C.

(b) The forecast presented to group B in round 12 of the experiment. The options were the same as those presented to group A.

Marimo *et al* (2015) followed a very similar methodology to Roulston and Kaplan (2009) but compared understanding of three forecast temperature presentations: a tabulated deterministic forecast, a tabulated forecast including also upper and lower bounds (defined in the same way as Roulston and Kaplan, 2009) and a graphical representation of the latter (figure 2). This form of graphical presentation was being used experimentally on a supplemental test/demonstration area of the Met Office website at the time that the experiments were conducted. Results and conclusions were very similar to those of Roulston and Kaplan (2009). Participants given the additional uncertainty information (groups B and C) on average performed better than those without (group A) regardless of a)

	Maximum Temperature (°C)				
Most likely	5	6	5	1	3
	Sat 1 Jan	Sun 2 Jan	Mon 3 Jan	Tue 4 Jan	Wed 5 Jan

(b)

	Maximum Temperature (°C)				
High range	6	10	14	3	11
Most likely	5	6	5	1	3
Low range	3	2	1	-1	-1
	Sat 1 Jan	Sun 2 Jan	Mon 3 Jan	Tue 4 Jan	Wed 5 Jan

Product description
Table shows temperature ranges for a five day forecast. Temperatures fall within the indicated range roughly 9 out of 10 times with the most likely temperature in the middle. Presentation format is to help the Met Office develop and improve a product.

(c)

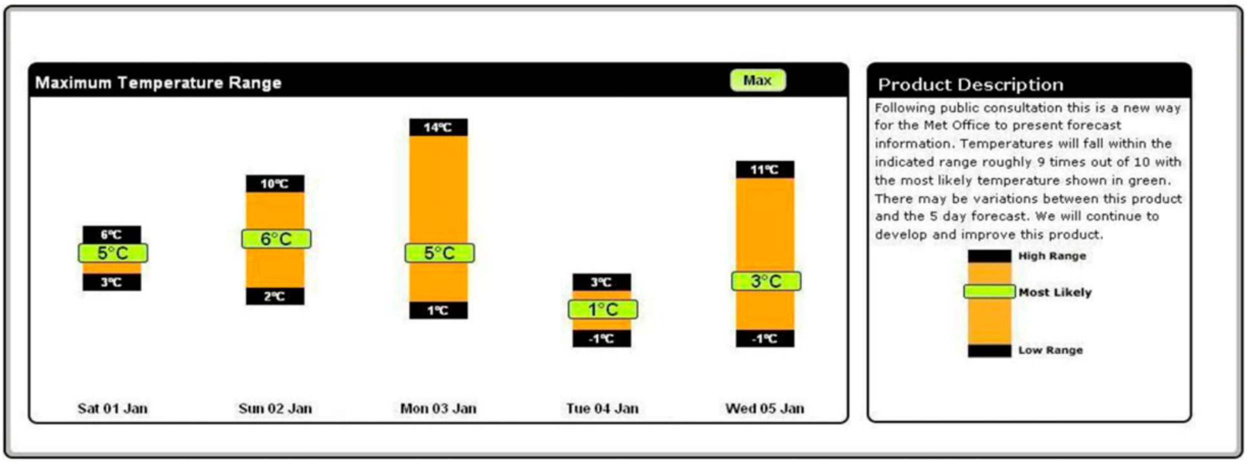

Figure 2. Forecasts presented to groups A–C in question 1 of the experiment by Marimo *et al* (2015). Participants were given the option to choose between two statements and receive 0.50 GBP. Statement A was the max temperature Monday is above 6°C. Statement B was the max temperature on Tuesday is above 0°C. (a) The forecast presented to group A in question 1 of the experiment. (b) The forecast presented to group B in question 1 of the experiment. (c) The forecast presented to group C in question 1 of the experiment.

gender or academic department. The significant difference came from the “swing” questions, indicating that the users were able to use the uncertainty information when it matters. For the non-swing questions, Group A did perform marginally better than Groups B and C, indicating that in some circumstances the additional information may have led to some confusion for some participants, but overall the benefits greatly outweighed this impact. Participants in Group C who received the graphical presentation performed marginally less well than those in Group B who received the same information in a table, although the difference was not statistically significant. The experiment also measured the time taken to make choices, and those in group C took significantly less time to make decisions than those in Group B. This may have influenced the slightly poorer performance but suggests that the graphical approach may be less costly in time taken on interpretation and understanding.

Marimo *et al* (2009) concluded with the following general recommendations to the Met Office: The Met Office should provide uncertainty information in temperature forecasts as it improves interpretation and understanding; there is a need to constantly assess and test different presentation formats that can be used to present weather forecasts to different users. The Met Office should engage users throughout the product development process in order to ensure information is communicated effectively and to minimize misinterpretations.

The final study in the series commissioned between the Met Office and the Exeter FEELE laboratory was conducted by Mu *et al* (2018). This paper focused specifically on the presentation of National Severe Weather Warning (NSWWS) warnings, comparing the decisions made by people seeing (i) the full warning matrix as presented in NSWWS warnings (Treatment 1, figure 3); (ii) the warning colour along with the warning matrix but without a tick in the box (Treatment 2) or (iii) just the warning colour (Treatment 3). It is worth noting that the participants receiving Treatments 2 and 3 received essentially the same information, the warning colour, but by seeing the matrix those receiving Treatment 2 were given a general indication of the complexity of information behind the colour.

Consistent with the previous studies, Mu *et al* (2018) found that participants that received information on both the likelihood and impact levels not only had higher expected profits but made decisions with higher expected payoffs given their information than participants receiving only the warning colour. More detailed analysis revealed that when participants received both warning colour and risk information (impact and likelihood level), they followed both the warning colour and risk information contained in the warning, but followed the colour more. Different participants were also given different payoff functions for the risk-based decisions they were asked to make, adding further richness to the analysis. Without going into too much detail, participants' behaviour was shown to vary significantly according to what payoff function they received, indicating that people were able to relate the risk information in the warning to particular decisions they were being asked to make, in order to minimise losses. In summary, Mu *et al* (2018) concluded that giving the impact level and likelihood level helps participants make better decisions. However, when participants cared more about impact than likelihood because their damages rise steeply at higher impact levels, those given only the warning colour (Treatment 3) more often chose the option that had higher expected payoffs conditional on their information. These results suggest that an effective warning system should not just have one presentation format, but should vary the format based upon the needs of those receiving it.

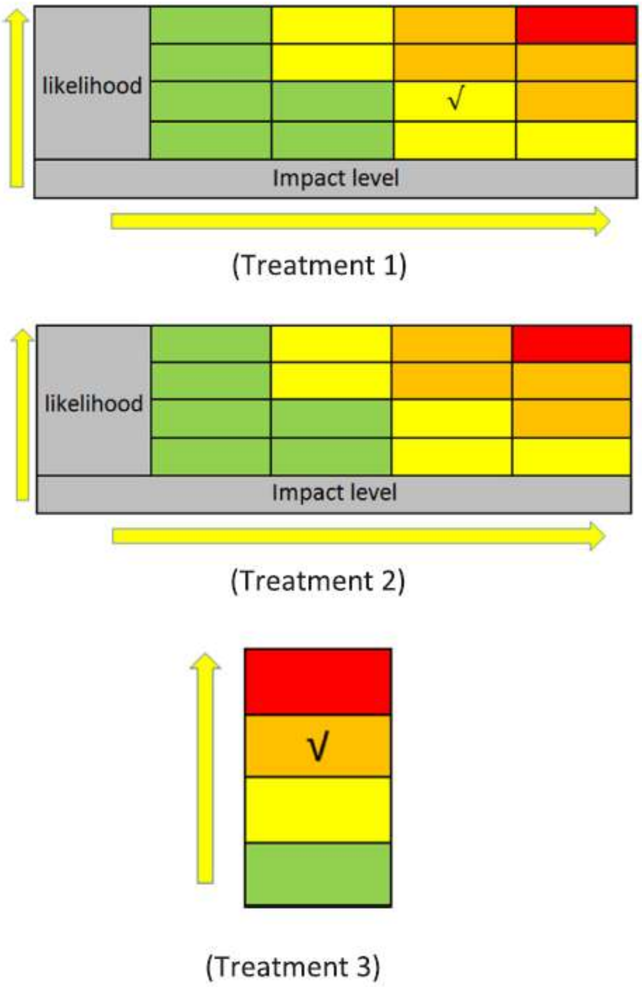

Figure 3: The form of the warnings presented in Treatments 1, 2 and 3 in the experiments of Mu *et al* (2018). In the matrix, likelihoods were 20%, 40%, 60%, and 80% (this information is not available to participants) and impact level corresponded from 1 to 4.

A3.2 The Weather Game

One concern raised about the FEELE work was that the experiments sampled university students and were therefore not representative of the general population. Encouragingly the results were consistent across students of different academic disciplines, but nevertheless they did sample disproportionately more academic sections of the population. One could argue that this is a useful sampling since these are representative of the future decision-makers in society, but it does leave open the question of whether presentation of probabilistic forecasts would work for a wider demographic. In order to address this question, the Met Office initiated a collaboration with Prof. David Spiegelhalter, Winton Professor of the Public Understanding of Statistics at the University of Cambridge, to conduct a study through the use of an online game which could be circulated through social media and attract a much wider participation. Funding was awarded from NERC (Natural Environment Research Council) for a student internship which was awarded to Elisabeth Stephens (a

PhD student at the University of Bristol). Stephens developed the *Weather Game* around the idea of putting the player in the shoes of an ice cream seller who has to make decisions on when and where to sell ice cream, based on forecasts of temperature and rainfall. By successfully attracting over 8000 separate players to participate, the Weather Game was able to trial several different probabilistic presentations of the forecasts, with different levels of complexity (figure 4).

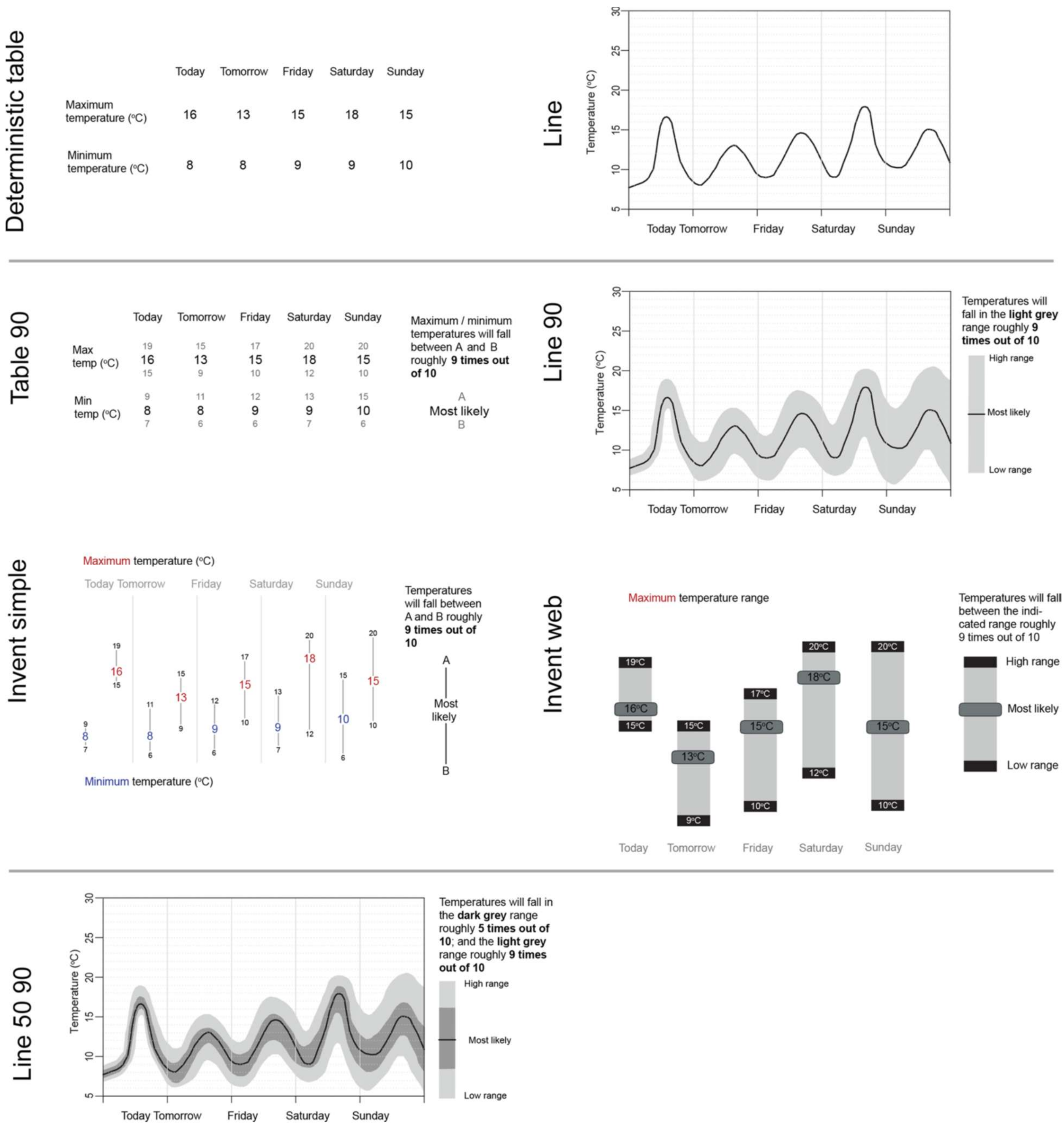

Figure 4a: Temperature forecast presentation formats used in Stephens *et al* (2019). Two different deterministic formats used for comparison (a table and a line graph); four different ways of presenting the 5th and 95th percentiles (Table 90, Line 90, Invent Simple, Invent Web; and a more complex fan chart (Line 50 90) representing the 25th and 75th percentiles as well as the 5th and 95th shown in Line 90.

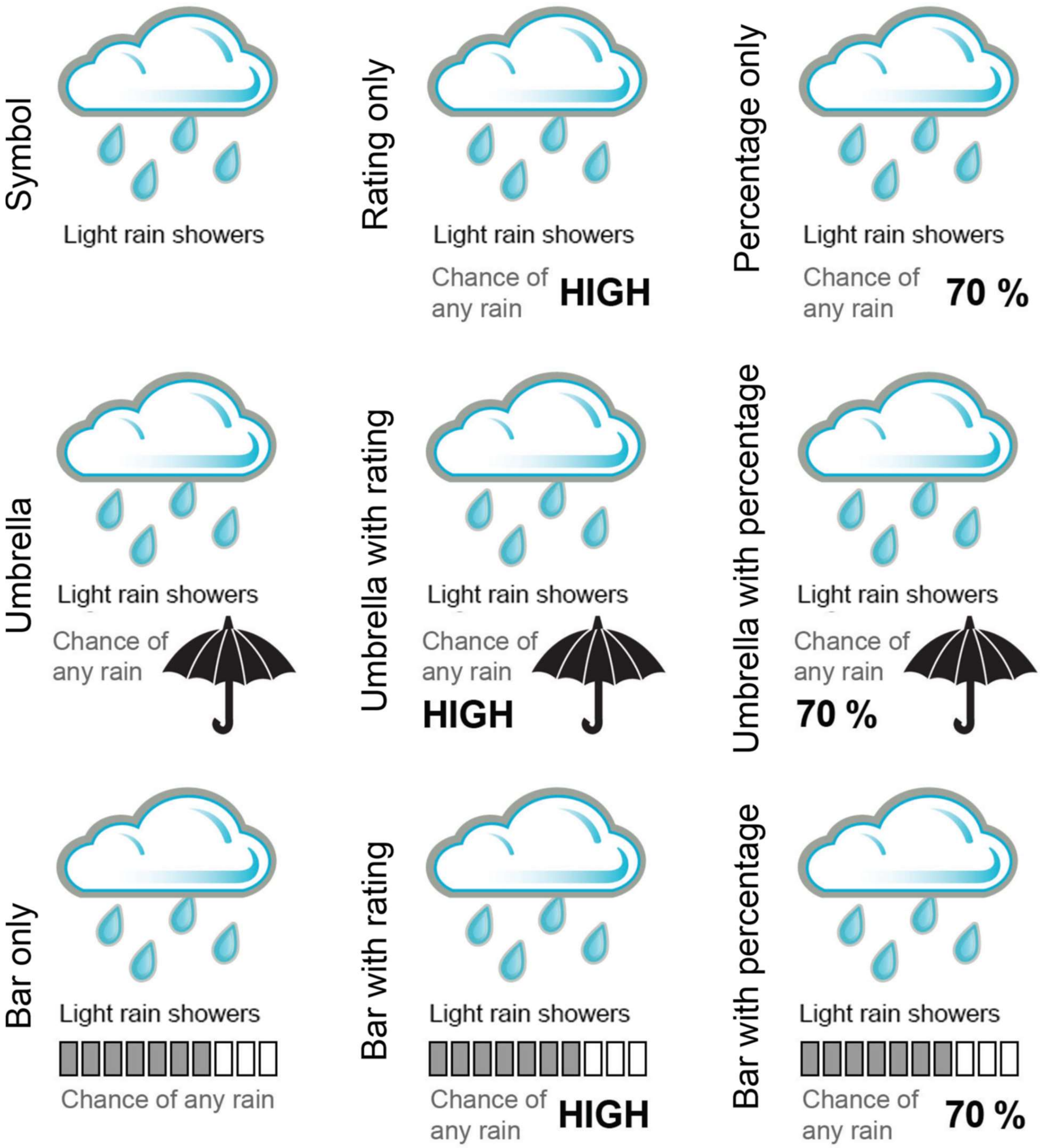

Figure 4b: Precipitation presentation formats used in Stephens *et al* (2019), with varying levels of information content. The rating is either low (0 %–20 %), medium (30 %–60 %), or high (70 %–100 %), and the percentage is to the nearest 10 %.

Importantly, the game tested not whether people thought they understood the forecasts, but what decisions they made and hence how much money the ice cream seller might make. As well as being asked to make decisions, participants were also asked how confident they were in their decisions, on a scale of 1 to 10; confidence was compared to the underlying probability. Results successfully demonstrated:

- One question for each of temperature and rainfall was designed such that the correct decision was clear from any forecast format. This was a useful easy question to introduce the decisions, but also showed that providing additional information on uncertainty did not lead to confusion and misinterpretation compared to simple deterministic forecasts – virtually all participants answered correctly.
- Where clear categorical deterministic forecasts were given, many people inferred a level of uncertainty, in that their confidence was not 100%. This is consistent with previous research (Joslyn and Savelli, 2010; Morss *et al.*, 2008) which also found that when purely deterministic forecasts are presented, the public will make their own estimate of uncertainty.
- For probability of rainfall, there was no clear overall best presentation, but participants answered better where a specific numerical value was given for probability (e.g. 40%) than where the probability was presented symbolically or with a High/Med/Low rating, even where the exact same information was provided symbolically. Thus, while providing a visual image of the probability is useful, it is better also backed up by the numerical value.
- People given only deterministic forecasts of rainfall (a symbol) significantly over-estimated the likelihood of rainfall, on average.
- For temperature forecasts, the most correct decisions were made by participants who received the most complex presentation of the forecast, the Line 50 90 fan chart in figure 4a.

Overall, Stephens *et al* concluded: *We find that participants provided with the probability of precipitation on average scored better than those without it, especially those who were presented with only the "weather symbol" deterministic forecast. This demonstrates that most people provided with information on uncertainty are able to make use of this additional information. Adding a graphical presentation format alongside (a bar) did not appear to help or hinder the interpretation of the probability… In addition to improving decision-making ability, we found that providing this additional information on uncertainty alongside the deterministic forecast did not cause confusion when a decision could be made by using the deterministic information alone. Further, the results agreed with the findings of Joslyn and Savelli (2010), showing that people infer uncertainty in a deterministic weather forecast, and it therefore seems inappropriate for forecasters not to provide quantified information on uncertainty to the public.*

References

Blastland, M., Freeman, A.L., van der Linden, S., Marteau, T.M. and Spiegelhalter, D. (2020), Nature Vol 587, 362-364.

Gigerenzer, G. 2002: Reckoning with Risk – Learning to Live with Uncertainty, Penguin Books, 2002.

Gigerenzer, G., Hertwig, R., Van Den Broek, E., Fasolo, B. and Katsikopoulos, K.V. (2005) A 30% chance of rain tomorrow: How does the public understand probabilistic weather forecasts? Risk Analysis, **25**, 623–629.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., and Woloshin, S.: Helping doctors and patients make sense of health statistics, Psychol. Sci. Publ. Int., 8, 53–96, https://doi.org/10.1111/j.1539-6053.2008.00033.x , 2007.

Golding, B (Editor) (2022): Towards the "Perfect" Weather Warning, Springer, DOI: https://doi.org/10.1007/978-3-030-98989-7, 2022.

Handmer, J. and Proudley, B.: Communicating uncertainty via probabilities: The case of weather forecasts, *Environ. Hazards-UK*, **7**, 79–87, https://doi.org/10.1016/j.envhaz.2007.05.002, 2007.

Joslyn, S. L. and Nichols, R. M.: Probability or frequency? Expressing forecast uncertainty in public weather forecasts, *Meteorol. Appl*., **16**, 309–314, https://doi.org/10.1002/met.121 , 2009.

Joslyn, S., Nadav-Greenberg, L., and Nichols, R. M.: Probability of Precipitation: Assessment and Enhancement of End-User Understanding, *B. Am. Meteorol. Soc*., **90**, 185–194, https://doi.org/10.1175/2008BAMS2509.1 , 2009.

Juanchich, M., and M. Sirota, 2019: Not as gloomy as we thought: Reassessing how the public understands probability of precipitation forecasts. J. Cognit. Psychol., 31, 116–129, https://doi.org/10.1080/20445911.2018.1553884.

Lazo, J. K., Morss, R. E., and Demuth, J. L.: 300 Billion Served: Sources, Perceptions, Uses, and Values of Weather Forecasts, B. Am. Meteorol. Soc., 90, 785–798, https://doi.org/10.1175/2008BAMS2604.1 , 2009.

Marimo P, Kaplan TR, Mylne K, Sharpe M (2015). Communication of uncertainty in temperature forecasts. *Weather and Forecasting*, *30*(1), 5-22. DOI: https://doi.org/10.1175/WAF-D-14-00016.1

Morss, R. E., Demuth, J. L., and Lazo, J. K. (2008): Communicating Uncertainty in Weather Forecasts: A Survey of the US Public, *Weather Forecast*., **23**, 974–991, https://doi.org/10.1175/2008WAF2007088.1, 2008.

Mu, Di, Todd R. Kaplan, Rutger Dankers (2018) Decision making with risk-based weather warnings, International Journal of Disaster Risk Reduction, 30, Part A, 59-73. DOI: https://doi.org/10.1016/j.ijdrr.2018.03.030.

Murphy, A. H., Lichtenstein, S., Fischhoff, B., and Winkler, R. L.: Misinterpretations of Precipitation Probability Forecasts, B. Am. Meteorol. Soc., 61, 695–701, https://doi.org/10.1175/1520-0477(1980)061<0695:MOPPF>2.0.CO;2, 1980.

Mylne, K. R., 2002: Decision-making from probability forecasts based on forecast value. *Meteor. Appl.,* **9,** 307–315.

Navigator (2014): Communicating Probabilistic Weather Information - Debrief notes. (Powerpoint presentation, available only inside the Met Office: https://metoffice-my.sharepoint.com/:p:/g/personal/ken_mylne_metoffice_gov_uk/EZe15dmevCJJhH_2Eavgh0gBClX3SzU-YRp27MAxG90phA?e=97CYKa)

Ramos, M-H., S. J. van Andel, and F. Pappenberger (2013) Do probabilistic forecasts lead to better decisions? Hydrol. Earth Syst. Sci., **17**, 2219–2232, www.hydrol-earth-syst-sci.net/17/2219/2013/ doi:10.5194/hess-17-2219-2013

Richardson, D. S., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quart. J. Roy. Meteor. Soc.,* **127,** 2473–2489.

Ripberger, J., Bell, A., Fox, A., Forney, A., Livingston, W., Gaddie, C. Silva, C., and Jenkins-Smith, H. (2022) Communicating Probability Information in Weather Forecasts: Findings and Recommendations from a Living Systematic Review of the Research Literature, *Weather, Climate and Society,* American Meteorological Society **14,** 481-498**.** DOI: 10.1175/WCAS-D-21-0034.1

Roulston, M.S. Gary E. Bolton, Andrew N. Kleit and Addison L. Sears-Collins (2006) A Laboratory Study of the Benefits of Including Uncertainty Information in Weather Forecasts, Weather and Forecasting, 21, 116-122, DOI: https://doi.org/10.1175/WAF887.1

Roulston, M.S. and T.R. Kaplan (2009) A laboratory-based study of understanding of uncertainty in 5-day site-specific temperature forecasts *Meteorol. Appl.* **16**: 237–244 DOI: 10.1002/met.113

Roulston, M.S. & L.A. Smith (2002) valuating Probabilistic Forecasts Using Information Theory, Monthly Weather Review, 130 1653–1660, DOI: 10.1175/1520-0493(2002)130%3C1653:EPFUIT%3E2.0.CO;2

Roulston, M.S. & L.A. Smith (2003) Combining dynamical and statistical ensembles, Tellus A: Dynamic Meteorology and Oceanography, 55:1, 16-30, DOI: 10.3402/tellusa.v55i1.12082

Steele E.C.C. , Brown H , Bunney C , Gill P, Mylne K, Saulter A, Standen J, Blair L, Cruickshank S, and Gulbrandsen M. 2021. "Using Metocean Forecast Verification Information to Effectively Enhance Operational Decision-Making." Paper presented at the Offshore Technology Conference, Virtual and Houston, Texas, August 2021. doi: https://doi.org/10.4043/31253-MS. (Note this publication is behind a paywall; therefore, for convenience, an internal preprint for which Met Office staff can access here: OTC-31253-MS_submit.pdf)

Stephens, E.M. Spiegelhalter, D.J. Mylne, K. Harrison, M.(2019) The Met Office Weather Game: investigating how different methods for presenting probabilistic weather forecasts influence decision-making. *Geoscience Communication*. **2** NO. 2, 2019.

Tak, S., Toet, A., and Van Erp, J. (2015): Public understanding of visual representations of uncertainty in temperature forecasts, *J. Cogn. Eng. Decis. Mak*., **9**, 241–262, https://doi.org/10.1177/1555343415591275 , 2015.

Walters, D., Barciela, R., Davies, P, Mylne, K., Petch, J., Roberts, N. and Wells, O., (2023) Exploiting Ensemble-based NWP - Classes of Use Cases for Ensemble Weather Forecasts - *in preparation.*