

Article type: Advanced Review

Communicating probabilistic information from climate model ensembles— lessons from numerical weather prediction

First author: Elisabeth M Stephens*, University of Bristol, liz.stephens@bristol.ac.uk
Second author: Tamsin L Edwards, University of Bristol
Third author: David Demeritt, King's College London

Abstract

Climate model ensembles are widely heralded for their potential to quantify uncertainties and generate probabilistic climate projections. However, such technical improvements to modelling science will do little to deliver on their ultimate promise of improving climate policymaking and adaptation unless the insights they generate can be effectively communicated to decision-makers. While some of these communicative challenges are unique to climate ensembles, others are common to hydro-meteorological modelling more generally and to the tensions arising between the imperatives for saliency, robustness and richness in risk communication. The paper reviews emerging approaches to visualising and communicating climate ensembles and compares them to the more established and thoroughly evaluated communication methods used in the numerical weather prediction domains of day to day weather forecasting (in particular probabilities of precipitation), hurricane and flood warning, and seasonal forecasting. This comparative analysis informs recommendations on best practice for climate modellers, as well as some further thoughts on key research challenges to improving the future communication of climate change uncertainties.

While there is now a high degree of confidence that the global climate is changing and will continue to do so over the next century and beyond if current development trajectories continue¹, projecting exactly what will change, when, where, and by how much is necessarily an uncertain business, which is a challenge for those charged with adapting to climate change. Robust climate policymaking

depends on climate scientists not only on improving the precision of their projections, but also effectively characterizing and communicating the associated uncertainties (and other limitations, such as ignorance and ambiguity)². To meet that need, climate modellers are increasingly adopting so-called 'ensemble' prediction (EP) techniques³⁻⁵. Rather than generating a single, 'best-guess' prediction, EP methods produce a suite of predictions, designed to represent the uncertainties associated with their forecasts. The many technical difficulties of ensemble climate modelling are well recognized, and there is an extensive scientific literature on them (reviewed by Hargreaves⁶).

However, much less attention has been given to the question we focus on in this paper: how effective is the communication of climate projections using EP techniques? Technical improvements in the science are of only limited value if the information and insights they generate cannot be communicated to inform decision-making. The huge volume and complexity of information now generated by climate models⁷ make the communication challenges particularly acute. With climate EP still a relatively new field, there are opportunities to learn from the experiences in numerical weather prediction (NWP) where there is both a longer track record of communicating EPs and an established tradition of research evaluating different methods for visualising and communicating forecast information.

Accordingly, in this paper we compare the communication of ensemble climate predictions with approaches used to communicate EPs from NWP. The paper is organised as follows. After an initial background discussion of EP and challenges of their communication, the paper reviews methods for visualising climate ensembles before turning to experiences of communicating such probabilistic information in four domains of NWP application: probabilities of precipitation (and day to day weather), hurricanes, floods, and seasonal forecasting. In the penultimate section we distil some lessons for ensemble climate prediction from the research on those NWP domains, before concluding with broader recommendations for future research and practice.

2. Ensemble Predictions and the Challenges of Communication

An ensemble is a group of model simulations designed to explore one (or more) of the four main sources of uncertainty associated with the output of a simulator. In climate and other kinds of hydrometeorological modelling, uncertainties arise from four main sources: boundary conditions, initial conditions, model structure, and model parameters (see Further Reading), and the last quarter century has seen steady growth in the development and use of ensembles to explore their relative influence in different hydro-meteorological domains¹. Boundary condition uncertainty is important in climate prediction due to the long timescales⁸, whereas initial condition uncertainty is dominant in NWP due to chaos⁹. Systematic explorations of boundary conditions with emissions scenarios have been standard for at least a decade¹⁰. Initial condition ensembles (here called ICEs) have been used in NWP since 1992¹⁰ and, to a lesser extent, in climate prediction^{1, 5}. Structural uncertainty has been explored with multi-model ensembles (MMEs) since the first global coupled climate model intercomparison project (CMIP) in the mid 1990s⁷, and parameter uncertainty has been explored with perturbed parameter ensembles (PPEs) for the last decade³. Nevertheless, there are still philosophical debates among frequentists and Bayesians about the interpretation of climate ensembles¹¹⁻¹⁶. These debates are beyond the scope of this paper, except where it affects choices about how to visualise and communicate their results.

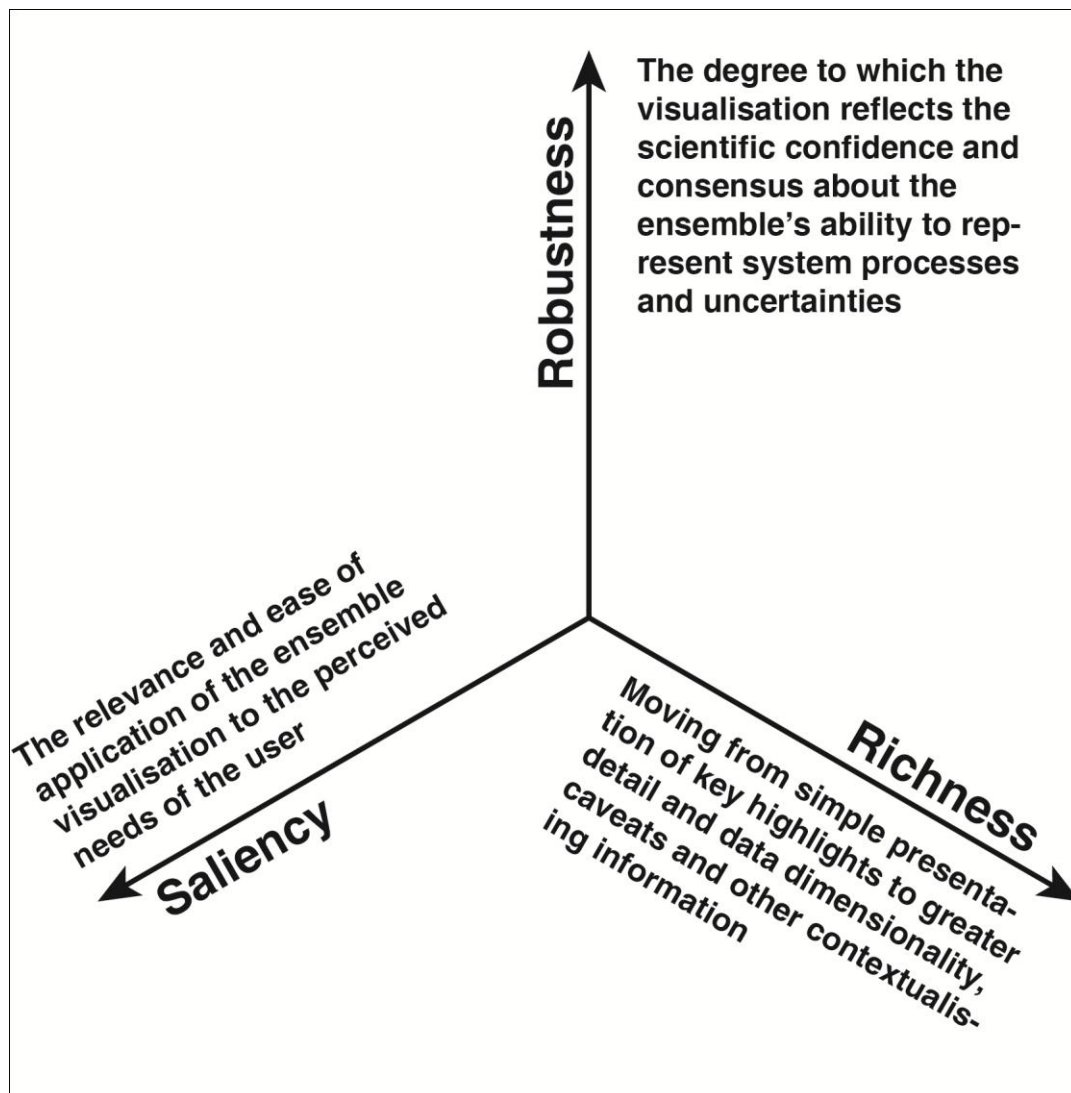


Figure 1: The three properties of communication

We argue that efforts to visualise and communicate EPs can be understood as involving three distinct properties: richness (amount of information communicated), robustness (the fidelity of the EP and the degree to which this is communicated), and saliency (interpretability and usefulness of the communication to a particular user). These may be viewed as a 3-dimensional space (Figure 1) in which the location of any given communication method depends on both design choices made and the limitations of the underlying EP. Our focus here is on the qualities of EP visualisation and communication, rather than on those of the underlying EPs themselves. But it is important to recognize that EPs are themselves representations, and might also be evaluated in terms of their richness, saliency, and robustness.

These three dimensions of communication are interlinked and often in tension. Some users may demand increases in informational richness (e.g. a full probability distribution rather than a range) that impact the ability of others to understand or use the information. Likewise concerns with robustness (e.g. limitations and ambiguities of the EP) might require reduced informational richness, given that highly contested or incomplete predictions should not be communicated with unwarranted precision. Such alterations in richness, in turn, also affect perceptions of saliency,

potentially decreasing it for users who want access to particular predictions, or increasing it for those who prefer simple, unambiguous results.

To manage these tensions, the IPCC has gone to great pains to ‘calibrate’ the language of its assessment reports. Guidance notes (most recently Mastrandrea et al.^{17, 18}) ascribe more precise definitions to expressions of qualitative ‘confidence’ (such as very high, for findings with high agreement and robust evidence) and quantitative ‘likelihood’ (such as very likely, for probability in the range 90-100%) that past research had found to be misleading or otherwise liable to multiple interpretation^{19, 20}. Our imperative of communicating ‘robustness’ broadly corresponds to this ‘confidence’, but incorporates the possibility of experts confidently communicating results that are not completely robust (discussed later). While such improvements in the clarity of the language of uncertainty are certainly welcome, these qualitative expressions are insufficient for communicating the wealth of quantitative information generated by climate ensembles. For this scientists have devised a number of visualisations, which balance the three imperatives in different ways.

3. Challenges and methods of visualising ensemble climate projections

The task of communicating ensemble climate projections involves a number of challenges that stem from what may be termed the ‘deep uncertainties’ (e.g. Kandlikar et al.,²¹) unique to this particular domain of numerical modelling. Unlike NWP, for example, future boundary conditions for climate projections are not simply uncertain, but fundamentally indeterminate insofar as they depend on future choices and behaviour. While the effects of different development pathways on future atmospheric greenhouse gas concentrations can be modelled using scenarios, those scenarios represent potential futures to which no relative probabilities can be assigned. In other words, future climate can only be projected (“what would happen if”) and not predicted (“what will happen”), as weather can be.

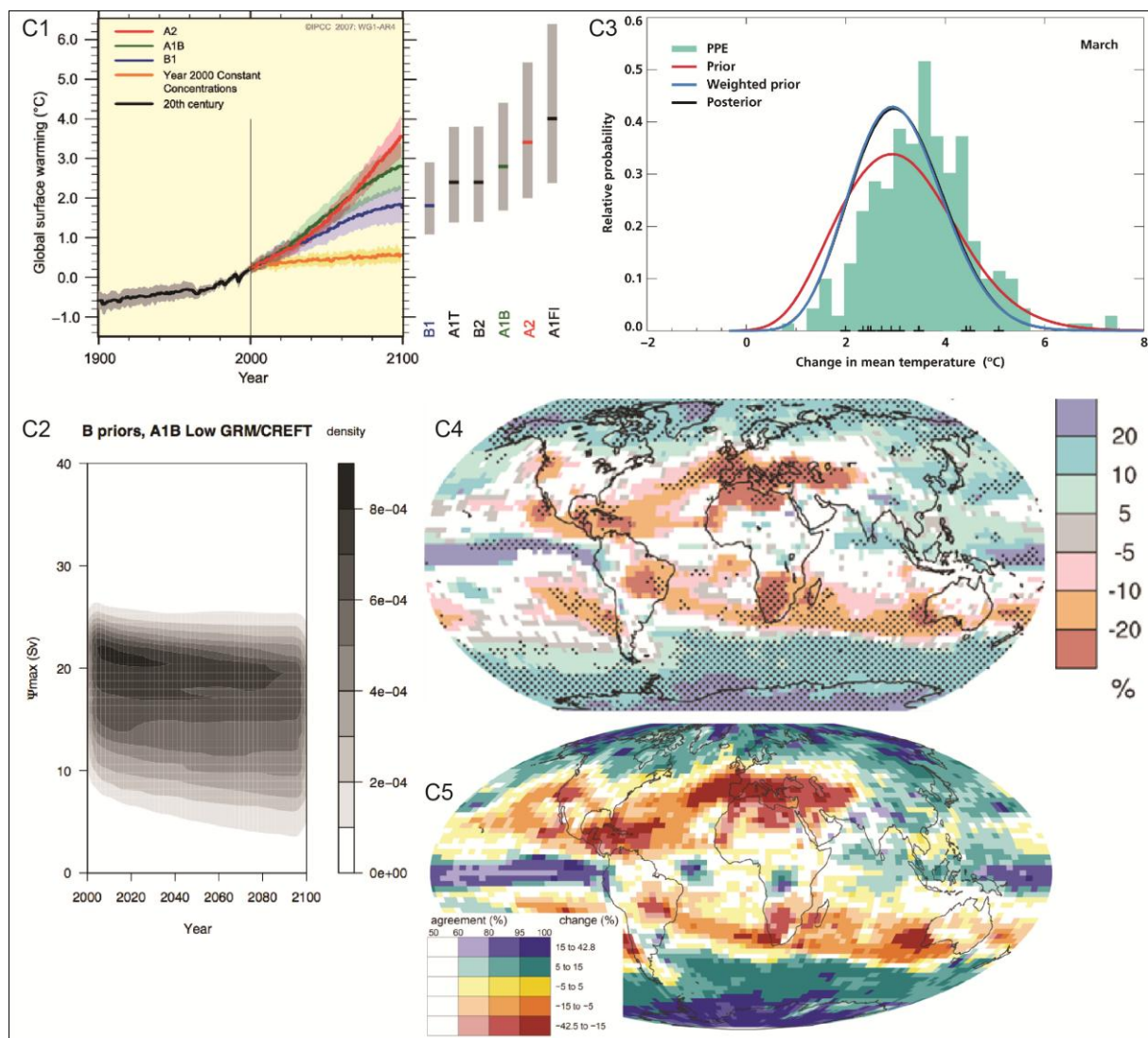


Figure 2: C1, Multi-model global means (solid lines) and ± 1 standard deviation range of individual model annual averages. (©IPCC¹). By permission of Cambridge University Press. C2, Probability density of the strength of the Meridional Overturning Circulation. (Challenor et al.²², © Oxford University Press 2010). By permission of Oxford University Press. C3, Changes in 20 year-mean surface air temperature over the HadSM3 grid box corresponding to Wales, in March, in response to doubled CO₂. (UKCP09, © UK Climate Projections, 2009). C4, Relative changes in precipitation (%) for the period 2090–2099, relative to 1980–1999 (©IPCC¹). By permission of Cambridge University Press. C5, New mapping technique illustrating change in precipitation (similarly to C4) with hues and percentage model agreement across the ensemble with saturation. (©Kaye et al.²³).

Furthermore the long temporal horizon of climate limits the number of ensemble members (due to computational expense) and, more importantly, does not allow ensemble predictions to be calibrated with repeated testing against observations (as is possible in NWP). These slow scientific progress in understanding of key climate processes such as the carbon cycle. The challenges of communicating these ‘deep’ scientific uncertainties in climate modelling are compounded by additional societal factors that complicate communication. First, climate science is heavily politicised. Special interest groups seek to advance their political cause by amplifying or dampening scientific uncertainties in line with their policy proclivities, and research has shown that those political biases also reinforce the way people seek out and credit new information about risk^{24, 25}.

Second, in the case of climate change it is not just scientific literacy that is a requirement; statistical understanding is also required to appreciate, for example, the effects of a two degree change in global mean temperature when diurnal and seasonal temperature variations are much greater. Third, the problems of long timescales and unknown future greenhouse gas concentrations increase the difficulty in communication not only to non-technical audiences¹⁸ but also experts from other fields in which model calibration and prediction are more straightforward¹⁹.

To meet those challenges, it is important to strike the right balance between our three properties of representation. First, selecting the appropriate degree of richness with which to represent very high dimensional probabilistic information on a two dimensional surface involves choices about dimensional reduction (for the outputs of interest) and level of detail (for the uncertainties explored). These choices must be made with an eye to their implications for both user saliency and the degree of robustness^{11, 17, 26}. Dimensional reduction and representation are relatively straightforward and involve the selection of variables and aggregation of spatio-temporal dimensions, although care must be taken not to disguise model inadequacy in the process: for example, plotting contours rather than 'blocky' maps could give the impression of the model resolution being greater than it is, and therefore suggests predictions are more precise. The choice of the level of detail is more complicated, as the four aspects of uncertainty (see Section 2) must also be summarised with robustness in mind. ICEs may be averaged, but emissions scenarios cannot because they represent distinct plausible futures with no relative probability assessment. Simple averaging of perturbed parameter and multi-model ensemble results is not straightforward, because it relies on good ensemble design in a well-defined space, which may not be the case for PPEs and is never the case for MMEs. The amount of information extracted from an EP can range from a full probability density function (pdf) to a histogram, an interval or percentile range, an order of magnitude estimate, a sign estimate, or a statement of complete ignorance^{17, 21}. Inevitable tensions arise between the needs of non-technical users and the risks of over-simplification of results. An example of dimensional reduction is shown in Figure 2, C1, compared with the more complete representation of time-evolving uncertainty shown in Figure 2, C2. The UK Climate Projections show a full pdf but also the original histogram on which it is based (Figure 2, C3).

Second, 'robustness' involves communicating the degree to which the EP is judged to represent reality: this judgement is based on an assessment of the type, amount, quality and consistency of evidence (including observations, models and theory), and the degree of agreement among experts (i.e. consensus) about its interpretation (e.g. expert elicitation^{27, 28}). Communication of robustness is particularly important in climate science, because ensembles cannot be calibrated in the same way as they are in NWP. Weather forecast uncertainties can be stated in terms of the average frequency of error against observations, but climate projection uncertainties must be represented in terms of expert assessment of the degree to which the model represents reality⁶. We separate robustness from confidence not only for the reasons described earlier but also because the latter is often used for the degree of ensemble agreement. For example, Figure 2, C4 shows ensemble agreement by overlaying a MME mean precipitation field with stippling in the regions where more than 90% of the model predictions have the same sign. Alternatively, a finer scale of agreement can be shown by using colour saturation (Figure 2, C5). Notably, neither gives an indication of the degree to which the EP adequately samples uncertainties. However, satisfying this robustness requirement by adding information richness will impact upon the saliency for some users. However, satisfying these robustness requirements by adding information richness may impact saliency for some users.

Third, ensemble communications should also be tailored to different users for saliency, taking into account user understanding and their requirements, as research suggests that users have sometimes struggled to see the relevance of climate ensembles for their purposes^{29, 30}. An innovative visualisation designed for a non-expert audience is the ‘migrating state’³¹, in which several northeastern US states ‘hop’ to more southern latitudes to demonstrate the potential future shift in local climate for two different emissions scenarios (Figure 3, C6), but the lack of information on robustness might frustrate an expert or sceptical audience. Another example that makes use of everyday experiences is a thermometer, with ranges of uncertainty for two different emissions scenarios (Figure 3, C7). The statistical nature of climate has prompted many attempts to communicate uncertainty in terms of betting odds: for example, interactive roulette wheels of future temperature change for two emissions scenarios (Figure 3, C8). Visualisations typically use colour to represent a dimension, but for general audiences care must be taken to consider colour vision impairment; useful guidance is given by Kaye et al.³².

There has been little research on the effectiveness of these visualisations^{33, 34}, unlike in NWP to which we now turn for comparative insight. Given the similarities between two fields, it follows there may be something to learn from these experiences.

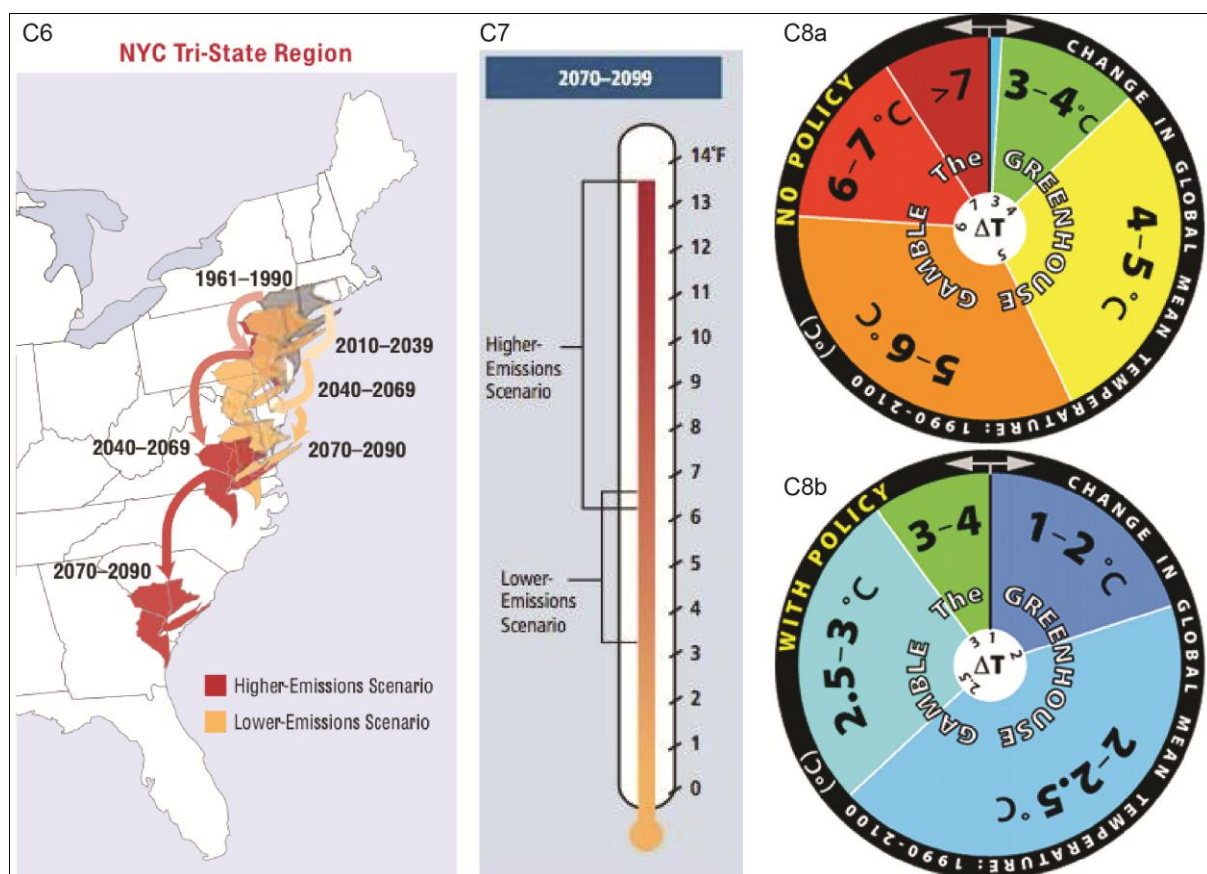


Figure 3: C6, Moving state graphic; C7, Thermometers showing projected temperature increases, both from Frumhoff et al.³¹ © 2007 Union of Concerned Scientists). By permission of Union of Concerned Scientists. C8, Roulette-style spinning wheels to depict estimated probability. Accessed 27th Feb 2012 from <<http://globalchange.mit.edu/resources/gamble>>. By permission of MIT Global Change Program.

4. Communicating ensembles from Numerical Weather Prediction

Despite the differences between NWP and climate ensembles there are obvious similarities in the communication challenges given there is considerable overlap in the underlying physical processes and the resulting uncertainties about their prediction, as well as many of the same societal pressures for greater accuracy and clarity. In this section we discuss some of the communication methods of NWP.

4.1. Site-specific weather forecasts

There is a long history in public weather forecasting of communicating probabilistic information. In the US, for example, quantitative Probability of Precipitation (PoP) forecasts have been provided by the National Weather Service since the 1960s². Probabilities were initially calculated using statistical methods, but in the 1990s NWP ensemble techniques were adopted¹¹, and ensembles are now in wide use internationally for operational weather forecasting.

While ensemble models generate information about a host of multidimensional weather properties, the chance of (any) rain is the most widely available probabilistic NWP forecast product, with likely precipitation amounts and their spatio-temporal distributions much less common. This percentage chance of rain is typically presented in terms of the PoP for a given location and represented as a number (Figure 4, W1), sometimes accompanied by a graphic (Figure 4, W2 & W3). Bar charts of PoP, such as Figure 4, W4, should be avoided so that probability is not confused with forecasts of rainfall amount. Rarely is the amount of rainfall specified (Figure 4, W5).

Because PoP is the most widely available probabilistic NWP forecast product, its communication and public understanding have been studied more than other forecast products. Much of this published literature has focussed on whether people have understood the reference class for PoP, that is, (in most cases) the probability of any rain falling somewhere in a given area over a particular period of time³⁵⁻³⁸. Misunderstandings arise where people confuse the probability with areal or temporal coverage, for example believing a 30% PoP to mean that it will rain over 30% of a specified area, or for 30% of the time. Studies have documented wide variations in the proportion of survey respondents able to identify the technically correct definition of PoP, from less than 20%³⁷ to nearly 80%³⁶. While some of this variation can be attributed to small sample sizes and differences in survey design³⁷, Gigerenzer et al.³⁵ used a consistent method to compare understanding of PoP in five cities and found significant differences in understanding that could not be attributed to an individual's length of prior exposure to probabilistic forecasts. However Morss et al.³⁷ question the significance of these artificial tests of people's ability to provide a technically correct definition of PoP. They argue that what really matters is the ability to form interpretations of PoP that, while perhaps not technically correct, still help them to make better decisions. Despite well documented problems with misinterpretation, particularly of the reference class, surveys consistently show that most Americans value PoP forecasts as salient for everyday decisions, like whether or not to bring an umbrella^{37, 39}.

The choice of format for presenting uncertainty information influences its understanding, and there is heated debate amongst risk communication experts about the merits of different approaches^{40, 41, 42}. In contrast to medical risk communication⁴³, Joslyn and Nichols⁴⁴ find that conditional probability (e.g. 10%) of PoP was easier for experimental subjects to understand than natural frequency (e.g. 1 in 10), even when a reference class was specified (e.g. it will rain in 1 out of every 10 days like this). This finding is replicated by the survey research of Morss et al.³⁷, who found a clear preference

among survey respondents for a percentage or non-numerical text rather than the communication of PoP in terms of relative frequency as often recommended for the communication of medical risks^{40, 41} (although there is still some dispute, see Woloshin and Schwartz⁴⁵). To build on the frequency versus probability debate in the communication of uncertainty, Joslyn et al.⁴⁴ used an experimental design to look at the effects of specifying the probability of no rain as well as visual representations of uncertainty (e.g. a pie icon, see Figure 4, W4) on the understanding of PoP. They found that inclusion of the chance of no rain significantly lowered the number of individuals making reference class errors. There was also some improvement when the pie icon was added to the probability, which they suggested subtly helps to represent the chance of no rain. Given the wide use of such icons in the media, Joslyn et al. called for more research on the communicative value of such icons and other visualisations of probability. With this in mind, the UK Met Office devised an online game to test ways of presenting PoP. Results suggest that decision making ability is no different between participants presented with only conditional probability, and those with conditional probability and a bar graphic (Figure 4, W2)⁴⁶.

Presentation of the temperature ensemble forecast (e.g. Figure 4, W5 & W6) has received much less attention, largely because it is less frequently provided by meteorological agencies. Laboratory studies of students in the US and UK⁴⁷⁻⁴⁹ show that people who are presented with temperature uncertainties are better able to make decisions on risk and reward than those without. In the study of Roulston et al.⁴⁸, those information about the standard error performed significantly better than those without (similarly Roulston and Kaplan⁴⁹). Joslyn and LeClerc⁴⁷ replicated these results and also found that participants provided with uncertainty information outperformed even those who were given categorical advice about the optimal course of action given the uncertainty. This experimental study provided the first empirical support for the claim often made about ensembles that people can make better decisions if given uncertainty information. It also suggested that increasing the richness of uncertainty information may increase trust as those presented with the uncertainty forecasts rated them as significantly more trustworthy than those presented with just the deterministic forecasts (although this result should be put into the context of the experimental laboratory setting).

Research on the best methods for communicating uncertainty in NWP has focussed on site-specific forecasts, where only a single probability or uncertainty distribution, perhaps with some additional temporal resolution, is communicated. However, the information that most users are presented with, at least initially, is usually some kind of synoptic weather prediction across a geographic area (for example, on a television weather forecast), but there has not been much published research to date on how best to visualize the spatial distribution of PoP or other probabilistic weather products.

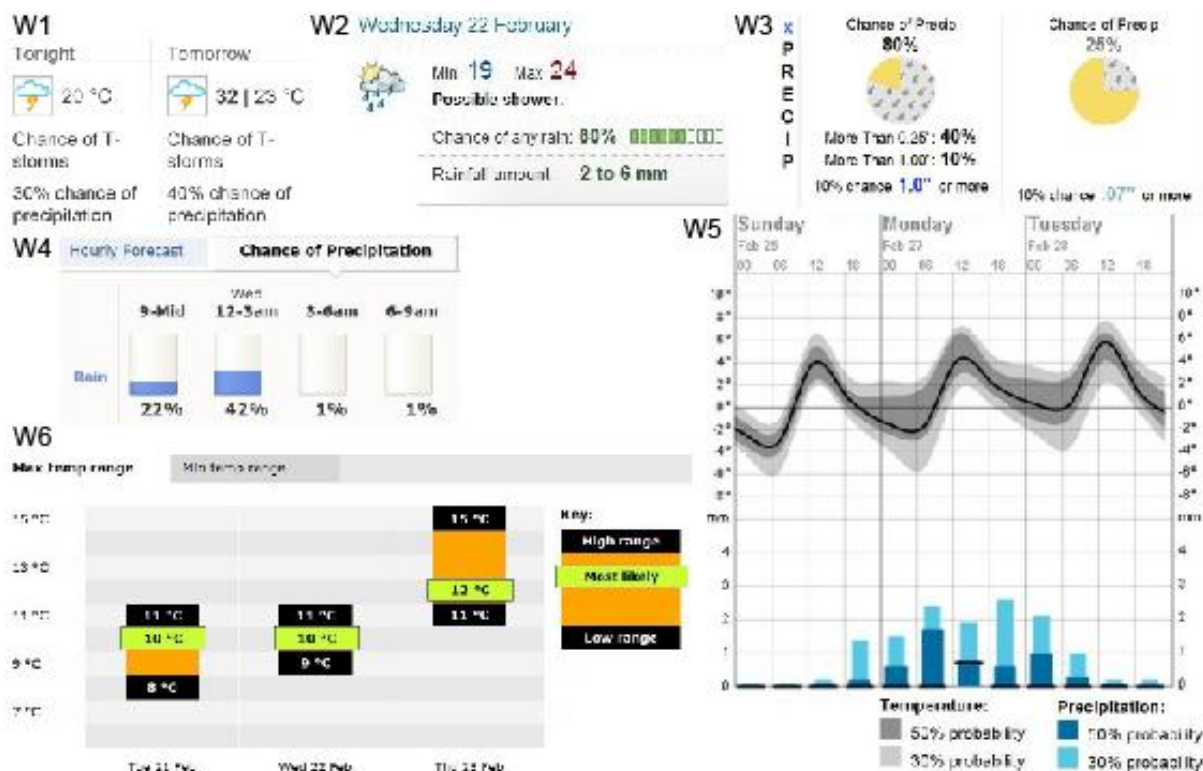


Figure 4: W1, PoP with no graphic, from wunderground.com forecast for Des Moines, IA, US. Accessed from: <http://www.wunderground.com/q/zmw:50301.1.99999>. W2: PoP with probability bar graphic for Sydney from the Bureau of Meteorology, Australia. Accessed from <http://www.bom.gov.au/nsw/forecasts/sydney.shtml>. W3: PoP with probability pie charts, University of Washington Probcast. Accessed from: <http://probcast.washington.edu/>. W4: PoP with probability bar, note how vertical bar and blue colour might cause confusion with the amount of rain. Accuweather, accessed from: <http://www.accuweather.com/en/us/new-york-ny/10017/weather-accupop/3712pc>. W5: Time series showing 50% and 80% probability range for temperature and precipitation amount, Norwegian Meteorological Institute. Accessed from: <http://www.yr.no/place/Norway/Oslo/Oslo/Oslo/long.html>. W6: Temperature range bar, showing 90% range for predicted maximum (and minimum in separate tab) temperatures. UK Met Office, accessed from: <http://www.metoffice.gov.uk/public/beta/weather/forecast/?tab=fiveDay&dayIdx=0&locId=350610>.

4.2. Hurricane Forecasts

Hurricane forecasting is perhaps the only area of EP where communication of spatial information has been carefully studied. The risk of hurricanes means that eye-catching graphics are widely adopted by the media: 57%-68% of survey respondents said the US National Hurricane Centre 'Cone of Uncertainty' (Figure 5, H1) was "very important" in their decision to evacuate⁵⁰. However, this graphic has been revised since its inception in 2000 in response to important lessons learned about its effectiveness. In particular members of the public often focused on the forecast track line and failed to appreciate both the uncertainty about it or the statistical meaning of the wider 'cone' of uncertainty about its projected course. As a result people living in the forecast cone but not near the track incorrectly consider themselves safe from harm⁵⁰. Another problem was that viewers often failed to understand that the hurricane would affect a much larger area than just the cone depicting the uncertainty about the track of the eye of the storm, whose sphere of influence was many times greater⁵¹.

Research into the Wind Speed Probabilities graphic has also raised questions about its interpretative flexibility and robustness. Some recipients interpret it as an indication of storm strength, storm extent, spatial hazard (rather than risk), or storm evolution over time, rather than the probability of a particular wind speed over a given time period⁵¹. Further, following a survey of Hurricane Ike survivors, Morss and Hayden⁵² registered doubts about the effectiveness of the underlying Saffir-Simpson storm scale metric since many people mistakenly believed it referred to storm risk and therefore did not evacuate when warned of a Category 2 storm with a large storm surge. There are clear indications that consideration should be paid to potential misinterpretation of the language used to describe hurricanes: for example, scientists use ‘growth’ to describe an increase in storm intensity, but this understandably leads to misinterpretations of storm size, and when storm size is presented as a radius it is often confused with diameter⁵¹.

Although roughly half of the participants in small ‘draw and tell’ focus groups said they would look at forecast graphics in the event of an approaching hurricane, Eosco⁵³ found that people presented with raw model forecast ‘spaghetti’ tracks (Figure 5, H2) were confused over ‘which one to believe’. Indeed, Broad et al.⁵⁰ note that the cone of uncertainty graphic was misinterpreted even in educational publications. It follows there is a need for rigorous pretesting of forecast graphics⁵⁴, as well as ongoing evaluation so problems can be identified and improvements made.

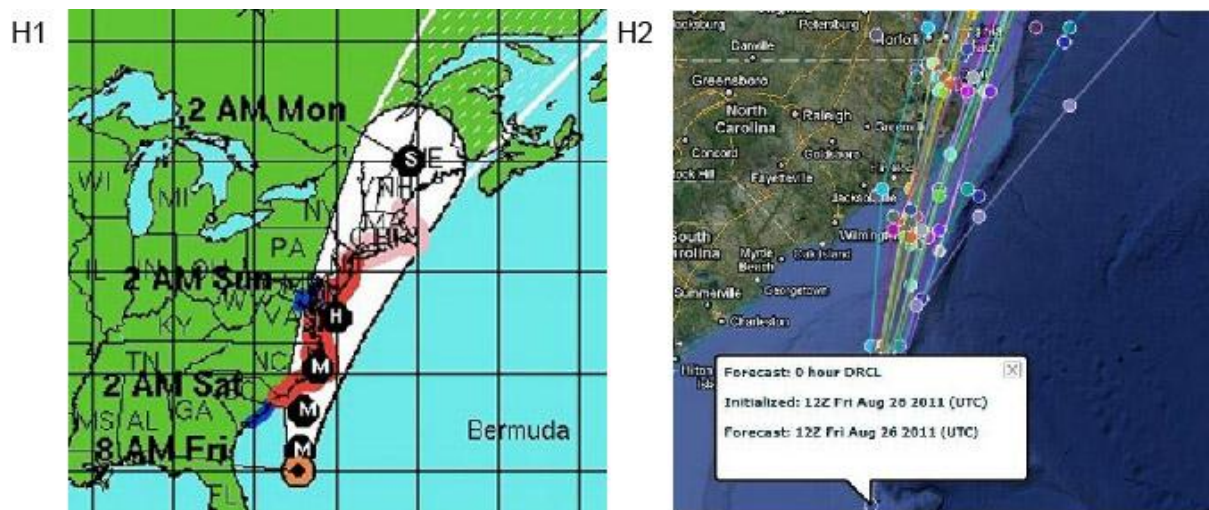


Figure 5: Figure H1: Spaghetti plot showing hurricane tracks. ABC Weather, accessed from: <http://www.wjla.com/blogs/weather/2011/08/hurricane-irene-path-projections-spaghetti-style-12544.html>. Figure H2: Hurricane Cone of Uncertainty, note that the estimated ‘best forecast track’ has now been removed to avoid confusion. National Hurricane Center (US), accessed from: <http://newsfeed.time.com/2011/08/26/hurricane-irenes-path-how-do-forecasters-predict-the-cone-of-uncertainty/>

4.3. Flooding

Building on nearly two decades of operational use in NWP, there is now growing international interest in developing coupled hydro-meteorological ensemble prediction systems (HEPS) for operational flood forecasting and warning⁵⁵. Quite apart from the technical challenges involved⁵⁶, another reason uptake has not been faster is that operational agencies are uncertain about how to communicate and use the resulting ensembles in flood incident management^{57, 58}. Several studies have documented divergent views among practicing hydrologists about the most important

information to extract from HEPS (i.e. ensemble mean, max/min values, summary statistics, hydrographic time series, etc.) and the appetite and ability of non-forecasters to make sense of it⁵⁹⁻⁶². Though it is much less remarked upon than the challenges of communicating to the lay public (but see Faulkner⁶³), communication of HEPS among hydrological experts and civil protection authorities is no less important, as intensive computational requirements mean that operational HEPS are likely to involve a central hub responsible for disseminating HEPS outputs to regional offices with responsibility for local forecasting and warning. In this context, nominally 'expert' hydrologists based in regional offices stand in the same relationship to HEPS as the lay public at large: dependent upon external information from a central HEPS that they are asked to take on trust without necessarily being able to interrogate it for themselves directly.

The European Flood Alert System (EFAS) is one of the longest running operational HEPS⁶⁴, having issued alerts on an experimental basis since 2005. Driven by the 51 member ECMWF ensemble, the EFAS models water balance at a 5km grid scale, and issues alerts to cooperating national forecasting agencies for 3-10 days ahead when critical thresholds are exceeded. EFAS alerts combine textual description of the synoptic situation with threshold exceedance maps in which pixel colours represent the level of EFAS threshold exceedance for that location and tabular information summarizing the number of ensemble members exceeding different EFAS threshold levels (Figure 6, F1), while a password protected website provides registered users with access to additional information. Independent research found warm support among EFAS users for tabular presentation of EFAS ensembles in terms of natural frequencies, which were seen as clear and easy to understand, but users also wanted greater richness and in particular conventional hydrographs, which effectively show the temporal evolution of flows (observed and projected) at a point, showing EFAS streamflow forecasts in m³/s, so as to understand the temporal evolution of the EFAS forecast and to enable comparison with their own, locally determined forecasts⁵⁹. However hydrographs were not initially provided by EFAS, partly because difficulties securing consistent European-level data for error correction and calibration meant that the EFAS hydrographs were not robust when compared with observed values, but also because of the desire to reinforce the institutional distinction between the role of EFAS as an early warning system, for which the salient information is threshold exceedance rather than precise prediction is key, and that of national agencies responsible for issuing more detailed, local scale flood forecasts⁶⁵. In response to user feedback, EFAS is now generating hydrographs for those locations where sufficient data is available⁶⁶, using a plume graph of the uncertainty in the error-corrected forecasts to supplement the tabular display of the number of threshold exceedances.

This tabular format has since been adopted to visualize ensemble flood forecasts in Switzerland⁶⁷. An evaluation by Frick and Hegg⁶⁸ found that Swiss civil protection officials valued probabilistic information from HEPS and judged their own understanding of it to have been improved through cartographic and tabular visualizations. However, uncertainty information did not lead to observable or self-reported improvements in the quality of their decisions over the course of the five month study period. Similarly Demeritt et al.⁶⁰ found reluctance among hydrologists and civil protection authorities to act on probabilistic warnings from HEPS.

One explanation for this hesitancy is that it stems from cognitive difficulties in interpreting the content of complex, information-rich HEPS forecasts. For instance, Priest et al.⁶⁹ found that in the UK emergency responders often could not understand the differences between various probabilistic

forecast products, and struggled to interpret even the very simplified form in which the Met Office's Extreme Rainfall Alerts and Flood Guidance Statement was communicated using a traffic light-based framework (Figure 6, F2); similar traffic light-based 'vigilance' maps are used by SCHAPI and Météo France (Figure 6, F3). Reflecting a more general predilection in the UK for various kinds of 'risk-based' policymaking^{70, 71}, the UK is unique in trying to incorporate a measure of potential impact as well as probability of occurrence in its flood and severe weather warnings. Other European HEPS platforms stick strictly to communicating the probability of occurrence⁶⁷. However, in a series of focus group exercises with European hydrologists about their preferences for HEPS warning formats, Pappenberger et al.⁶² found some support for the idea that flood forecasts should ideally incorporate some measure of vulnerability and impacts, along with other information about current observations and past model performance, but found little consensus on the best way to visualize HEPS. There was support for hydrographs, but a range of views about the appropriateness for different audiences of providing a full 'spaghetti' style graph of all ensemble members (e.g. Figure 6, F4) as opposed to reduced form visualisations such as the ensemble mean and the 10% and 90% confidence intervals provided in Austria on the publicly accessible website (Figure 6, F5).

While training and improved visualization might overcome these cognitive obstacles to acting on HEPS, research has also identified a set of organizational and political obstacles to doing so^{65, 72}. Probabilistic forecasting not only communicates forecast uncertainty but the very provision of that information also serves to shift responsibility for managing that uncertainty from forecasters onto forecast recipients. While this shift is sometimes welcomed as empowering local decision makers, it can also challenge the existing structures and organizational cultures for emergency planning and response, which, particularly in Napoleonic code countries, like Germany^{73, 74}, can involve highly legalistic standards for public safety and rigid response protocols based on binary distinctions between normal conditions and an exceptional state of emergency requiring extraordinary response. In this context civil protection authorities may well demand deterministic predictions issued at high degrees of certainty, and forecasters, in turn, see it as their professional duty to provide iron-clad deterministic predictions, rather than some probabilistic forecast of the likelihood of error⁶⁰.

approach, the Met Office now provides a 1-month and 3-month outlook for Civil Contingency Planners on its website (Figure 7, S1) which, in presenting the separate ensemble members against past observations, is quite open about the uncertainty and the forecast capabilities. There is also detailed user guidance available on the website, which has been accompanied by briefings to individual customers.

Recent research found that emergency responders in the UK did not find these new ways of presenting seasonal forecasts to be particularly compelling. As a result they were often ignored, partly because of concerns about the skill and robustness of the seasonal forecasts themselves, but also because the coarse spatio-temporal resolution (chosen to convey the lack of robustness) meant the forecasts did not provide information at scales salient for operational decision-making⁷².

In the USA the National Weather Service (NWS) has adopted a slightly different approach for presenting seasonal EPs of temperature, precipitation and hurricane activity. Perhaps reflecting the NWS's larger forecast area, isoline maps are used to present a seasonal outlook (Figure 7, S2) for the probability of the next three months being above or below the climatological average, with more detailed information for specific states and a detailed guide for non-technical users also provided.

Hartmann et al.⁷⁷ and Power et al.⁷⁸ describe some considerations of communication of seasonal forecasts in more detail. The approaches adopted by the NWS and Met Office, the former categorising the predicted trend, and the latter presenting the full distribution, reflect lessons learnt from oversimplification of the forecast, and the need to ensure uncertainties are not ignored in the communication process. It will be interesting for both the NWP and climate communities to monitor the success of these communications, particularly in the backdrop of the UK Parliament demanding better communication of uncertainties⁷⁹.

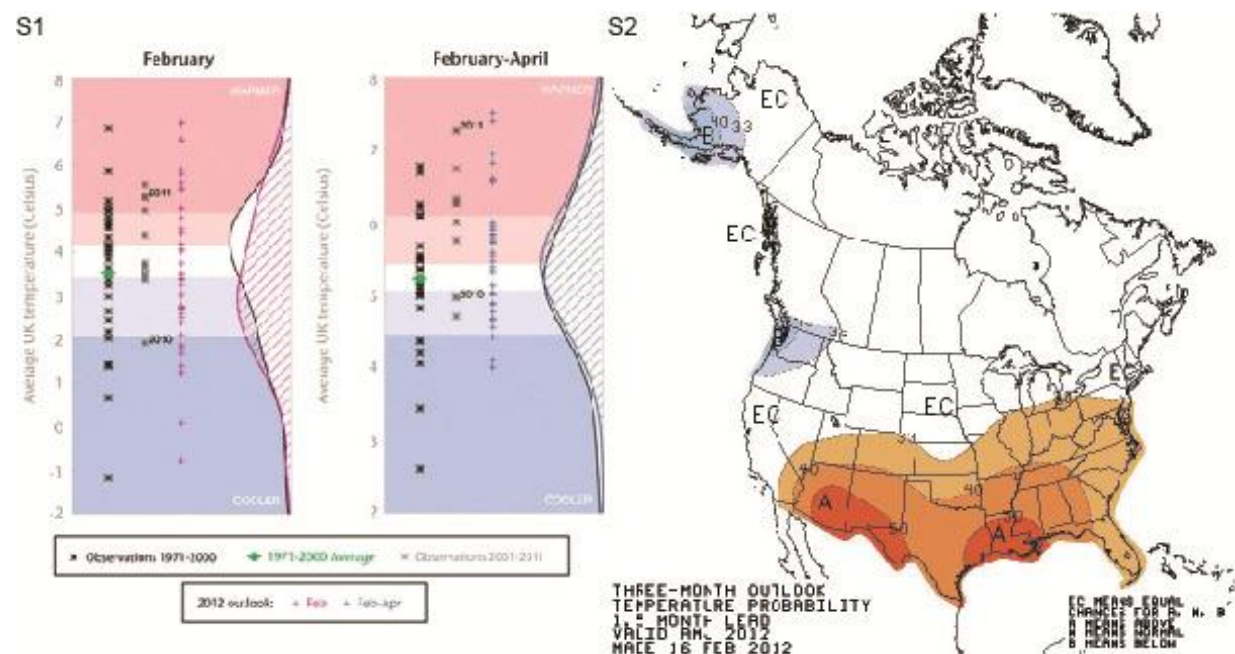


Figure 7: Figure S1: 1-month and 3-month UK outlook for temperature in the context of the observed climatology. UK Met Office Seasonal Outlook, <http://www.metoffice.gov.uk/publicsector/contingency-planners>

Figure S2: National Weather Service, monthly to seasonal outlooks,
http://www.cpc.ncep.noaa.gov/products/predictions/long_range/lead01/off01_temp.gif

5. Discussion

While the challenges of communicating climate EPs are not identical to those of NWP, lessons can be learned from their comparison. Both domains face fundamental challenges in balancing user requirements for saliency with the need for information richness to represent the multidimensional array of uncertain information, and for robustness in the communication of the limitations of the EP. Tensions are inevitable between the needs of users and the limitations of ensembles. One clear lesson from the NWP experience is the importance of multiple formats for representing EPs: there is no ‘one size fits all’⁴². Whereas an expert user of climate predictions may demand an information-rich display such as a time-varying probability distribution, some might argue that the climate ensemble cannot provide information of this saliency if the ‘uncertain uncertainty’ is taken into account. This is where climate ensembles diverge from their weather counterparts; given the short timescale of weather, it is possible to calibrate an NWP ensemble and thus assess whether the full range of uncertainty has been accounted for. Although calibration for extreme events remains challenging (e.g. Stephenson et al.⁸⁰), in general a calibrated weather ensemble can be relied upon to provide a robust probability distribution.

By contrast, climate EPs are marked by deep uncertainties and so the need for communication of robustness is greater to adequately convey how much or little confidence can be placed in the EP. This leaves a massive communication challenge. Should a narrow ensemble spread be communicated as a small uncertainty? How does one communicate ignorance and the possibility for rare or unforeseen surprises^{13, 21, 42, 81}? Such aspects are not only unsampled by the ensemble, but not even imagined or at least not representable. These problems lead Stainforth et al.⁸² to recommend that a climate EP be presented only as a “lower bound of maximum uncertainty”. Similar challenges arise in flood forecasting where robustness of the ensemble in representing the true range of uncertainty is also controversial⁶⁰, and the Stainforth et al. description of the ensemble as a lower bound of maximum uncertainty might equally apply. In their concern with communicating ignorance and the possibility for surprise, the climate ensemble community might have some lessons to offer their counterparts in NWP.

In general, the key challenge in communicating EP for NWP purposes is the balance between richness and saliency; presenting the probability distribution in a way that is meaningful for the user, whereas in climate science the tensions between robustness and saliency are more acute. Particularly for variables where there is little agreement amongst ensemble members, or where the degree of independence between ensemble members is not known, it may be preferable to present each climate model’s result individually rather than obscure the message about the lack of robustness by averaging over the ensemble. However, preliminary research³⁰ on the understanding and use of climate EPs for adaptation planning suggests that these increases in richness, required to convey the lack of robustness, may confuse users and reduce the perceived saliency of EPs. One approach to displaying agreement is the tabular presentation used by several flood forecasting agencies (e.g. Figure 6, F1) to show the number of ensembles or ensemble members above a given threshold value. A balance must be struck between communicating robustness, which requires more

'bandwidth' to represent the nuances of the science, and user requirements for saliency, which tend to involve dimensional reduction and less detail.

Rather than relying on intuition about best practice, the NWP community has benefited from an evidence-based approach to assessing how best to present ensemble information. While EPs in NWP are presented both in relatively raw forms such as spaghetti plots (Figure 6, F4 and Figure 5, H2), and also smoothed information such as fan charts (Figure 4, W5) and the cone of uncertainty (Figure 5, H1), research exploring the experiences of flood forecasting (with PPEs) and hurricane forecasters (often MMEs) provide evidence that many users find such raw data representations difficult to understand⁵³ and that saliency can be improved with less information-rich displays of key summary statistics^{60, 83}. As yet there has been comparatively little research on the communicative effectiveness of different ways of presenting climate ensembles. For instance, more research is required to determine whether the perceived saliency of statistical information can be increased by translating it into experiential and emotional information that draw on personal experiences, consider cultural context, and affect emotions (though one must be aware of the cognitive and cultural biases that can influence interpretation^{17, 28, 84, 85}.) Some examples for climate could include analogies using past events such as the changing frequency of a historical climatic extreme⁸⁶, and familiar representations such as the migrating state map, thermometer range and roulette wheel (Figure 3). However, these simple examples are limited in their capacity to represent a wide range of uncertainties, so care must be taken to effectively communicate robustness to ensure that the content remains a faithful representation of the science.

Another lesson from the NWP literature is the importance of interactivity and user engagement to improve the communication (and therefore use) of ensemble information^{46, 87}. The communication and understanding of climate ensembles could be addressed in the same way⁸⁸. Indeed, the 'ClimatePrediction.net' distributed computing project⁵, C-ROADS simulator (<http://climateinteractive.org>) and UK Climate Projections 2009 interactive website⁸⁹ all offer interactive experiences that provide an opportunity to engage users and improve interpretation. The workshop-style end-user engagement seen in the HEPS literature⁶² is a good step towards improving understanding. In fact, Nobert et al.⁸³ argue that one of the reasons that Sweden has enjoyed such success with its HEPS is the commitment to engaging with its users and seeking their advice both on the best ways of visualising HEPS but also on the information they need from HEPS to inform operational decisions.

It is also important to recognize how words and phrases used by scientists to represent the state of their knowledge can be liable to ambiguity and misinterpretation, as Drake et al.⁵¹ have shown for hurricane forecasts. Perhaps these difficulties should not be surprising, since scientists themselves have often used ambiguous and contradictory translations of IPCC guidelines on calibrated language²⁶. Understanding climate EPs requires both scientific and statistical literacy; it has been estimated that the most recent IPCC Working Group I Summary for Policy Makers¹ requires 17 years of education to be understood⁸⁸. But while literacy is clearly important, a key lesson from the NWP literature is that small technical misunderstandings of EPs do not necessarily affect the decision-making ability of recipients³⁷. Perhaps what matters more is for end-users to be able to use what they garner from EPs to inform their deliberations, rather than for them to be able to reiterate scientific technicalities.

Conclusion

The NWP literature has demonstrated that conclusions from research in other fields (such as health) are not necessarily transferrable (e.g. whether to use frequencies or relative probabilities). Conversely, despite its unique characteristics, the process of addressing communication challenges in climate science has focused mainly on the wider social and decision science literature rather than collecting empirical evidence specific to communicating climate predictions^{33, 34}. There is therefore a clear research gap for climate focussed studies that follow similar lines to that carried out for PoP, hurricanes and floods; using behavioural economics experiments and workshop-style end-user engagement to improve communication of EPs. However, there will need to be differences in how such empirical studies are conducted due to the effect of the heavily politicised nature of climate change and the fact that, compared to climate projections, people have their own experiences of multiple outcomes of weather forecasts that will influence communication and understanding. Studies should also be undertaken to look at the ability of experience-based activities and user engagement to improve the understanding, interpretation and communication of EPs.

Achieving the right balance between the three communication imperatives (Figure. 1) of saliency for different user groups, information richness, and adequate representation of robustness should be seen as one of the key challenges for the communication of climate EPs. There is plenty of scope for research to improve understanding of the requirements for their salient communication. The most recent IPCC guidance note¹⁷ is a first step towards an improved use of language, and future studies could replicate NWP (and health sector) research to determine whether conditional probabilities or frequencies should be used for informing decision making under climate change. Additional consideration would also need to be given to the communication of deep uncertainties. In considering the balance between the three communication imperatives, an obvious area for study is the difference in the perception of robustness, interpretation and use of different levels of information richness, e.g. raw ensemble output (e.g. spaghetti plots) and smoothed versions of the same data (e.g. fan charts). Addressing whether smoothing leads to overconfidence in the robustness of predictions and whether raw model output is used as intended will help define requirements for salient communication. Best practice for uncertainty quantification of climate EPs are outside the scope of this paper (e.g. Hargreaves⁶), but the challenges arising from the long timescales of climate change do propagate into their communication. Accordingly, there needs to be dialogue between all stakeholders in the production and dissemination of climate science to reach a consensus on how to achieve salient communication in the face of controversy over the extent to which the ensemble can be said to represent reality.

Acknowledgements

Elisabeth Stephens and David Demeritt are both supported by the EU FP7 KULTURisk project (www.kulturisk.eu) via grant FP7-ENV-2010-265280. David Demeritt is also supported by the ESRC (RES-189-25-0286). Thanks also to comments from two anonymous reviewers that have greatly helped in revising the paper.

References

1. Solomon SDQMMZCMMKBAMTaHLM. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel

on Climate Change. 2007. Available at:
http://www.ipcc.ch/publications_and_data/ar4/wg1/en/contents.html.

2. NRC NRC. Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts. 2006.
3. Andronova NG, Schlesinger ME. Objective estimation of the probability density function for climate sensitivity. *Journal of Geophysical Research* 2001, 106:PP. 22,605-622,611-PP. 622,605-622,611.
4. Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, Collins M, Stainforth DA. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* 2004, 430:768-772.
5. Stainforth DA, Aina T, Christensen C, Collins M, Faull N, Frame DJ, Kettleborough JA, Knight S, Martin A, Murphy JM, et al. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* 2005, 433:403-406.
6. Hargreaves JC. Skill and uncertainty in climate models. *Wiley Interdisciplinary Reviews: Climate Change* 2010, 1:556-564.
7. Meehl GA, Covey C, Taylor KE, Delworth T, Stouffer RJ, Latif M, McAvaney B, Mitchell JFB. THE WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research. *Bulletin of the American Meteorological Society* 2007, 88:1383-1394.
8. Hawkins E, Sutton R. The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society* 2009, 90:1095-1107.
9. Lorenz EN. Climatic Predictability. The physical basis of climate and climate modeling. GARP Publication Series, Geneva: World Meteorological Organization, 16:132-136.
10. Nakicenovic NJ, Alcamo G, Davis B, de Vries J, Fenhann S, Gaffin K, Gregory A, Gruebler ea. Special Report on Emissions Scenarios, Working Group III of the Intergovernmental Panel on Climate Change. 2000, Page 595.
11. Parker WS. Predicting weather and climate: Uncertainty, ensembles and probability. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics* 2010, 41:263-272.
12. Smith LA. What Might We Learn from Climate Forecasts? *Proceedings of the National Academy of Sciences of the United States of America* 2002, 99:2487-2492.
13. Stainforth DA, Allen MR, Tredger ER, Smith LA. Confidence, Uncertainty and Decision-Support Relevance in Climate Predictions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2007, 365:2145-2161.
14. Rougier J. Probabilistic Inference for Future Climate Using an Ensemble of Climate Model Evaluations. *Climatic Change* 2007, 81:247-264.

15. Sexton DMH, Murphy JM. Multivariate probabilistic projections using imperfect climate models. Part II: robustness of methodological choices and consequences for climate sensitivity. *Climate Dynamics* 2011, 38:2543-2558.
16. Sexton DMH, Murphy JM, Collins M, Webb MJ. Multivariate probabilistic projections using imperfect climate models part I: outline of methodology. *Climate Dynamics* 2012, 38:2513-2542.
17. Mastrandrea MDCB, F. FT, O. S, L. EK, J. ED, H. F, E. H, J. KK, R. MP, G. M, et al. Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of UncertaintiesIPCC WGII - Cross-Working Group Consultations. 2010. Vol. AR5 Guidance Note Annex, Pages 4-4. Available at: <http://www.ipcc-wg2.gov/meetings/CGCs/index.html>.
18. Mastrandrea MDCBFTFSOEKLEDJFHHEKKJM. Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of UncertaintiesIPCC WGII - Cross-Working Group Consultations. Vol. AR5 Guidance Note Annex, Pages 4-4. Available at: <http://www.ipcc-wg2.gov/meetings/CGCs/index.html>.
19. Wallsten TS, Budescu DV, Rapoport A, Zwick R, Forsyth B. Measuring the vague meanings of probability terms. *Journal of Experimental Psychology-General* 1986, 115:348-365.
20. Moxey LM, Sanford AJ. Communicating quantities: A review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology* 2000, 14:237-255.
21. Kandlikar, Risbey J, Dessai S. Representing and communicating deep uncertainty in climate-change assessments. *Comptes Rendus Geosciences* 2005, 337:443-455.
22. Challenor P, McNeill D, Gattiker J. Assessing the probability of rare climate events. In: O'Hagan A, West M, eds. *The Oxford Handbook of Applied Bayesian Analysis*. Oxford: Oxford University Press; 2010, 403-430.
23. Kaye N. An assessment of mapping techniques to visualise uncertainty in climate data. 2010. Vol. 81.
24. Demeritt D. Science studies, climate change and the prospects for constructivist critique. *Economy and Society* 2006, 35:453-479.
25. Kahan DM, Jenkins-Smith H, Braman D. Cultural cognition of scientific consensus. *Journal of Risk Research* 2010, 14:147-174.
26. Mastrandrea M, Mach K. Treatment of uncertainties in IPCC Assessment Reports: past approaches and considerations for the Fifth Assessment Report. *Climatic Change* 2011, 108:659-673.
27. Hall J, Twyman C, Kay A. Influence Diagrams for Representing Uncertainty in Climate-Related Propositions. *Climatic Change* 2005, 69:343-365.
28. Morgan MG, Mellon C. Certainty, uncertainty, and climate change. *Climatic Change* 2011, 108:707-721.
29. Demeritt D, Langdon D. The UK Climate Change Programme and communication with local authorities. *Global Environmental Change* 2004, 14:325-336.

30. Tang S, Dessai S. Usable science? The UK Climate Projections 2009 and decision support for adaptation planning. *Environmental Science and Policy* In Review.
31. Frumhoff PC, McCarthy JJ, Melillo JM, Moser SC, Wuebbles DJ. Confronting Climate Change in the U.S. Northeast: Science, Impacts, and Solutions. A report of the Northeast Climate Impacts Assessment.
32. Kaye NR, Hartley A, Hemming D. Mapping the climate: guidance on appropriate techniques to map climate variables and their uncertainty. *Geoscientific Model Development Discussions* 2011, 4:1875-1906.
33. Fischhoff B. Applying the science of communication to the communication of science. *Climatic Change* 2011, 108:701-705.
34. Pidgeon N, Fischhoff B. The role of social and decision sciences in communicating uncertain climate risks. *Nature Climate Change* 2011, 1:35-41.
35. Gigerenzer G, Hertwig R, van den Broek E, Fiasolo B, Katsikopoulos KV. "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis* 2005, 25:623-629.
36. Handmer J, Proudley B. Communicating uncertainty via probabilities: The case of weather forecasts. *Environmental Hazards* 2007, 7:79-87.
37. Morss RE, Demuth JL, Lazo JK. Communicating Uncertainty in Weather Forecasts: A Survey of the US Public. *Weather and Forecasting* 2008, 23:974-991.
38. Murphy AH, Lichtenstein S, Fischhoff B, Winkler RL. Misinterpretations of Precipitation Probability Forecasts. *Bulletin of the American Meteorological Society* 1980, 61:695-701.
39. Lazo JK, Morss RE, Demuth JL. 300 Billion Served. Sources, Perceptions, Uses, and Values of Weather Forecasts. *Bulletin of the American Meteorological Society* 2009, 90:785-+.
40. Gigerenzer G. *Reckoning with Risk: Learning to Live with Uncertainty*. London: Penguin; 2003.
41. Lipkus IM. Numeric, verbal, and visual formats of conveying health risk: Suggested best practices and future recommendations. *Medical Decision Making* 2007, 27:696-713.
42. Spiegelhalter D, Pearson M, Short I. Visualizing Uncertainty About the Future. *Science* 2011, 333:1393-1400.
43. Krämer W, Gigerenzer G. How to Confuse with Statistics or: The Use and Misuse of Conditional Probabilities. *Statistical Science* 2005, 20:223-230.
44. Joslyn SL, Nichols RM. Probability or frequency? Expressing forecast uncertainty in public weather forecasts. *Meteorological Applications* 2009, 16:309-314.
45. Woloshin S, Schwartz LM. Communicating Data About the Benefits and Harms of Treatment A Randomized Trial. *Annals of Internal Medicine* 2011, 155:87-U70.

46. Stephens E, Spiegelhalter D, Mylne K, Harrison M. Impact of presentation method and socio-demographics on decision-making using a Probability of Precipitation forecast. *Bulletin of the American Meteorological Society* In Review.
47. Joslyn SL, LeClerc JE. Uncertainty Forecasts Improve Weather-Related Decisions and Attenuate the Effects of Forecast Error. *Journal of Experimental Psychology: Applied* 2011, Advance Online Publication.
48. Roulston MS, Bolton GE, Kleit AN, Sears-Collins AL. A laboratory study of the benefits of including uncertainty information in weather forecasts. *Weather and Forecasting* 2006, 21:116-122.
49. Roulston MS, Kaplan TR. A laboratory-based study of understanding of uncertainty in 5-day site-specific temperature forecasts. *Meteorological Applications* 2009, 16:237-244.
50. Broad K, Leiserowitz A, Weinkle J, Steketee M. Misinterpretations of the "Cone of Uncertainty" in Florida during the 2004 Hurricane Season. *Bulletin of the American Meteorological Society* 2007, 88:651-+.
51. Drake L. Hurricane forecast and end-users: bridging the communication gap. 2010 National Hurricane Conference 2010.
52. Morss RE, Hayden MH. Storm Surge and "Certain Death": Interviews with Texas Coastal Residents following Hurricane Ike. *Weather, Climate and Society* 2010, 2:174-189.
53. Eosco G. Pictures may tell it all: The use of draw-and-tell methodology to understand the role of uncertainty in individuals' hurricane information seeking processes. Fifth Symposium on Policy and Socio-economic Research. Second AMS Conference on International Cooperation in the Earth System Sciences and Services 2010.
54. Orlove BS, Broad K, Meyer R. Assessing the Effectiveness of the Cone of Probability as a Visual Means of Communicating Scientific Forecasts. *AGU Fall Meeting Abstracts* 2010, -1:0658-0658.
55. Cloke H, Thielen J, Pappenberger F, Nobert S, Bálint G, Edlund C, Koistinen A, de Saint-Aubin C, Sprockereef E, Viel C, et al. Progress in the implementation of Hydrological Ensemble Prediction Systems (HEPS) in Europe for operational flood forecasting. *ECMWF Newsletters* 2009, Pages 20-24. Available at: <http://www.ecmwf.int/publications/newsletters/pdf/121.pdf>.
56. Cloke HL, Pappenberger F. Ensemble flood forecasting: A review. *Journal of Hydrology* 2009, 375:613-626.
57. Australia Co. Flood Warning. 2009. Vol. Manual 21. Available at: [http://www.em.gov.au/Documents/Manual%2021-Flood%20Warning\(2\).PDF](http://www.em.gov.au/Documents/Manual%2021-Flood%20Warning(2).PDF).
58. Dale M, Ji Y, Wicks J, Mylne K, Pappenberger F, Cloke H. Applying probabilistic flood forecasting in Flood Incident Management. 2012. Vol. SC0900032. Available at: <http://evidence.environment-agency.gov.uk/FCERM/en/Default/HomeAndLeisure/Floods/WhatWereDoing/IntoTheFuture/Scienc>

eProgramme/ResearchAndDevelopment/FCRM/Project.aspx?ProjectID=c0899b01-6fcd-4775-9aeb-a2dcc8ec4d39&PageID=7ef5c014-c33b-4eb4-af59-ec6f0a09d80f.

59. Demeritt D, Nobert S, Cloke HL, Pappenberger F. The European Flood Alert System (EFAS) and the communication, perception and use of ensemble predictions for operational flood risk management. *Hydrological Processes* 2012:n/a–n/a–n/a–n/a.
60. Demeritt D, Nobert S, Cloke H, Pappenberger F. Challenges in communicating and using ensembles in operational flood forecasting. *Meteorological Applications* 2010, 17:209-222.
61. Ramos M-H, Mathevet T, Thielen J, Pappenberger F. Communicating uncertainty in hydro-meteorological forecasts: mission impossible? *Meteorological Applications* 2010, 17:223-235.
62. Pappenberger F, Stephens E, Thielen J, Salamon P, Demeritt D, van Andel SJ, Wetterhall F, Alfieri L. Visualizing probabilistic flood forecast information: expert preferences and perceptions of best practice in uncertainty communication. *Hydrological Processes* 2012:n/a–n/a–n/a–n/a.
63. Faulkner H, Parker D, Green C, Beven K. Developing a translational discourse to communicate uncertainty in flood risk between science and the practitioner. *Ambio* 2007, 36:692-703.
64. Roo AD, Thielen J, Salamon P, Bogner K, Nobert S, Cloke H, Demeritt D, Younis J, Kalas M, Bódis K, et al. Quality control, validation and user feedback of the European Flood Alert System (EFAS). *International Journal of Digital Earth* 2011, 4:77-90.
65. Demeritt D, Nobert S. Responding to early flood warning in the European Union. In: Meyer CO, De Franco C, eds. *Forecasting, Warning and Responding to Transnational Risks*. London: Palgrave Macmillan; 2011.
66. Bogner K, Pappenberger F. Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. *Water Resources Research* 2011, 47:24 PP.-24 PP.
67. Bruen M, Krahe P, Zappa M, Olsson J, Vehvilainen B, Kok K, Daamen K. Visualizing flood forecasting uncertainty: some current European EPS platforms—COST731 working group 3. *Atmospheric Science Letters* 2010, 11:92-99.
68. Frick J, Hegg C. Can end-users' flood management decision making be improved by information about forecast uncertainty? *Atmospheric Research* 2011, 100:296-303.
69. Priest S, Parker D, Hurford A, Pardoe A, McCarthy S, Tapsell S. *Surface Water Flood Warning Scoping Project—Final Report*. 2011. Vol. SCSC080034/R1.
70. Rothstein H, Borraz O, Huber M. From the 'neurotic' to the 'rationalising' state: risk and the limits of governance. Responding to early flood warning in the European Union. In: De Franco C, Meyer CO, eds. *Forecasting, Warning and Responding to Transnational Risks*. London: Palgrave Macmillan; 2011.
71. Rothstein H, Downer J. 'RENEWING DEFRA': EXPLORING THE EMERGENCE OF RISK-BASED POLICYMAKING IN UK CENTRAL GOVERNMENT. *Public Administration* 2012:no–no–no–no.

72. Demeritt D. The perception and use of public weather services by emergency and resiliency professionals in the UK. 2012, Pages 36-36.
73. Lange H, Garrelts H. Risk Management at the Science–Policy Interface: Two Contrasting Cases in the Field of Flood Protection in Germany. *Journal of Environmental Policy & Planning* 2007, 9:263-279.
74. Krieger K. Putting Varieties of Risk-Based Governance into Institutional Context: The Case of Flood Management Regimes in Germany and England in the 1990s and 2000s. PhD Thesis. King's College London. 2011.
75. Alleyne R. Britain will have first decent 'barbecue summer' in three years with temperatures regularly above 80 F. *The Daily Telegraph* 2009. Available at: <http://www.telegraph.co.uk/topics/weather/5250745/Britain-will-have-first-decent-barbecue-summer-in-three-years-with-temperatures-regularly-above-80-F.html>.
76. Eden P. So much for our barbecue summer. *The Daily Telegraph* 2009. Available at: <http://www.telegraph.co.uk/comment/personal-view/5934444/So-much-for-our-barbecue-summer.html>.
77. Hartmann HC, Pagano TC, Sorooshian S, Bales R. Confidence Builders: Evaluating Seasonal Climate Forecasts from User Perspectives. *Bulletin of the American Meteorological Society* 2002, 83:683-698.
78. Power SB, Plummer N, Alford P. Making climate model forecasts more useful. *Aust. J. Agric. Res.* 2007, 58:945-951.
79. Committee SaTS. Science in the Met Office Report. 2012. Available at: <http://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/news/120221-met-office-rpt-published/>.
80. Stephenson DB, Casati B, Ferro CAT, Wilson CA. The extreme dependency score: a non-vanishing measure for forecasts of rare events. *Meteorological Applications* 2008, 15:41-50.
81. Curry J. Reasoning about climate uncertainty. *Climatic Change* 2011, 108:723-732.
82. Stainforth DA, Downing TE, Washington R, Lopez A, New M. Issues in the interpretation of climate model ensembles to inform decisions. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 2007, 365:2163-2177.
83. Nobert S, Demeritt D, Cloke H. Informing operational flood management with ensemble predictions: lessons from Sweden. *Journal of Flood Risk Management* 2010, 3:72-79.
84. Patt AG, Schrag DP. Using specific language to describe risk and probability. *Climatic Change* 2003, 61:17-30.
85. Marx SM, Weber EU, Orlove BS, Leiserowitz AA, Krantz D, Roncoli C, Phillips J. Communication and Mental Processes: Experimental and Analytic Processing of Uncertain Climate Information. SSRN eLibrary 2007.

86. Stott PA, Stone DA, Allen MR. Human contribution to the European heatwave of 2003. *Nature* 2004, 432:610-614.
87. Pappenberger F, Stephens E, Thielen J, Salamon P, Demeritt D, van Andel SJ, Wetterhall F, Alfieri L. Visualizing probabilistic flood forecast information: expert preferences and perceptions of best practice in uncertainty communication. *Hydrological Processes* 2012:n/a-n/a.
88. Sterman J. Communicating climate change risks in a skeptical world. *Climatic Change* 2011, 108:811-826.
89. Murphy JMea. Climate change projections. In UK Climate Projections. UKCP09: <http://ukclimateprojections.defra.gov.uk>. 2009, Pages 1-194.

Figure captions

Figure 1: The three imperatives for visualisation

Figure 2: C1, Multi-model global means (solid lines) and ± 1 standard deviation range of individual model annual averages. (©IPCC1). By permission of Cambridge University Press. C2, Probability density of the strength of the Meridional Overturning Circulation. (Challenor et al.22, © Oxford University Press 2010). By permission of Oxford University Press. C3, Changes in 20 year-mean surface air temperature over the HadSM3 grid box corresponding to Wales, in March, in response to doubled CO₂. (UKCP09, © UK Climate Projections, 2009). C4, Relative changes in precipitation (%) for the period 2090–2099, relative to 1980–1999 (©IPCC1). By permission of Cambridge University Press. C5, New mapping technique illustrating change in precipitation (similarly to C4) with hues and percentage model agreement across the ensemble with saturation. (©Kaye et al.23).

Figure 3: C6, Moving state graphic; C7, Thermometers showing projected temperature increases, both from Frumhoff et al. 31 © 2007 Union of Concerned Scientists). By permission of Union of Concerned Scientists. C8, Roulette-style spinning wheels to depict estimated probability. Accessed 27th Feb 2012 from <<http://globalchange.mit.edu/resources/gamble>>. By permission of MIT Global Change Program.

Figure 4: W1, PoP with no graphic, from wunderground.com forecast for Des Moines, IA, US. Accessed from: <http://www.wunderground.com/q/zmw:50301.1.99999>. W2: PoP with probability bar graphic for Sydney from the Bureau of Meteorology, Australia. Accessed from <http://www.bom.gov.au/nsw/forecasts/sydney.shtml>. W3: PoP with probability pie charts, University of Washington Probcast. Accessed from: <http://probcast.washington.edu/>. W4: PoP with probability bar, note how vertical bar and blue colour might cause confusion with the amount of rain. Accuweather, accessed from: <http://www.accuweather.com/en/us/new-york-ny/10017/weather-accupop/3712pc>. W5: Time series showing 50% and 80% probability range for temperature and precipitation amount, Norwegian Meteorological Institute. Accessed from: <http://www.yr.no/place/Norway/Oslo/Oslo/Oslo/long.html>. W6: Temperature range bar, showing 90% range for predicted maximum (and minimum in separate tab) temperatures. UK Met Office, accessed from: <http://www.metoffice.gov.uk/public/beta/weather/forecast/?tab=fiveDay&dayIdx=0&locId=350610>.

Figure 5: Figure H1: Spaghetti plot showing hurricane tracks. ABC Weather, accessed from: <http://www.wjla.com/blogs/weather/2011/08/hurricane-irene-path-projections-spaghetti-style-12544.html>. Figure H2: Hurricane Cone of Uncertainty, note that the estimated 'best forecast track' has now been removed to avoid confusion. National Hurricane Center (US), accessed from: <http://newsfeed.time.com/2011/08/26/hurricane-irenes-path-how-do-forecasters-predict-the-cone-of-uncertainty/>

Figure 6: F1, EFAS forecast for southern Poland. Courtesy of EFAS, Joint Research Centre, European Commission, Ispra, Italy. F2, Flood Guidance Statement, joint Met Office / Environment Agency Flood Forecasting Centre. F3, the Vigicrues flood risk used by the SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations/ Central Service for Hydrometeorology and Flood Prediction Support) in France. F4, spaghetti plot of forecasted precipitation, Austrian Fire Service & Civil Defence Early Warning Centre and F5, their publicly accessible simplification of the uncertainty.

Figure 7: Figure S1: 1-month and 3-month UK outlook for temperature in the context of the observed climatology. UK Met Office Seasonal Outlook, <http://www.metoffice.gov.uk/publicsector/contingency-planners>, Figure S2: National Weather Service, monthly to seasonal outlooks, http://www.cpc.ncep.noaa.gov/products/predictions/long_range/lead01/off01_temp.gif

Further Reading/Resources

1) Glossary of terms

Boundary conditions – model inputs that determine the evolution of the system; e.g. emissions scenarios

Boundary condition ensemble – ensemble in which different boundary conditions are used for each simulation

Ensemble – group of models, or group of simulations generated with different models or model inputs

Emissions scenario – plausible future trajectory of factors that influence climate, including emissions of greenhouse gases and air pollutants, and changes in land-use

EP – ensemble prediction

Hydrograph – graph showing rate of flow versus time for a given location

ICE – initial condition ensemble

Initial conditions – model inputs that determine the starting state of the system, such as three-dimensional fields of atmospheric and ocean temperatures

Initial condition ensemble – ensemble in which different initial conditions are used for each simulation

Initial condition uncertainty – uncertainty due to imperfectly known initial conditions, in particular due to uncertainty in observing the present state of the atmosphere and ocean

MME – multi-model ensemble

Multi-model ensemble – ensemble of models of different structures, usually a group of the models developed by research and meteorological institutes around the world

Parameters – tunable ‘control dials’ of the model, typically there to represent processes that are not included explicitly due to finite resolution or imperfect knowledge; e.g. threshold of relative humidity for cloud formation

Parameter uncertainty – uncertainty due to model parameters for which the best settings are not known

Perturbed parameter ensemble – ensemble in which different values of the model control parameters are used for each simulation

PPE – perturbed parameter ensemble

Structural uncertainty – uncertainty about the adequacy of a model in describing reality, due to its finite spatial and temporal resolution and the physical, chemical or biological processes that are missing or imperfectly represented

2) Summary tables of visualisation types

Table 1: Climate ensemble visualisations

	Variables / data dimensionality	Information richness	Target Audience	URL of Visualisations
Uncertain timeseries	Various; timeseries	Time evolution of percentiles or density of probability distribution	Policymakers	Figure 2, C1 Figure 2, C2
Climatic envelope	Two variables (various); 2D contour plot	Percentiles of joint probability distribution	Policymakers, public	http://ukclimateprojections.defra.gov.uk/22625
Stippled map	Various; 2D (lat-lon) map	Mean and three categories of % sign agreement (0-60%; 61-	Policymakers	Figure 2, C4

		89%; 90-100%)		
Colour saturated map	Various; 2D (lat-lon) map	Mean and % sign agreement or signal-to-noise ratio	Various	Figure 2, C5
Histogram/PDF	Various; frequency or probability	Marginal frequency distribution / probability distribution function	Policymakers, public	Figure 2, C3
Percentiles	Various	Percentiles of probability distribution	Various	IPCC ¹ http://www.ipcc.ch/publications_and_data/ar4/wg1/en/ch10s10-5.html#10-5-1 (Box 10.2, Fig. 1, top right) http://ukclimateprojections.org.uk/content/view/1931/500/
Migrating states	Regional temperature	Mean	Policymakers, public	Figure 3, C6
Thermometer	Temperature	Mean and/or range	Policymakers, public	Figure 3, C7
Roulette wheel	Global mean temperature	Frequency	Public	Figure 3, C8

Table 2: Visualisation of site-specific weather forecasts

	Variables / data dimensionality	Information richness	Target Audience	URL of Visualisations
Probability of Precipitation (PoP) or	Varies, usually for a specific time period	Probability of threshold	Public etc	Fig.W1 http://www.weather.com/weather/hourbyhour/graph

Chance of Precipitation		(any rain) Precision or Probability varies		/USNY0996 Fig.W2 http://www.wunderground.com/q/zmw:10001.2.99999
PoP with graphic	Varies, usually for a specific time period			Fig.W3 http://www.bom.gov.au/ns/w/forecasts/sydney.shtml Fig.W4 http://www.accuweather.com/us/ny/new-york/10017/forecast-accupop.asp?fdays=1 Fig.W5 http://probcast.washington.edu/
Rainfall distributions	Time series	Percentiles (middle 50% & 80%)	Public	Fig.W6 http://www.yr.no/place/Norway/Oslo/Oslo/Oslo/long.html
Temperature fancharts	Time series	Percentiles (middle 50% & 80%)	Public	Fig.W6 http://www.yr.no/place/Norway/Oslo/Oslo/Oslo/long.html
Temperature Range Plot	Time series	Median and 90% range	Public	Fig.W6 http://www.metoffice.gov.uk/public/beta/weather/forecast/?tab=fiveDay&dayIdx=0&locId=350610

Table 3: Visualisation of hurricane forecasts

	Variables / data dimensionality	Information richness	Target Audience	URL of Visualisations
Spaghetti plots	Time and space	Paths from individual models	Public	Fig.W8 http://www.wjla.com/blogs/weather/2011/08/hurricane-irene-path-projections-spaghetti-style-12544.html
Cone of uncertainty	Time and space	Derived fan from individual paths	Public	Fig.W9 http://newsfeed.time.com/2011/08/26/hurricane-irenes-path-how-do-forecasters-predict-the-cone-of-uncertainty/

Table 4: Visualisation of probabilistic flood forecasts

	Variables / data dimensionality	Information richness	Target Audience	URL of Visualisations
Spaghetti hydrograph	Flow in m ³ /sec over time	Time evolution of percentiles or density of probability distribution	expert users,	Fig.W10 http://www.nwrfc.noaa.gov/espdp/espdp.cgi
Simplified spaghetti hydrograph	flow in m ³ /sec over time	Mean, median, 5-95% confidence intervals	Emergency responders, general public	Fig.W11 http://wwwi2.ymparisto.fi/i2/04/I049411001y/wgen.html ; http://swissrivers.ch/
Threshold exceedance maps	Pixels where EFAS thresholds exceeded over next 3-10 days	Spatial distribution of categorical levels of threshold exceedance, but neither the values, nor distribution of ensemble members across different threshold levels is shown	Hydrologists,	http://floods.jrc.ec.europa.eu/efas-flood-forecasts
Tables of ensemble members exceeding a given threshold over time at a given point	Number of threshold members exceeding various threshold levels	No spatial distribution or values, but shows the temporal evolution of the signal	Emergency responders	Fig.W12 Fig.W13
Traffic light based flooding hazard map	Spatial distribution of likelihood of fluvial flooding in next 24 hours	Color coded probability categories (green, yellow, amber red) of the likelihood of flooding	General public, emergency responders	Fig.W14 http://www.vigicrues.gouv.fr/
Flood Guidance Statement	Spatial distribution of risk of flooding	im	Emergency planners	Fig.W15 http://www.ffc-environment-agency.metoffice.gov.uk/services/FGS_User_Gui

	over next 5 days			de.pdf
--	------------------	--	--	------------------------

Table 5: Visualisation of seasonal forecasts

	Variables / data dimensionality	Information richness	Target Audience	URL of Visualisations
1-month / 3-month outlook	Seasonal for UK	Predicted distributions for temperature and precipitation	Expert (Contingency planners)	Fig.W16 http://www.metoffice.gov.uk/publicsector/contingency-planners
Monthly to seasonal outlooks	Seasonal for USA	Temperature and Precipitation, normal, above normal, below normal	Includes summary for non-technical users	Fig.W17 http://www.cpc.ncep.noaa.gov/products/predictions/long_range/lead01/off01_temp.gif Fig.W18 http://www.cpc.ncep.noaa.gov/products/predictions/long_range/lead01/off01_temp.gif
Hurricane Outlook (text-based forecast)	Seasonal for a particular area	Probability of above normal / normal / below normal		http://www.cpc.ncep.noaa.gov/products/outlooks/hurricane.shtml

1) Expanded figure captions

Figure C1: Multi-model global means (solid lines) and ± 1 standard deviation range of individual model annual averages (shading) of surface warming for the 20th century and SRES scenarios A2, A1B and B1. The grey bars at right indicate the best estimate (solid line within each bar) and the likely range assessed for six SRES scenarios. (IPCC, 2007, © Intergovernmental Panel on Climate Change 2007). By permission of Cambridge University Press.

Figure C2: Probability density of the strength of the Meridional Overturning Circulation through the 21st century. (Challenor et al.²² © Oxford University Press 2010). By permission of Oxford University Press.

Figure C3: Changes in 20 year-mean surface air temperature over the HadSM3 grid box corresponding to Wales, in March, in response to doubled CO₂. Green histogram shows 280 perturbed physics simulations of HadSM3. Black ticks show corresponding changes simulated by 12 multi-model ensemble members. Red curve shows the distribution obtained by emulating responses across the full parameter space of surface and atmospheric processes in HadSM3. The red curve also includes the broadening effect of adding the variance (but not the mean) of discrepancy. Blue curve shows the effects of weighting the emulated responses according to observational constraints see (Section 3.2.9). Black curve shows the posterior distribution, which includes the shift arising from adding in the mean effect of discrepancy. (©UKCP09).

Figure C4: Relative changes in precipitation (%) for the period 2090–2099, relative to 1980–1999. Values are multi-model averages based on the SRES A1B scenario for December to February. White areas are where less than 66% of the models agree in the sign of the change and stippled areas are where more than 90% of the models agree in the sign of the change. (©IPCC¹). By permission of Cambridge University Press.

Figure C5: New mapping technique illustrating change in precipitation (similarly to C4) with hues and percentage model agreement across the ensemble with saturation. (©Kaye et al.³²)

Figure C6: Red arrows track what summers could feel like in the NYC Tri-State region over the course of the century under the higher-emissions scenario. Yellow arrows track what summers in these states would feel like under a lower-emissions scenario. (Frumhoff et al.³¹, © 2007 Union of Concerned Scientists). By permission of Union of Concerned Scientists.

Figure C7: These “thermometers” show projected increases in regional average summer temperatures for three time periods: early-, mid-, and late- century. (Frumhoff et al.³¹, © 2007 Union of Concerned Scientists). By permission of Union of Concerned Scientists.

Figure C8a&b: The roulette-style spinning wheels depict the estimated probability, or likelihood, of potential temperature change (global average surface temperature) over the next 100 years. The face of each wheel is divided into coloured slices, with the size of each slice representing the estimated probability of the temperature change in the year 2100 falling within that range. The Greenhouse Gamble wheel on the left is the “no policy” or reference case, in which it is assumed no action is taken to try to curb the global emissions of greenhouse gases. The Greenhouse Gamble wheel on the right is the “with policy” case, which assumes that policies are enacted to limit cumulative emissions of greenhouse gases over the century to 4.2 trillion metric tons, measured in CO₂-equivalent. Accessed 27th Feb 2012 from <<http://globalchange.mit.edu/resources/gamble>>. By permission of MIT Global Change Program.

Figure W1: PoP with no graphic, from wunderground.com forecast for Des Moines, IA, US. Accessed from: <http://www.wunderground.com/q/zmw:50301.1.99999>.

Figure W2: PoP with probability bar graphic for Sydney from the Bureau of Meteorology, Australia. Accessed from <http://www.bom.gov.au/nsw/forecasts/sydney.shtml>.

Figure W3: PoP with probability pie charts, University of Washington Probcast. Accessed from: <http://probcast.washington.edu/>

Figure W4: PoP with probability bar, note how vertical bar and blue colour might cause confusion with the amount of rain. Accuweather, accessed from: http://www.accuweather.com/en/us/new-york-ny/10017/weather-accupop/3712_pc.

Figure W5: Time series showing 50% and 80% probability range for temperature and precipitation amount, Norwegian Meteorological Institute. Accessed from: <http://www.yr.no/place/Norway/Oslo/Oslo/Oslo/long.html>

Figure W6: Temperature range bar, showing 90% range for predicted maximum (and minimum in separate tab) temperatures. UK Met Office, accessed from: <http://www.metoffice.gov.uk/public/beta/weather/forecast/?tab=fiveDay&dayIdx=0&locId=350610>

Figure H1: Spaghetti plot showing hurricane tracks. ABC Weather, accessed from: <http://www.wjla.com/blogs/weather/2011/08/hurricane-irene-path-projections-spaghetti-style-12544.html>

Figure H2: Hurricane Cone of Uncertainty, note that the estimated 'best forecast track' has now been removed to avoid confusion. National Hurricane Center (US), accessed from: <http://newsfeed.time.com/2011/08/26/hurricane-irenes-path-how-do-forecasters-predict-the-cone-of-uncertainty/>

Figure F1: Further tabular detail for a selected point in southern Poland from the EFAS forecast for 12:00h UTC on 17 May 2010 when severe flooding resulted in 2.5 billion euros in damages. The first two rows classify the various EFAS river flow forecasts produced for that point using the deterministic rainfall forecasts from DWD (Deutscher Wetterdienst, the German national meteorological service) and ECMWF, with purple indicating flows in excess of the EFAS Severe Alert Level (SAL) corresponding to a simulated flood event with a return period of >20 yr., red indicating flows in excess of the EFAS High Alert Level (HAL) and yellow in excess of the Medium Alert Level. The numbers in the subsequent boxes indicate the number of EFAS ensemble members produced using the ECMWF ensemble (EUE) and the COSMO-LEPS limited area ensemble (COS) that exceed the HAL and Severe Alert Levels (SAL). Courtesy of EFAS, Joint Research Centre, European Commission, Ispra, Italy

Figure F2: "The Flood Guidance Statement issued in England and Wales by the joint Met Office / Environment Agency Flood Forecasting Centre provides a simple cartographic display. This same risk matrix is now also used by the UK Met Office as part of its National Severe Weather Warning Service to communicate the likelihood and impact of severe weather events"

Figure F3: The Vigicrues flood risk used by the SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations/ Central Service for Hydrometeorology and Flood Prediction Support) in France to communicate the risk of flooding over the next 24 hours on main rivers. The Green, yellow, orange, and red pixels represent escalating levels of hazardousness that call for corresponding levels of vigilance in response to the threat. These colour codes do not explicitly distinguish the probability of flooding from its magnitude, which can lead to confusion.

Figure F4: In Austria, emergency services personnel working in the Abteilung Feuerwehr und Zivilschutz Landeswarnzentrale [the Fire Service & Civil Defence Early Warning Centre] have additional access to much richer EP outputs, including 'spaghetti' plots of the 51-member ALADIN-

LAEF of convective rainfall, which are the light colored lines in this plot which also shows the deterministic forecast (Hauptlauf) in black and the observed in red.

Figure F5: In Austria, the public has accessed to simplified HEPS forecasts of streamflow, with the blue line showing observation, the green line a 'best guess' forecast, and the two grey lines the 10% and 90% confidence intervals.

Figure S1: 1-month and 3-month UK outlook for temperature in the context of the observed climatology. UK Met Office Seasonal Outlook,
<http://www.metoffice.gov.uk/publicsector/contingency-planners>

Figure S2: National Weather Service, monthly to seasonal outlooks,
http://www.cpc.ncep.noaa.gov/products/predictions/long_range/lead01/off01_temp.gif