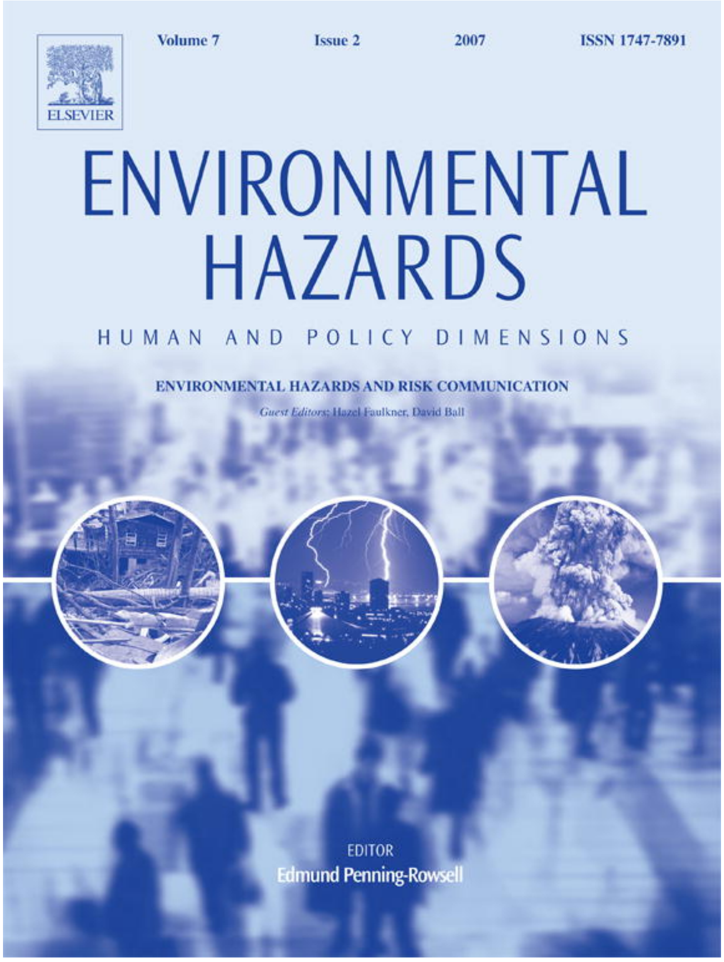
Volume 7
Issue 2
2007
ISSN 1747-7891
ELSEVIER
ENVIRONMENTAL HAZARDS
HUMAN AND POLICY DIMENSIONS
ENVIRONMENTAL HAZARDS AND RISK COMMUNICATION
Guest Editors: Hazel Faulkner, David Ball
EDITOR
Edmund Penning-Rowsell

ELSEVIER

Environmental Hazards 7 (2007) 115–127

ENVIRONMENTAL HAZARDS

www.elsevier.com/locate/hazards

Papers

Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting

David Demeritt[a,*], Hannah Cloke[a], Florian Pappenberger[b,1], Jutta Thielen[c], Jens Bartholmes[c], Maria-Helena Ramos[c]

[a]*Department of Geography, King's College London, Strand, London WC2R 2LS, UK*
[b]*Department of Environmental Sciences, Lancaster University, Lancaster LA1 4YQ, UK*
[c]*Institute for Environment and Sustainability, Joint Research Centre, European Commission, Via E. Fermi 1, I 21020 Ispra (VA), Italy*

Abstract

Under the auspices of the World Meteorological Organization, there are a number of international initiatives to promote the development and use of so-called ensemble prediction systems (EPS) for flood forecasting. The campaign to apply these meteorological techniques to flood forecasting raises important questions about how the probabilistic information these systems provide can be used for what in operational terms is typically a binary decision of whether or not to issue a flood warning. To explore these issues, we report on the results of a series of focus group discussions conducted with operational flood forecasters from across Europe on behalf of the European Flood Alert System. Working in small groups to simulate operational conditions, forecasters engaged in a series of carefully designed forecasting exercises using various different combinations of actual data from real events. Focus group data was supplemented by a follow-up questionnaire survey exploring how flood forecasters understand risk, uncertainty, and error. Results suggest that flood forecasters may not instinctively use ensemble predictions in the way that promoters of EPS perhaps think they should. The paper concludes by exploring the implications of these divergent 'epistemic cultures' for efforts to apply ensemble prediction techniques developed in the context of weather forecasting to the rather different one of flood forecasting.

Keywords: Epistemic cultures; European Flood Alert System; Ensemble forecasting; Risk communication

1. Introduction

Flood forecasting is an important mechanism for reducing the damaging effects of flood events. In response to the blistering criticism it received for its handling of the Easter 1998 floods in England and Wales (e.g. Bye and Horner, 1998), the Environment Agency, the official government body responsible, has completely overhauled its flood forecasting and warning systems, while at the European level, the European Commission has generously supported a number of Framework research programmes to improve flood forecasting capacity. Similar developments are taking place in the United States. The strategic plan for the US National Weather Service (2001) calls for substantial investment in developing an Advanced Hydrologic Prediction System to improve its flood forecasting capacity. Despite now well-established problems with communicating and making effective use of such forecasts in actual emergencies (Parker and Handmer, 1998; Pielke, 1999), the bulk of the research and funding continues to be devoted to the largely technical aims of reducing uncertainty and improving the predictive skill of forecast models.

One of the most promising avenues of scientific research is the development of so-called ensemble prediction systems (EPS). Ensemble prediction methods were developed by meteorologists to overcome the problem of deterministic weather forecasting in the face of uncertainty and chaos (Leith, 1974, Ehrendorfer, 1997). In contrast to

---

[*] Corresponding author.
E-mail address: david.demeritt@kcl.ac.uk (D. Demeritt).
[1] European Centre for Medium Range Weather Forecasting, Shinfield Pack Reading, RG2 9AX, UK

1747-7891/$ - see front matter © 2007 Published by Elsevier Ltd.
doi:10.1016/j.envhaz.2007.05.001

traditional deterministic modelling techniques that produce a single forecast, an EPS generates a suite, or ensemble, of predictions. Depending on the precise design of the system, ensemble members may vary in their assumptions about initial conditions, boundary conditions, model parameterization, model structure, or some combination thereof so as to reflect the uncertainty about them (Buizza et al., 2005). The computational demands of preparing an ensemble of differently perturbed model runs for each individual forecast usually mean that EPS operate at a coarser grid scale that resolves less physical detail than comparable deterministic modelling set-ups. Nevertheless, EPS often exhibits greater skill at forecasting meteorological phenomena than deterministic forecasting models. For example, using an EPS, the European Centre for Medium-Range Weather Forecasting (ECMWF) has been able to extend the time horizon over which it can forecast precipitation from 48 h to up 4–6 days in summer and a further 3 days in winter, when rainfall is easier to forecast skillfully (Buizza et al., 1999; Lalaurette & van der Grijn, 2003).

Moreover, generating a suite of forecasts, rather than a single deterministic forecast, also provides a way to quantify and thereby communicate the uncertainty about them. Unlike a deterministic forecast, an ensemble of predictions can be converted into a probability distribution function. Working with probability distribution functions requires some technical sophistication, not least to understand the assumptions involved in converting a discrete number of ensemble members into a full probability distribution, but they provide a much more explicit and precise way of inferring the likelihood of any given future outcome than qualitative descriptions, such as 'chance of late afternoon thundershowers', which do not specify exactly what that chance is. However, past research has also documented how much even experts, including meteorologists (Murphy et al., 1980; De Elia and Laprise, 2005), sometimes struggle to understand probabilities and probabilistic forecasts correctly (Tversky and Kahneman, 1974).

This ability to quantify forecasting uncertainty is particularly welcomed by sophisticated forecast users with well developed cost-loss functions (Palmer, 2002). For instance, short-term energy demand for heating and air conditioning is sensitive to temperature, and sudden surges in electricity demand represent a significant financial and operational risk for electricity generating companies. Armed with probabilistic forecasts, such firms can optimize their operations by buying or selling energy futures so as to balance the costs of having excess generation capacity against losses potentially avoided from being caught short and having to buy in at short notice expensive capacity from elsewhere (Altalo and Smith, 2004; Taylor and Buizza, 2003). For these two reasons EPS are now widely regarded by meteorologists as the state-of-the-art technique for weather forecasting (Gneiting and Raftery, 2005). EPS are now in daily operational use by national weather services around the world including Canada, the United States, and Australia and in Europe by ECMWF as well as the UK Met Office, Meteo France, and Deutscher Wetterdienst (DWD). For example, the UK Met Office now provides, as part of the National Severe Weather Warning Service (Hymas, 1993), a First-Guess Early Warning, estimating probabilities of severe weather 'objectively' from the ECMWF EPS. Its aim is to make forecasters more confident about issuing warnings more frequently and earlier than they otherwise might if reliant solely upon deterministic forecasts whose uncertainty was not quantified (Legg and Mylne, 2004).

Building on the established success of ensemble weather forecasting, there are now several initiatives to promote the application of such meteorological techniques to hydrological modelling, and in particular to real-time flood forecasting. At present, the vast majority of flood forecasting systems are deterministic in design and input, but this is changing fast. In the United States, for example, the National Weather Service launched its Advanced Hydrological Prediction Services programme in 1997 with the explicit aim of developing an ensemble flood forecasting capacity. Its latest plan calls for an EPS for forecasting floods to be fully operational by 2013 (US National Weather Service, 2001). Likewise in Europe, the European Commission (2002) is funding scientists at its Joint Research Centre (JRC) to experiment with the incorporation of such ensemble techniques into the EU-wide European Flood Alert System (EFAS). These developments are being facilitated through the World Meteorological Organization sponsored Hydrological Ensemble Prediction EXperiment (HEPEX, 2006). Its goal is to "bring together the international hydrological and meteorological communities … to produce reliable hydrological ensemble forecasts that can be used with confidence by the emergency management and water resources sectors to make decisions that have important consequences for economy, public health and safety" (Franz et al., 2005).

As the HEPEX manifesto implies, there are important differences between the meteorological and hydrological science communities. Whereas the use of ensembles in operational flood forecasting is still in its infancy, such techniques are widely used and even more widely accepted by weather forecasters. The reasons for this differential uptake of EPS by hydrologists are complex and reflect important philosophical and institutional differences in disciplinary approach to physically based modelling (Demeritt, 2001; Pappenberger and Beven, 2006). Although some cognate techniques, such as Monte Carlo sampling and updating algorithms, have long been used in flood forecasting (Lees et al., 1994), EPS were first developed by meteorologists during the early 1990s. Thus it should be no surprise both that ensemble forecasting techniques are better developed in meteorology and that meteorologists are more familiar with how best to extract, use, and communicate the important information conveyed by an EPS than hydrologists, who have not enjoyed the same wealth of data and computational power necessary to

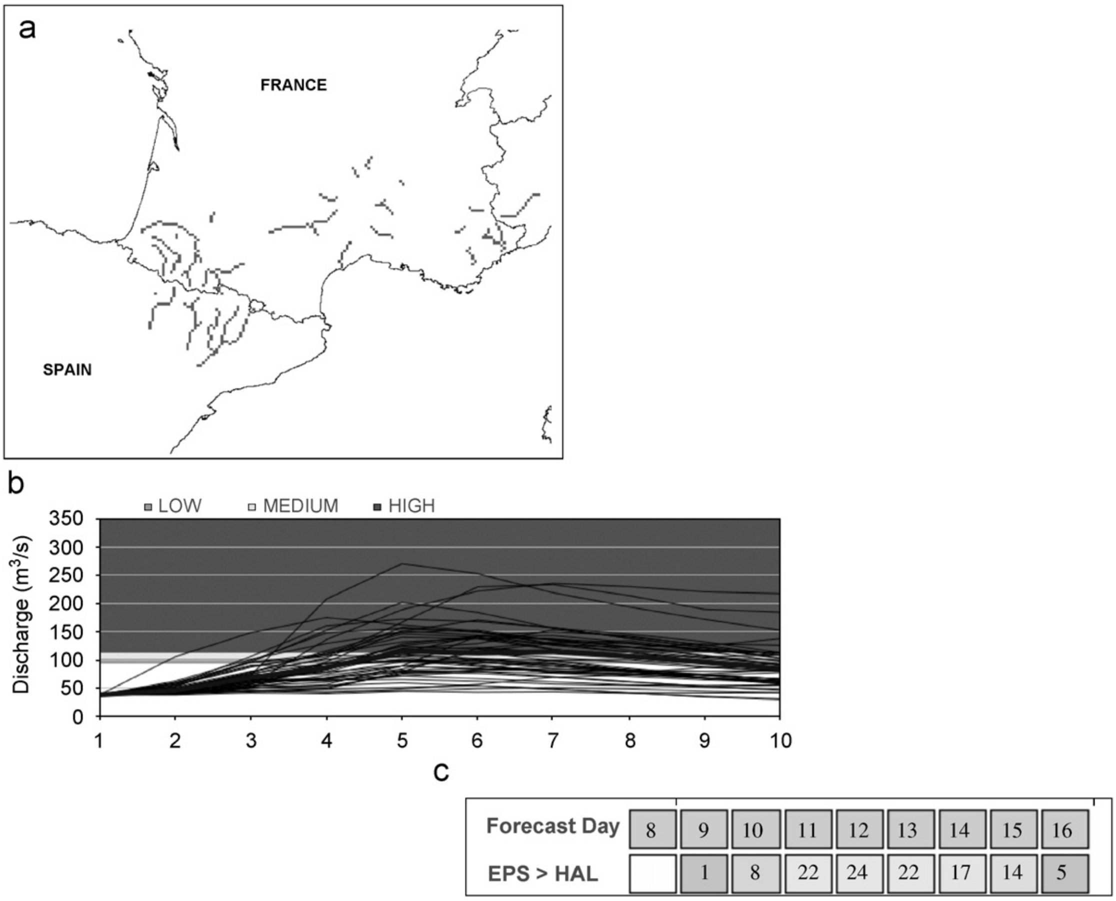

Fig. 1. Illustration of a possible EPS flood forecast representation as used in EFAS. (a) Threshold map: river pixels that are over High Alert Level (HAL) are shaded. (b) Spaghetti plot of ensemble discharges for a chosen point. (c) Ensemble flood alert summary: boxed numbers reflect the number of EPS members that forecast discharge values that exceed the HAL.

develop and validate EPS at a scale fine enough to be useful for flood forecasting.

Whatever the precise reasons for it, this widespread unfamiliarity with EPS poses considerable challenges for system designers, given the range of possible ways of communicating the results from an EPS, such as spaghetti plots, threshold maps, and summary tables (Fig. 1). As the NRC (2006, pp. 55–56) recently noted, ''[i]n planning these services [i.e. EPS], it is critical to ensure that the products and modes of presentation… are communicating what is intended''. However ensemble flood forecasts are to be communicated, there are also questions about whether and how useful hydrologists actually find such probabilistic information for what in operational terms is typically a binary decision (or set of decisions) about whether or not to issue a flood warning. Recent studies by Rayner et al. (2005) and by Morss et al (2005) have both documented significant cultural and institutional constraints to water resource managers making the best use of innovative decision-support technologies. While one of the attractions of ensemble forecasting is its potential to provide an ''early warning'' of possible flood events, there is an inevitable trade-off between false alarms and false negatives, that is unforeseen flood events. Hammond (1996) calls this the duality of error: for any given level of forecasting accuracy, increasing the number of early warnings will inevitably also increase the number of false alarms. Promoters of probabilistic predictions and ensemble forecasting have often assumed that forecast users share such a bias towards early warnings and false negatives, but the operational contexts of flood forecasting may well result in different responses to the duality of error.

To explore these wider questions about the perceptions of risk, uncertainty, and error in flood forecasting and their implications for the use of EPS, we here draw from material generated during a series of forecasting exercises and focus group discussions with operational flood forecasters from across Europe. The exercises were originally conducted as part of a 2-day workshop held by the JRC to inform its experimental development of EFAS. That workshop is the subject of a separate report (Thielen et al., 2005), which details its purposes, design, and the specific feedback obtained from EFAS users about the design and potential operational uses of the prototype system, the effectiveness of its proposed protocols for and methods of communicating medium-range flood forecasts to cooperating national authorities, and the provision of training for those users. Rather than focusing narrowly on

EFAS per se, this paper draws selectively from the data generated during that workshop to explore the various understandings of flood forecasting, EPS, and of risk, uncertainty, and error expressed during the workshop. In describing those perceptions, our purpose here is not the prescriptive one of spelling out the correct way to understand those contested concepts or use EPS—that would be a tall order indeed given the considerable differences of opinion among the six authors contributing to this paper. Rather, our purpose here is to document those perceptions and reflect on their implications for the practice of flood forecasting and the use of EPS in risk communication.

After describing our data and methods, we describe how participating forecasters made sense of the EPS in the context of what might be called, after Knorr-Cetina (1999), the 'epistemic culture' of flood forecasting. Then we describe how they used EPS to make forecasts and how, in turn, their forecasting decisions are influenced by concerns about wider public responses to them. The paper concludes with some reflections on the contrasts between the epistemic cultures of flood forecasting and meteorology and their implications for efforts to apply ensemble prediction techniques developed in the context of weather forecasting to the rather different one of flood forecasting.

2. Research design

The data for this study were collected in November 2005 during a workshop hosted by the EFAS development team at the JRC in Ispra, Italy. As we noted above, the workshop was designed to provide the JRC team responsible for the experimental development of EFAS with some very specific feedback about the proposed content of EFAS alerts and the protocols for communicating them to cooperating national water authorities. However, the discussions also ranged much more broadly, and in this paper we draw on the various data generated at the workshop to explore some more general issues about ensemble forecasting and perceptions of risk, uncertainty, and error.

In addition to the research team, the workshop was attended by eleven hydrological forecasters representing nine hydrological organizations in eight EU-member states with operational responsibility for forecasting on the Rivers Ebro, Po, Loire, Rhone, Garonne, Seine, Meuse, Rhine, Elbe, Oder, Danube, Sava, and Drava. This sample was selected both to represent something of the institutional diversity of flood forecasting in the European Union and the variety of hydrological regimes with which forecasters cope, from the major continental river systems of central Europe to the relatively more responsive and 'flashy' catchments typical of both mountain regions and dry Mediterranean climates.

Participating forecasters varied widely in their background and prior experience. Though all our informants were actively involved in producing real-time flood forecasts of some sort, the types of forecasts they issued (i.e. rainfall, river discharge, river level), and the kinds of models and modelling set-ups they were accustomed to using (i.e. meteorological, hydraulic, hydrological modelling; model dimensions such as 1D, 2D; model type such as mechanistic or data based) were diverse, reflecting the range of different disciplines, from meteorology and engineering to physical geography, involved in flood forecasting. While one participant reported having been working in operational flood forecasting for less than 2 years, the three most experienced had been doing it for more than 15, and the modal answer to our survey question about prior experience of real-time flood forecasting was 2–5 years. The majority (64%) had prior experience preparing flood alerts to be sent to civil authorities, but fewer (45%) had any experience communicating a flood forecast or warning to the media. All but one of our participants came from hydrological organizations that had previously signed formal memoranda of understanding with the JRC to receive the experimental ensemble flood forecast alerts being developed there as part of EFAS, but only four participants had any actual personal experience using ensemble forecast products (whether from EFAS or direct from weather forecasting agencies such as DWD and ECMWF) in the preparation of their own real-time forecasts.

At the workshop, the 11 participants were divided into four small groups of 2–4 people each, so as to spread forecasters by national origin and recreate the approximate size of actual operational flood forecasting teams. Groups were asked to complete a series of simulated forecasting exercises based on information about three case studies of real events that occurred in the summer of 2005 in the Elbe and Danube basins. The case study events were selected to illustrate both the potential and the problems of medium range ensemble forecasting. Whereas one case study was a 'surprise' event, in which, compared to conventional deterministic methods, probabilistic flood forecasting techniques might have provided some advanced notice, another was a false alarm in which the eventual flooding event was much less severe than the EPS indicated was possible. The third was typical of the many flooding events that devastated central Europe that summer, but in this case study all groups (with the exception of the control group—see below) were given ensemble but not deterministic precipitation forecasts. For each case study, all the groups were provided with:

- Maps of the river basins in question showing topography, the river network, and the location of the discharge gauging stations;
- Information about the meteorological situation, including recent precipitation measurements, synoptic charts, and radar, as well as 7-day deterministic precipitation forecasts from DWD and 10-day deterministic and EPS forecasts from ECMWF;
- Observed discharges at selected discharge gauging stations for the previous 7 days;

- Forecasted river discharges at each gauging station over the next 2 days from a nominally "local" deterministic hydrological model;
- Prototype EFAS flood alerts reports, providing information, in several different forms (see Fig. 1, above), about the results of its deterministic flood forecasting (using the DWD and ECMWF deterministic forecasts) and its EPS-based (using the ECMWF ensemble) flood forecasting.

Furthermore, for each case study, a different small group was designated as a 'control', which did not receive the various EPS-based precipitation and flood forecast products provided to the other three small groups.

With the assistance of a facilitator to keep them to task and guide them through the information provided, each small group was asked to deliberate on the hydrological situation, to write a brief forecast bulletin, such as they might regularly do for their own hydrological services, and decide whether or not to issue a "low", "medium", or "severe" warning to the appropriate civil protection authorities. It was up to the participants to define what low, medium, and severe meant. This task was then repeated at two further daily time steps, each backed by updated forecasts and discharge data for 12:00 and 00:00 of that day. Discussions were conducted in English, recorded, and subsequently transcribed and coded to identify emergent themes. The facilitators were also charged with observing how the groups used the information provided to them, which provided another source of data.

Each group had 2 hours to complete each of the three case studies. As we discuss in more detail below, this time pressure, combined with participants' unfamiliarity with both the catchments and the EPS and other data being applied to them, is not typical of actual operational conditions. As with all experiments, ours is an artificial one, and this leads to an inevitable analytical challenge in determining the extent to which the responses to EPS we observed simply reflect the novelty of the experimental environment rather than responses likely to occur in operational contexts.

In order to reinforce the inferences we draw on the basis of observing the ways forecasters responded to EPS during the small-group exercises, this paper also draws on two other sets of data, which we use to 'triangulate' and thereby cross-check the results of our small-group exercises. First, we conducted two one-hour plenary debriefings with all workshop participants to discuss their experiences of the workshop exercises and of working with the ensemble forecast products. These plenary sessions, which were recorded, transcribed, and coded in the same way as the small group exercise discussions, provided a valuable opportunity to check our initial impressions of how participants made sense of the forecasting exercises with participants themselves, and thereby confirm and enrich our understanding of their understandings. Second, participants also completed a questionnaire survey asking them a variety of basic information about their personal background and institutional experience as well as 50 closed and 8 open-ended questions about their attitudes to flood modelling, uncertainty, and hydrological science. Responses to these questions provided another method for empirically accessing the understandings of our informants and thereby assuring the validity and robustness of our conclusions about them. While their survey responses are largely consistent with the views expressed in the workshop exercises and plenary discussions, there are some interesting contradictions that we explore in more detail below.

3. Making sense of EPS

During the small-group exercises, participating forecasters initially struggled to make sense of the EPS and integrate it with the other information provided so as to make their forecasts in the time allowed. Although the groups were composed of strangers, working together in what was, for most, a second language, their shared scientific background and professional experience, along with a common interest in learning more about EPS, provided the basis for group solidarity and effective working arrangements. It took very little time for the groups to apply themselves to completing the tasks assigned to them. They found the exercises challenging to complete, as this exchange from the plenary discussions reveals:

> forecaster 6: we spent a lot of time on the first case … because we didn't know the material, how to the maps, how to start.
> forecaster 3: It was the same for us. We didn't know material, the graphs, the material.

Despite, or indeed perhaps because of, the wealth of new information at their disposal, the groups tended to focus on just a small subset of what was available about each case study. Doswell (2004, p. 1117) has compared weather forecasting to trying to drink from a fire hose: "the torrent of data and products derived from the data threatens to overwhelm" the forecaster. In the exercises groups typically ignored the synoptic meteorological maps, summary climatological information, rain gauge data, and radar, which in a more familiar operational context they might be expected to have used more to help them assess both how much previously forecasted rain had actually fallen and where. Instead, after spending the first 5 or 10 minutes of each case study poring over the map of the river network, trying to familiarize themselves with the area they were dealing with, groups focused the vast majority of their time and attention on just two kinds of information: first, the latest precipitation forecasts, especially, when it was available, the ensembles, and second, the hydrographs showing the latest discharge measurements and a deterministic forecast of discharges for the next 48 hours, as well as, when it was available, the EFAS ensemble discharge outputs. The novelty and comparative unfamiliarity of the various ensemble precipitation and discharge forecast outputs meant that they occupied a disproportionate amount of the time of those groups who had access to them. But the control groups that were denied access to the ensembles were little different: at each time step they too focused predominantly on the latest precipitation forecasts and discharge information.

Over the course of each case study, this focus became even more pronounced. By the last time step of each case study, most groups were looking at nothing else. In part this was an artefact of the exercise design. As one participant explained "normally [at our forecasting center] we have more [computer] screens than people", whereas during the workshop the groups had paper print outs of the precipitation forecasts and discharge information, and with the assistance of the facilitator could look at a single computer with a CD containing weather forecast animations, radar, and other additional data. The result, as that same flood forecaster went on to explain, was an inevitable narrowing of focus:

> Of course I think we don't use most of the actual information from the CD because … if you must ask, 'show me this, this, and this' and you have a lot of different informations, it's not enough time 1 h to prepare the forecast…In the normal situation, every time we have some basic information which we have every day and we are sure that the information is correct everyday. And we check this information the first time. Next we look for additional information [murmurs of agreement around room]. Now in the practice there all was additional information, because all was new, new forms, new system, and we spent a lot of time, looking at [it] simply for information, and local discharge forecast.

This informant's emphasis on the novelty of information, like the groups' tendency to focus most of their attention on the latest forecasts and discharge information, reflects what psychologists call the 'added information bias' (Nichols, 1999). This bias can lead forecasters to overemphasize the significance of new information relative to what is already available or to assume that additional information will improve performance when in practice it may only generate confusion (Slovic et al, 1977). In the specific case of weather forecasting, Stewart et al. (1992) found that additional information, particularly if unfamiliar, did not necessarily improve forecasting skill because of the difficulties of assimilating it properly.

This tendency to emphasize the value of new information was compounded by the fact that during the exercises groups were making flood forecasts for unfamiliar catchments. In the absence of much prior experience either with the catchments themselves or the models used in the forecasting exercise, groups relied heavily on the new information provided at each time step in order to assess past forecasts and decide how much confidence to place in future ones. This is where EPS can play an important role. Unlike deterministic forecasts, EPS makes the uncertainties associated with its forecasts more explicit and thereby reduces reliance on what Michael Polanyi (1967) famously called the scientist's 'tacit ways of knowing' based on intuition, experience, and craft skill.

Participants recognized this promise. They welcomed the capacity of EPS to supplement the forecaster's own expert judgment with a more formal, quantitative representation of forecast uncertainty. Surveyed about the advantages for their organization of using EPS, they offered replies such as "to know the uncertainty of the forecast" and "better uncertainty estimation". In the exercises, the groups certainly appreciated having the ensembles. Indeed, having had access to them once, groups often experienced serving subsequently as a 'control group' denied access to them as a loss. As one person recalled, "we were last as the control group and it was a great confusion for us [to] work without EPS". But participants also insisted that local knowledge and personal experience were still required to make sense of the EPS properly. Recalling the second case study, forecaster 1 commented that the ensemble precipitation forecasts:

> covered in fact almost the whole climatic range, from medium or very low up to high or extreme range. It means that the distribution of the medium term forecast covers the whole climatic range. Which means we don't know much more than the climate, although we are not looking that far ahead in time, but one week or a couple of days.

In this context, where the EPS is indicating a wide range of possible future system states, it is very difficult to make a forecast with much confidence. Even simple summary statistics, like the ensemble mean, may not be reliable aids in interpreting EPS outputs, because there is a positive correlation between the error of the ensemble mean and the ensemble spread (Buizza et al., 2005). That is, the wider the spread, the greater the error of the ensemble mean, because the uncertainty about outlier events is greater than that about the more typical system states that the model has been tuned to reproduce (Lalaurette and van der Grijn, 2003). This represents a particular problem for flood forecasters as it is precisely those extreme events that they are most concerned with forecasting accurately. Given how EPS has been promoted as helping forecasters to do just that (e.g. Legg and Mylne, 2004; US National Weather Service, 2001; NRC, 2006), it is thus somewhat ironic that EPS is much less effective at predicting extremes than at forecasting more 'normal' system states.

Faced with the wide range of possible future outcomes often shown by the EPS during our exercises, forecasters tended to lean even more heavily on their personal experience in deciding what forecast to issue. As one of the group facilitators observed during the plenary discussion:

> most of you obviously relied far more on your experience of the catchment than anything else. …

> You just know obviously that your experience is exceeding the model capability in that sense far more.

Such personal knowledge and experience are important because of the considerable technical difficulties of accurately modelling the frequency of extreme events. However, decision research has consistently found that this kind of intuitive reasoning can be unreliable (Gilovich et al., 2002; Tversky and Kahneman, 1974). Compared to objective forecasting methods, intuitive forecasting produces a higher average error, though the range of those errors may be narrower than the very large, if rare, errors that result from a model being applied outside the range of conditions to which it has been tuned (Doswell, 2004; Stewart, 2000). Our participants themselves sometimes reflected on the limitations of their own intuitions. Having failed to forecast a flood event in the first case study, forecaster 9 commented:

> My problem is that I am more expert for the big catchments. I'm dealing with the Rhine itself, the Danube itself, and something like that. So, um, I'm not used to this small catchment, and here at the Danube [pointing to the main branch of the river on the map] there is no problem at all. Until now…

The experience was not enough to shake this forecaster's faith in the value of local knowledge compared to formal models, such as EPS. Rather, she insisted, the problem was that she did not have the right kind of local knowledge.

The importance of such personal knowledge was something of an article of faith in the 'epistemic culture' of flood forecasting. During the case studies, participants repeatedly emphasized its importance to good flood forecasting. "Hydrologists need to know their rivers", said forecaster 8 to murmurs of assent from other members of his small group. Something very similar was said in another small group, where forecaster 9 commented, "Real forecasting centres know their catchment very well, each tributary". To peals of laughter from the plenary, forecaster 2 made light of the difficulties caused by his group's ignorance of the Rivers Sava and Drava on the upper Danube, which were the focus of the second case study, designed to see how participants would respond to a false alarm:

> Well we did evacuate [laughter throughout room]. Because we had the impression that this area responded quite quickly, especially the upper parts. We also got a little confused with these words Sava and Drava, especially when we also used the French ça va [chortles in background]. So we talked about the Sava ça va [laughter throughout room].

The crescendo of laughter from the plenary at hearing this speaker confess to having issued an evacuation order when none had actually been necessary was not simply a response to this dose of wit at the end of a long and tiring day of work. It was also an expression of solidarity with the difficulties encountered by the speaker in having to make forecasts without such local knowledge.

The local knowledge possessed by good flood forecasters was not just an empirical knowledge of particular local catchments and their physical geography. It was also a knowledge of the digital worlds of the computer modelling set-ups used to forecast floods. There are at least two ways in which this too was a local knowledge of sorts. First, validating these models involves tuning model parameters so as to produce the best fit (however defined) with locally available data. This process of parameterization means that even the same basic modelling platform may well perform quite differently when used in different places. This implied 'uniqueness of place' is widely recognized in the epistemic culture of rainfall-runoff modelling (e.g. Beven, 2006). Thus, almost all of our participants (92%, with none disagreeing) agreed or strongly agreed with the statement, "it is important for forecasters to understand the workings of the models they use". This emphasis on the personal knowledge of the forecaster points to a second way in which forecasters' knowledge of the virtual world of computer set-ups is a local one. During one case study, forecaster 9 explained why:

> I know my model much more than this model here, and so I know in which parts I can believe the model and in which parts I have to be um. Maybe I know this is not so good in this part. And I have no idea about this here.

What this quotation is gesturing at is the way in which through experience forecasters develop a craft based 'feel' for their own models that allows them to know more about them than they can even say or would otherwise be immediately apparent to a less experienced outsider looking simply at their formal diagnostic statistics. Flyvbjerg (2001) calls this sort of holistic, practical knowledge 'phronesis', and he argues that it is what distinguishes the performance of the novice who must stick mechanically to the rules from that of the true expert who knows when to improvise and think outside of the envelope of convention.

Initially our flood forecasters placed great confidence in the nominally 'local' deterministic discharge model used in the case studies. However, they soon became much more critical of it. After seeing, in the first case study, the 'local' discharge forecasts miss the subsequently observed values by over $30\,m^3\,s^{-1}$, which could mean, in some places, a difference in water level of around half a metre, or the difference between bankfull conditions and houses flooded, two groups independently concluded that its forecasts were not very skilful. As one group explained

> forecaster 9: the hydrological forecast is very bad [laughter]
> forecaster 10: yeah. Really really bad [more laughter from both] forecaster 9: and maybe um, it's still bad [inaudible]

Participants felt more than able to make judgments, sometimes quite critical, about the quality of the discharge model they were using. By contrast, they were much less discriminating users of the EPS, despite their insistence, both in discussion and in their survey responses, on the importance of having intimate feel for the forecasting models they use. In part, their curiosity was deterred by the design of the exercise. Groups were only provided with printed maps of the forecasted 10 day accumulated rainfall as well as a sheet of postage stamp sized maps showing the rainfall forecasted to fall in Europe during the 10 day forecast period by each of the 51 ECMWF ensemble members, rather than the more spatially detailed, daily forecasts they sometimes asked for. It was possible for them to ask the facilitator to access the CD so as to zoom in to some portion of any one of those individual ensemble members, but hardly any groups requested this additional information, both because the process of accessing it was cumbersome and because it was the property of the ensemble as a whole, rather than of any one of its individual members, that they were seeking to understand better.

In this way, the very character of the EPS also served to deter close scrutiny. Unlike a deterministic discharge forecast that could be immediately compared against measured discharge, it was hard to judge the skill of the EPS simply by looking. In keeping with their emphasis on the local knowledge and craft skill of good flood forecasting, our participants believed that in order to use the EPS properly they would need regular access to it so that they could develop an intuitive feel for what it was telling them:

> forecaster 2: Because you could in principle, make it everyday. Then people get used to this information. To see it everyday, well what does the ensemble say. And then at certain times, they all go red and it's an alert let's say.
> facilitator: would you consider that useful? Even if there is no high alert reached, for example, would you consider it for daily forecasting? Would like to have them daily?
> forecaster 3: I think I would like to learn more about how to use the information in the right way.
> forecaster 11: Like a training, a training, so you can
> forecaster 3 [interrupting]: in this way if you go like this, you can see each day if they are repeating the same forecast and ah…
> facilitator: but in principle we are not really allowed to do that. We would have to check that with ECMWF, if they would allow us to do that on a daily basis, because in principle we agreed only during flood events.[2]

Notwithstanding these limitations, it was surprising how uncritically the EPS was received. At least some participants were aware of technical debates about EPS. In response to an open-ended survey question about their potential disadvantages, one participant referred to the contentious assumptions they involve:

> EPS starts from the point that taking a neighbourhood of initial conditions in meteorological model can resolve the uncertainty of the meteorological forecast, but the 'real' IC [initial conditions] may be far away [from] the used IC and its neighbours

But that kind of critical scepticism was very much the exception during the workshop. Instead, participants tended to take the EPS results they were given at face value, though further research would be required to confirm whether that were also true in actual operational contexts. Their confidence is reminiscent of what sociologist of science Donald Mackenzie (1990, p. 371) has called the 'certainty trough' (Fig. 2). Mackenzie suggests that his schematic diagram represents "the distribution of certainty about any established technology". In the case of climate models, Shackley and Wynne (1995) argue that policy-makers and other model users may place undue confidence in them because of their failure to appreciate the tacit judgments and uncertainties involved in their construction. Likewise, no one questioned the definition of the 'high and severe' alert thresholds from the EFAS hydrological model, despite the fact that those thresholds are actually based on uncalibrated model runs, rather than the "measured data" that 82% of respondents agreed or strongly agreed should be the basis "for issuing flood alerts".[3] Thus there is at least a danger that hydrologists may misunderstand and place inappropriate levels of confidence in the formal representations of uncertainty provided by an unknown EPS-based flood forecasting system.

4. Flood forecasting with EPS

Although the groups spent most of their time during the group exercises poring over the EPS, they were not, in fact, central to the groups' decision-making process. Instead, it was the deterministic discharge forecasts that was the key. How the groups responded to the EPS depended on whether or not it seemed to confirm the deterministic forecast. As one of the moderators recalled:

---

[2] As the facilitator's comments indicate, the licensing agreement under which ECMWF provides its EPS to EFAS forbids EFAS from providing its end users with direct access to the precipitation ensembles it uses to drive its own discharge forecasting model. Instead, EFAS end users can only receive this information, which both forecasters themselves and a

(*footnote continued*)
recent report from the US NRC (2006) both insist is crucial in building forecaster skill and confidence in EPS, in the event of an actual flood alert.

[3] The EFAS thresholds are calculated from simulations of a 14-year series of observed rainfall. If such an uncalibrated model is used it is assumed that the model, which is an approximation of reality, reproduces the non-linearity of nature, which is not the case for most hydrological models (Beven, 2006). Also it is questionable whether the error structure of the forecast and the interpolated observed field are similar enough to allow such an estimation (see Pappenberger and Beven, 2006).

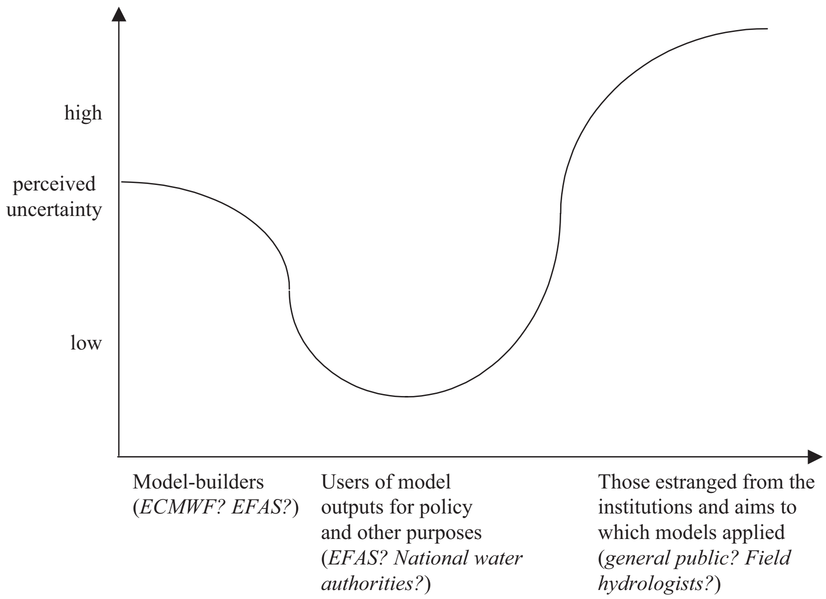

Fig. 2. The 'certainty trough' in the relative confidence accorded to forecasting models, such as EPS, by model builders, policy users of model outputs, and the general public. After Mackenzie (1990) and Shackley and Wynne (1995).

> they were looking at the EPS a lot. But when I asked at the end, I had the impression they were doing a decision on the EPS, but then they said, 'Yeah, but this is just to support a little bit more our deterministic forecast', even though they spent maybe 70% of their time looking at the wonderful spaghetti plots.

Cognitive psychologists call this an 'anchoring bias' (Nichols, 1999), whereby a single piece of information provides the central reference point, or anchor, around which the perception of other sources of information is then adjusted (Tversky and Kahneman, 1974). In the small-group exercises, the anchor point was the deterministic discharge model forecast, which the groups then sought to confirm by referring to the EPS. In the face of uncertainty about their models and model predictions, flood forecasters warmly welcomed this use for the EPS, particularly when it confirmed the results of their deterministic model:

> The most important information for us was the EPS. Because it was very comfortable to say 'yes we are sure because the most of the EPS forecasts say the same information' … We are not afraid to say this (forecaster 5)

The vast majority (72%, with none disagreeing) of participants agreed or strongly agreed that "best way to use ensembles is to confirm the results of our deterministic flood model".

There are two potential problems with this use of EPS in flood forecasting. First it is prone to a 'confirmation bias' whereby decision makers tend to overemphasize information that confirms some prior belief, in our case of flood forecasting, the deterministic forecast (Nichols, 1999). This tendency was clearly evident in the way that our flood forecasters responded to the EPS. Participants repeatedly explained how seeing the EPS "helps us to be sure" (forecasters, 10 and 5). Forecasters were always happy to see the EPS broadly agree with the deterministic forecast. Not only did it give them more confidence in their deterministic forecast, it also gave them licence to stop looking for other information, which might undermine that confidence.

This then points to the second problem with the way our informants used EPS. Despite their hopes that it might be used "to confirm deterministic forecasting", they found that in practice, such confirmation, in the form of an ensemble tightly clustered around the value predicted by a deterministic forecast, was rarely available when it was actually needed most. In situations of high uncertainty when forecasters most wanted some way to confirm their deterministic forecast, the spread of the ensemble members was likely to be widest (Buizza et al., 2005). This should come as no surprise, because displaying the uncertainty surrounding a forecast is precisely what EPS is designed to do.

However, our flood forecasters were not used to seeing uncertainty displayed in this way, and they found it difficult to know how to interpret the EPS. Promoters of EPS often insist that compared to traditional deterministic forecasting, the additional information provided by EPS should help to improve the quality of forecast decision making (i.e. AMS (American Meteorological Society), 2002; Palmer, 2002; Legg and Mylne, 2004). To some extent this view was endorsed by participants themselves. All but one (92%) of our participants agreed that ensembles "provide flood forecasters with valuable information", but they were less clear about exactly what that information was (more skillful forecast? early warning? fuller representation of uncertainty?) or how useful it might be for their operational purposes. By the end of the exercise just over half (52%) were still not sure if "existing EPS are more skilful than the deterministic models we use for rainfall forecasting".

During the case studies, participants often found that rather than helping them to confirm the deterministic discharge forecasts they were used to working with, that in practice, the EPS just confused them:

> forecaster 9: if the EPS shows the same direction like the other forecast, it makes us more confident
>
> facilitator: but if it doesn't?
>
> forecaster 9: much more confused

Dealing with such discrepancies was a particular challenge in the first case study, in which, as forecaster 1 recalled, "the ECMWF forecasts were rather different [from the DWD]". One of our groups felt confident in discriminating between the two: "as this region is mountainous and orographic precipitation is significant we tend to believe more, to rely more on the German [forecast]", which they knew to have a smaller grid size and thus be more likely to resolve for the local effects of topography on rainfall distribution. Other groups, however, were not able to call upon this level of expertise about the EPS. In the absence either of such expertise or the local knowledge of the models and catchments favoured in the epistemic culture of flood forecasting, groups often felt as though they had no rational basis for making a forecast. During the plenary discussions, one moderator observed of the groups he worked with:

> at the end point it was belief [laughter around room]. It was something more religious [more laughter]. That you believe in the EPS when the deterministic says something similar but you don't believe when it says something else [laughter]. So it's something like this.

In this exchange, laughter helped to relieve the evident discomfort at the connection being made here between religious faith and science.

In making their forecasts, participating forecasters did not always welcome the additional information provided by the EPS, particularly when it called into question a decision already anchored around a deterministic forecast. Indeed, during the third case study several groups expressed relief at only having the ECMWF ensembles, rather than, as in previous case studies, both the ECMWF and the DWD forecasts:

> facilitator: in that case [i.e. scenario 3] we didn't give you the DWD forecast, only ECMWF forecast. Did you miss the DWD or were you happy not to have confusing other information?
>
> [laughter throughout room]
>
> forecaster 6: We are happy.
>
> forecaster 9: We are happy, because now we have the same signal. Maybe if the German weather forecast is showing the same signal we have been much more happy.

5. Issuing warnings

Flood forecasters understood the provision of early warnings of possible flood events as one of their most important professional responsibilities. In response to the survey questions, they consistently expressed a preference for early precautionary action. Most (64%) agreed with the statement, "Flood forecasters should risk issuing warnings that may prove to be false so as to provide as much advanced warning as possible about any possible flood event". Likewise only 18% agreed that, "Flood forecasters should not issue flood warnings until they are certain that there will be a flood". By providing flood forecasters with earlier indications of low probability flood events, EPS offers them the potential to fulfill this precautionary role even more effectively.

During the small-group exercises, however, flood forecasters consistently hesitated to issue early public warnings on that basis alone. As forecaster 1 explained to the moderator during the third case study:

> we decided that we shall issue warning for tributaries Vedea and Arges because floods are expected within short time. Although some danger is indicated and even high danger on Alt River, but it is more than four days ahead, which is rather uncertain so we shall keep quiet about [giggles from other group members]

In the face of uncertainty, they tended to adopt a wait and see approach, as this exchange shows:

> forecaster 5: The most strange situation was when we had two quite different deterministic forecasts. [murmurs of agreement]
> facilitator: And then what did you do?
> forecaster 5: I choose nothing [laughter all around room]
> forecaster 10: And wait for the next forecast

Thus, for flood forecasters erring on the side of caution tended to mean waiting to be sure, rather than issuing immediate warnings to the public.

In part this wait and see approach reflected the added information bias we have already discussed. Forecasters typically hoped that additional information would make them more certain about whether or not a flood event was likely to occur. But their tendency to wait for confirmation rather than risk issuing a potentially false alarm also reflects an underlying attitude to what Hammond (1996, p. 22) calls the "duality of error". This is the balance between two kinds of mistakes: predicting an event that does not actually occur (a false positive) and failing to predict correctly an event that does occur (a false negative). The frequency of these two kinds of error is inversely related. For any given level of prediction accuracy, the number of false negatives can only be reduced by

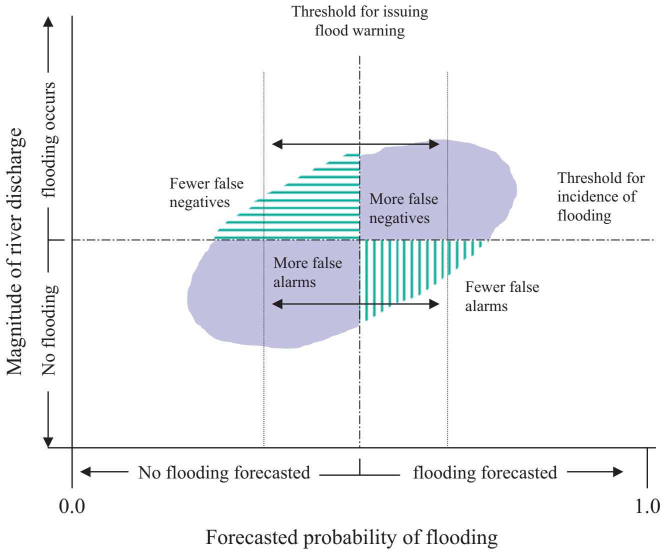

Fig. 3. The duality of error. For any given level of forecasting accuracy, any shift in the warning threshold to alter the number of false alarms (shown in lower right) will have the opposite effect on the number of false negatives (shown in upper left).

increasing the number of false positives, as shown in Fig. 3. By consistently waiting for additional to confirm any decision about issuing a flood warning, our flood forecasters were not only seeking to reduce the frequency of errors, represented in Fig. 3 by the elliptical shaped dispersion around a 45° line of perfect predictions. Flood forecasters were also, in effect, expressing a strong practical aversion to false positives very different to the preference for precautionary action they consistently stated in response to survey questions. In terms of Fig. 3, this aversion to false alarms would be represented by shifting the decision threshold for issuing flood warnings to the right.

This attitude to false alarms and forecasting error seems also very different to the prevailing attitude of meteorologists. As Doswell (2004, p. 1119) notes, in the epistemic culture of "weather forecasting, false negatives are seen as a less-desirable outcome than false positives ('false alarms'), because they are associated with the unfavourable notion of an unforecast weather event, perhaps with casualties as a result". By contrast, flood forecasters were very conscious of the possible costs—to the public and to their institutional credibility—of issuing early warnings that later proved to be false, as this exchange suggests:

> forecaster 9: For example, at Cologne there are some walls, 10 meters above the gauging station in Cologne. If you are making a forecast of 10 meter and 20 cm, um, you bring all the people on top of the roofs. So you have to be sensitive to the region.
>
> facilitator: How is the decision for when the forecast is for 10 m 20 cm different than when it is 9 m 20 cm or 9 m and 80 cm? Are you more or less likely to issue an early warning?
>
> forecaster 9: um. Maybe it's good to have an early warning for the more than 10 meters, but we are dealing with the short term forecasts and not only the red green and pink [i.e. alert levels] We are dealing [points to sheet] really with the meters. They want to know it in centimeters. Is this a 10 m flood? Or 9 m and 70 cm? It's not easy to say.

There was a feeling among participating flood forecasters that *their* end users would not tolerate mistakes. In their study of American water authorities, Rayner et al (2005) found a similarly institutionalized aversion to error, which was amplified by concerns about what Rothstein et al (2006) call 'institutional risk' to reputation and credibility. If forecasting mistakes were to be made, it was best to make them using tried and true methods, rather than be exposed to the risk of additional opprobrium for having departed from convention with some new experimental approach.

In turn those beliefs about wider public and institutional responses to warnings then influenced the decision of whether or not to issue a warning. While flood forecasters tended to hesitate in the face of countervailing evidence before issuing a warning, once a warning had been issued it tended to stay up longer than the evidence itself might otherwise indicate. As forecaster 1 explained, we continued to maintain our warning "to be on the side of safety and just knowing users' psychology... if you lower the alert level and then [have to] rise [it] again, it would be worse".

6. Conclusion

While there are limits to what can be concluded about the attitudes and behaviour of an entire profession from such a small sample working through a series of artificial simulation exercises, our findings do raise some questions about the uses and potential effectiveness of EPS in flood forecasting. One of the attractions of EPS is its capacity to provide an early warning of possible flood events. While supporters of the precautionary principle often point to the importance of early warnings for averting disasters (e.g. Harremoës et al., 2002), they sometimes fail to acknowledge the costs of such a pre-emptive risk communication strategy. In any forecasting system there is an inevitable trade-off between false alarms and false negatives, that is, unforecasted flood events. Promoters of ensemble flood forecasting have typically assumed that hydrologists share the same bias towards providing early warnings and bearing the associated costs of the resulting false alarms as the meteorological community (e.g. NRC, 2006). However, our results here suggest that the epistemic culture of flood forecasting may be based on some rather different attitudes to risk, uncertainty, and error. These attitudes have important implications for EFAS and the many other international efforts to apply ensemble prediction techniques initially developed in the context of weather forecasting to the rather different one of real-time flood forecasting.

While flood forecasters welcomed EPS, they did not use it in the precautionary way that promoters of ensemble forecasting often imagine they would or should. Despite the fact that EPS is not designed for this purpose, flood forecasters' first instinct was to use the EPS to confirm their own local deterministic models. When there was a discrepancy between the two, flood forecasters tended to adopt a 'wait and see' attitude during the simulation exercises rather than issuing early flood warnings whenever the EPS suggested a heightened possibility of future flooding. Future research will be required to confirm whether these same instincts hold in actual operational conditions, but it would seem that for flood forecasters playing it safe means waiting to be sure.

In part this response reflected concerns about the costs of 'crying wolf'. In addition to the perhaps understandable desire to protect their own institutional reputation and credibility, the reluctance of flood forecasters to issue early warnings also reflected important beliefs about the cognitive and other capacities of the groups to whom they issue flood warnings. During the forecasting exercises as well as in the subsequent debriefing discussions and the questionnaire surveys, flood forecasters repeatedly expressed the view that general publics, and even emergency service organizations, would not tolerate false alarms. Whereas promoters of EPS argue that probabilistic forecasts are more valuable than simple, deterministic ones because they allow recipients to optimise their exposure to risk through hedging or other behaviour to optimise their cost-loss function (i.e. AMS, 2002; Palmer, 2002), flood forecasters believed that their users either would not, or could not, respond in that kind of rational and utility-maximizing way to a flood warning. Rather than empowering forecast users with additional information, participating flood forecasters feared that even relatively 'expert' emergency service agencies would simply be confused by probabilistic forecasts and react either with panic or utter indifference. This finding suggests that while EPS offers operational forecasters with a useful in-house pre-alert function, their aversion to false alarms means that it is unlikely to provide the basis for the kind of early warning of flood events sometimes hoped for by its promoters.

Forecasters' tendency to wait and see was reinforced by several persistent cognitive biases structuring the way they understood and used EPS. Their expectations were initially anchored around the deterministic local discharge forecasts, which were then adjusted depending on whether or not the EPS seemed to confirm them. However, EPS is not designed to provide such confirmation. When the EPS and deterministic forecasts differed, as they often did, our flood forecasters tended to be confused and to wait for additional information to clarify the situation and confirm their local deterministic forecasts. This tendency to seek out information confirming their prior belief and to ignore or discount contradictory information reflects what psychologists call a confirmation bias. In an operational context of forecasters working with their own locally developed and trusted discharge models we might expect this confirmation bias to be even more prevalent than was in our experimental set-up. Such a confirmation bias will tend to limit the potential for EPS to sensitize forecasters to low-probability high consequence 'surprise' events.

In keeping with their emphasis on local knowledge and hands-on experience in the epistemic culture of flood forecasting, our respondents insisted that the best way to overcome these biases and develop skill in using EPS more effectively (and personal confidence in that skill) would be through working with it every day so as to develop an intuitive 'feel' for what it might be telling them. This desire has important implications for efforts to encourage the take-up of EPS by flood forecasters, given the licencing and intellectual property right restrictions on the dissemination of proprietary ensemble precipitation forecasts. For instance, under the terms of its licencing agreement, EFAS is only allowed to provide cooperating national river authorities with EPS rainfall forecasts from ECMWF in the event of an actual flood alert. Our research suggests that without the opportunity to work regularly with EPS that flood forecasters are unlikely to have either the confidence or the skill necessary to use it to its fullest potential.

References

Altalo, M.G., Smith, L.A., 2004. Using ensemble weather forecasts to manage utilities risk. Environmental Finance 20 (October), 8–9.

AMS (American Meteorological Society), 2002. Enhancing weather information with probability forecasts. Bulletin of the American Meteorological Society 83, 450–452 Last accessed 7 August 2006 from 〈http://www.ametsoc.org/POLICY/enhancingwxprob_final.html〉.

Beven, K.J., 2006. A manifesto for the equifinality thesis. Journal of Hydrology 320 (1–2), 18–36.

Buizza, R., Hollingsworth, A., Lalaurette, F., Ghelli, A., 1999. Probabilistic predictions of precipitation using the ECMWF ensemble prediction system. Weather and Forecasting 14, 168–189.

Buizza, R., Houtekamer, P.L., Toth, Z., Pellerin, G., Wei, M., Zhu, Y., 2005. A comparison of the ECMWF, MSC and NCEP global ensemble prediction systems. Monthly Weather Review 133, 1076–1097.

Bye, P., Horner, M., 1998. Easter 1998 Floods Report by the Independent Review Team to the Board of the Environment Agency, vol. 1. Environment Agency, Bristol.

Demeritt, D., 2001. The construction of global warming and the politics of science. Annals of the Association of American Geographers 91, 307–337.

Doswell III, C.A., 2004. Weather forecasting by humans—heuristics and decision making. Weather and Forecasting 19, 1115–1126.

Ehrendorfer, M., 1997. Predicting the uncertainty of numerical weather forecasts: a review. Meteorologische Zeitschrift, Neue Folge 6, 147–183.

Elia, R., Laprise, R., 2005. Diversity in interpretations of probability: implications for weather forecasting. Monthly Weather Review 133, 1129–1143.

European Commission, 2002. Communication from the Commission to the European Parliament and the Council: the European Community response to the flooding in Austria, Germany, and several applicant countries—a solidarity-based initiative. COM, Brussels, p. 481.

Flyvbjerg, B., 2001. Making Social Science Matter: Why Social Inquiry Fails and How It Can Succeed Again. Cambridge University Press, New York.

Franz, K., Ajami, N., Schaake, J., Buizza, R., 2005. Hydrological ensemble prediction focuses on reliable forecasts. Eos 86 (25), 21.

Gilovich, T., Griffin, D., Kahneman, D. (Eds.), 2002. Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge University Press, New York.

Gneiting, T., Raftery, A.E., 2005. Weather forecasting with ensemble methods. Science 310, 248–249.

Hammond, K.R., 1996. Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice. Oxford University Press, New York.

Harremoës, P., Gee, D., MacGarvin, M., Stirling, A., Keys, J., Wynne, B., Guedes Vaz, S. (Eds.), 2002. The Precautionary Principle in the 20th Century: Late Lessons from Early Warnings. Earthscan, London.

HEPEX, 2006. About HEPEX, last accessed 17 May 2006 from 〈http://hydis8.eng.uci.edu/hepex/〉.

Hymas, K., 1993. The Meteorological Office National Severe Weather Warning Service. Meteorological Magazine 122, 53–61.

Knorr-Cetina, K., 1999. Epistemic Cultures: How the Sciences Make Knowledge. Harvard University Press, Cambridge, MA.

Lalaurette, F., van der Grijn, G., 2003. Ensemble forecasts: can they provide useful early warnings? ECMWF Newsletter No. 96, European Centre for Medium-range Weather Forecasts, Reading, UK, pp. 10–18.

Lees, M., Young, P.C., Ferguson, S., Beven, K.J., Burns, J., 1994. An adaptive flood warning scheme for the River Nith at Dumfries. In: White, W.R., Watts, J. (Eds.), River Flood Hydraulics. Wiley, Chichester, pp. 65–75.

Legg, T.P., Mylne, K.R., 2004. Early warnings of severe weather from ensemble forecast information. Weather and Forecasting 19, 891–906.

Leith, C.E., 1974. Theoretical skill of Monte-Carlo forecasts. Monthly Weather Review 102, 405–418.

MacKenzie, D., 1990. Inventing Accuracy: An Historical Sociology of Nuclear Missile Guidance. MIT Press, Cambridge, MA.

Morss, R.E., Wilhelmi, O.V., Downton, M.W., Gruntfest, E., 2005. Flood risk, uncertainty, and scientific information for decision making: lessons from an interdisciplinary project. Bulletin of the American Meteorological Society 86, 1593–1601.

Murphy, A., Lichtenstein, S., Fischhoff, B., Winkler, R.L., 1980. Misinterpretation of precipitation probability forecasts. Bulletin of the American Meteorological Society 61, 695–701.

NRC, 2006. Toward a New Advanced Hydrologic Prediction Service (AHPS). National Academies Press, Washington, DC last accessed 21 July 2006 from 〈http://www.nap.edu/catalog/11598.html/〉.

Nichols, N., 1999. Cognitive illusions, heuristics, and climate prediction. Bulletin of the American Meteorological Society 80, 1385–1397.

Palmer, T.N., 2002. The economic value of ensemble forecasts as a tool for risk assessment: from days to decades. Quarterly Journal of the Royal Meteorological Society 128, 147–174.

Pappenberger, F., Beven, K., 2006. Ignorance is bliss: or seven reasons not to use uncertainty analysis. Water Resources Research 42, W05302.

Parker, D.J., Handmer, J.W., 1998. The role of unofficial flood warning systems. Journal of Contingencies and Crisis Management 6, 45–60.

Pielke Jr., R.A., 1999. Who decides? Forecasts and responsibilities in the 1997 Red River flood. Applied Behavioral Science Review 7, 83–101.

Polanyi, M., 1967. The Tacit Dimension. Anchor Books, Doubleday & Co., Garden City, NY.

Rayner, S., Lach, D., Ingram, H., 2005. Weather forecasts are for wimps: why water resource managers do not use climate forecasts. Climatic Change 69, 197–227.

Rothstein, H., Huber, M., Gaskell, G., 2006. A theory of risk colonization: the sprialling regulatory logics of societal and institutional risk. Economy and Society 35, 91–112.

Shackley, S., Wynne, B., 1995. Integrating knowledges for climate change: pyramids, nets and uncertainties. Global Environmental Change 5, 113–126.

Slovic, P., Fischhoff, B., Lichtenstein, S., 1977. Behavioral decision theory. Annual Review of Psychology 28, 1–39.

Stewart, T.R., 2000. Uncertainty, judgement and error in prediction. In: Sarewitz, D., Pielke, Jr., R.A., Byerly, R. (Eds.), Prediction: Science, Decision Making, and the Future of Nature. Island Press, Washington, DC, pp. 41–57.

Stewart, T.R., Heideman, K.F., Moninger, W.F., Reagan-Cirincione, P., 1992. Effects of improved information on the components of skill in weather forecasting. Organizational Behavior and Human Decision Processes 53, 107–134.

Taylor, J., Buizza, R., 2003. Using ensemble weather prediction in energy demand forecasting. International Journal of Forecasting 19, 57–70.

Thielen, J., Ramos, M.-H., Bartholmes, J., de Roo, A., Pappenberger, R., Cloke., H., Demeritt, D., 2005. Summary Report: First Workshop on the Use of Ensemble Prediction System in Flood Forecasting. Joint Research Centre of the European Commission, Ispra, Italy.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185, 1124–1130.

US National Weather Service, 2001. Strategic Plan for Weather, Water, and Climate Services 2000–2005. National Oceanic and Atmospheric Administration, Washington, DC Updated with changes approved as of December 2001, last accessed 7 April 2006 from 〈http://www.weather.gov/sp/strplan.htm〉.