

# Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers and end-users

Vanessa J. Fundel<sup>1</sup>  | Nadine Fleischhut<sup>2,3</sup>  | Stefan M. Herzog<sup>2</sup>  | Martin Göber<sup>1,3</sup> | Renate Hagedorn<sup>1</sup>

<sup>1</sup>Deutscher Wetterdienst, Offenbach, Germany

<sup>2</sup>Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

<sup>3</sup>Hans Ertel Centre for Weather Research, Offenbach, Germany

## Correspondence

Vanessa J. Fundel, Deutscher Wetterdienst, Frankfurter Strasse 135, 63067 Offenbach, Germany.

Email: vanessa.fundel@dwd.de

## Funding information

Hans Ertel Centre for Weather Research, a network of universities, research institutes and the Deutscher Wetterdienst, and funded by the BMVI (Federal Ministry of Transport and Digital Infrastructures).

Today's ensemble weather prediction systems provide reliable and sharp probabilistic forecasts—yet they are still rarely communicated to outside users because of two main worries: the difficulty of communicating probabilities to lay audiences and their presumed reluctance to use probabilistic forecasts. To bridge the gap between the forecasts available and their use in day-to-day decision making, we encourage scientists, developers, and end-users to engage in interdisciplinary collaborations. Here, we discuss our experience with three different approaches of introducing probabilistic forecasts to different user groups and the theoretical and practical challenges that emerged. The approaches range from quantitative analyses of users' revealed preferences online to a participatory developer–user dialogue based on trial cases and interactive demonstration tools. The examples illustrate three key points. First, to make informed decisions, users need access to probabilistic forecasts. Second, forecast uncertainty can be understood if its visual representations follow validated best practices from risk communication and information design; we highlight five important recommendations from that literature for communicating probabilistic forecasts. Third, to appreciate the value of probabilistic forecasts for their decisions, users need the opportunity to experience them in their everyday practice. With these insights and practical pointers, we hope to support future efforts to integrate probabilistic forecasts into everyday decision making.

## KEYWORDS

decisions under uncertainty, probabilistic forecast, risk communication, user dialogue, weather

## 1 | INTRODUCTION

Probabilistic forecasts have been available for many years within weather services (e.g., probabilities for exceeding precipitation thresholds or for the occurrence of thunderstorms). Yet with the exception of standardised weather forecasts for airports and several flood forecasting offices (Frick and Hegg, 2011), probabilistic forecasts are rarely communicated to outside users. Historically, uncertainty in weather forecasts – in the form of odds or verbal terms – was first communicated in the United Kingdom, France, and the United States

about 200 years ago (Murphy, 1998). More than 150 years later, the advent of numerical weather prediction models heralded a new era and transformed these initial ideas into operational probabilistic forecast products. While the early probability forecasts were based on subjective and statistical interpretations of the intrinsic uncertainty in single numerical forecasts, the estimation of uncertainty later relied on the combination of several forecast runs of one numerical weather model (a “time-lagged ensemble”; Branković *et al.*, 1990) or of numerical forecasts from different operational centres (a “poor-man's ensemble”; Ebert, 2001). The first global

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

medium-range ensemble prediction systems (EPSs) came into operation in the 1990s (Houtekamer *et al.*, 1996; Molteni *et al.*, 1996; Palmer *et al.*, 1992; Toth and Kalnay, 1993; 1997). The early EPSs suffered from underdispersiveness, indicating that the systems failed to correctly represent all sources of uncertainty (Buizza *et al.*, 2005). Since then, intensive research and development efforts in data assimilation, initial condition perturbation, and model physics representation have continuously improved the performance of operational EPSs (Bauer *et al.*, 2015; Palmer, 2018 give overviews and historical context). In order to provide reliable and sharp forecasts that can increase the economic value of risk assessments and decision making (Palmer, 2002), statistical calibration methods are used to complement physical modelling (e.g. Roulston and Smith, 2003; Hamill *et al.*, 2004; Gneiting *et al.*, 2005; Raftery *et al.*, 2005).

While the calibration of operational probabilistic forecasts still leaves room for improvement, the main obstacle to making probabilistic forecasts commonplace is the difficulty of communicating probabilities to lay audiences. The worry is that uncertainty may be misunderstood, resulting in risky behaviour or loss of trust. However, there is growing evidence from research in risk communication and decision making that probabilistic information can facilitate decision making (Joslyn and LeClerc, 2013) and that presenting probabilities in transparent formats can fundamentally improve laypeople's understanding of them (Hoffrage *et al.*, 2000; Gigerenzer *et al.*, 2007). In fact, probabilistic information is actually favoured by laypeople (Morss *et al.*, 2008). It can increase trust in forecasts (LeClerc and Joslyn, 2015) and may improve decision making (Roulston *et al.*, 2006). Forecast uncertainty must be presented clearly for these benefits to take effect; when, for instance, it is expressed using poorly defined verbal probability expressions, interpretations vary dramatically across people and situations (Weber and Hilton, 1990; Budescu *et al.*, 2014; Pardowitz *et al.*, 2015). Nonetheless, the current practice of withholding probabilistic information rather than striving to express it transparently impedes a shared decision-making process between meteorological experts, institutions, and the public (Gigerenzer and Muir Gray, 2011 discuss this issue for medicine). Rather than enabling institutions and the public to make informed decisions, meteorologists often find themselves – reluctantly – deciding on behalf of others by issuing deterministic warnings without knowing the particular needs of their end-users.

In order to encourage wider dissemination of probabilistic forecasts, the World Meteorological Organisation published guidelines on communicating forecast uncertainty and explicitly emphasised the need to address the challenge of communicating and responding to “low probability, high impact” events (Gill, 2008). Despite these efforts, the challenge of bridging the gap between the information available from weather services and its integration in day-to-day decision making persists (Taylor *et al.*, 2018). With the aim of fully exploiting the information in ensemble

predictions, the German national weather service (Deutscher Wetterdienst, DWD) has augmented its personnel resources to intensify both the user dialogue and the development of ensemble-based, user-optimised forecast products. The Hans Ertel Centre for Weather Research (Simmer *et al.*, 2016) enables the DWD to build long-term collaborations with social and behavioural scientists. This presents a unique opportunity to explore weather-related risk communication (i.e. the communication of probabilistic information) in interdisciplinary research groups, and to develop, in cooperation with end-users, methods and approaches for introducing probabilistic weather information.

In this article, we present three practical applications that communicate probabilistic weather forecasts to three professional user groups that differ considerably in their background, tasks and weather-dependent decisions. The applications were developed and analysed in interdisciplinary collaborations of DWD staff, key users of the DWD's services, and social and behavioural scientists outside the DWD. For each application, we used a different approach to introduce probabilistic forecasts and to tailor them to users' decision processes. The main goal of this article is to share and discuss our experience with the various approaches, the strengths and limitations we encountered, and the theoretical and practical challenges that emerged. In addition, we provide pointers to best practices and relevant literature from risk communication research, social and behavioural science to support future efforts to integrate probabilistic forecasts into everyday decision making.

## 2 | THREE DIVERSE PRACTICAL APPLICATIONS

One of the main duties of the DWD is safeguarding citizens and critical infrastructures in Germany by providing timely and reliable meteorological services and weather warnings. To achieve this, the DWD collaborates closely with agencies and parties responsible for emergency management and critical infrastructure.

In the following, we present our efforts to meaningfully integrate probabilistic weather forecasts into the decision-making processes of (a) emergency managers, (b) road workers, and (c) electrical transmission grid operators. The main duty of all three user groups is to ensure the security and safety of citizens and critical infrastructure. They prevent and react to possible weather-related hazards as part of their day-to-day business. In the face of this uncertainty, probabilistic forecasts would allow them to make informed decisions (Murphy, 1966; Palmer, 2002). We therefore provided them with such forecasts – even if they initially did not request them or indicated that they preferred deterministic forecasts. In doing so, we followed a shared decision-making approach that aims to inform decision makers rather than to withhold information about uncertainty and make decisions

on their behalf. After all, forecasters possess neither the users' domain knowledge about the concrete decisions for which the forecasts are used nor their understanding of the constraints in responding to forecasts and actual weather conditions.

The user groups differ considerably with respect to their educational background, their daily tasks and the weather-dependent decisions they need to make:

1. Emergency managers: decision makers in a large number of control and dispatch centres across Germany, responsible for coordinating resources and personnel across emergency services before, and as a reaction to, severe weather events. Their educational background is heterogeneous; most of them completed an apprenticeship prior to upgrading their skills to become emergency managers.
2. Road workers: "hands-on" practitioners with a specialised apprenticeship and very detailed local knowledge about the weather. The head of a large road operation centre is often educated in civil engineering or similar disciplines. This user group has only a few weather-dependent decisions to make.
3. Transmission system operators (TSOs): staff at the four German TSOs, with regional to nationwide responsibility, educated in physics, mathematics, or engineering, and trained to perform complex tasks. Their work comprises a large number of different weather-dependent tasks and also includes the development of automatic applications.

To tailor probabilistic forecasts to the needs of each user group and to ensure their sustained use, our work was guided by the following four questions:

1. Who are the users and what are their tasks and weather-dependent decisions?
2. How can we introduce probabilistic forecasts and tailor them to users' decision processes?
3. What kind of visualisation of probabilistic weather forecasts is suitable for users to make these decisions?
4. How valuable are the new probabilistic forecasts for users' decisions?

As the disparate characteristics and tasks of the user groups posed unique challenges, we used different approaches to develop each practical application. In the first example, we analysed users' revealed preferences by collecting and evaluating users' interactions with a dedicated online information platform. Here, the approach was to provide different representations in parallel in order to passively observe and quantify which ones users selected under real operational constraints. The second example introduced probabilistic forecasts along with a suitable trial case that was jointly defined during a workshop with a selected sample of users. Here, the approach emphasised the training aspect and direct personal interaction with the users. It focused on applying probabilistic forecasts to one particular type of decision. In the third example, we collaborated closely with users in expert teams to consider the entire range of possible

applications of probabilistic forecasts from scratch. In contrast to the first two user groups, this group has an education in physics, mathematics, or engineering and is trained to develop solutions for complex problems. Here, the approach on the one hand is to jointly incorporate probabilistic forecasts into existing automatic decision applications and, on the other hand, to integrate tailored probabilistic forecasts into a demonstration tool that showcases their potential benefits and allows us to quickly implement changes in the displayed information and realise new requirements.

Both the underlying weather forecast database and the forecast horizons (from short-term to medium-range) differ amongst the three applications. The first two examples comprise probabilistic forecasts from a model output statistics system, while the third example uses probabilistic forecasts from a mesoscale ensemble prediction system. In the DWD's assessment, all are sufficiently reliable (calibrated) for events that occur about 1–10 days per year at a location; for less frequent events, verification results are not stable. Once users can provide appraisals based on long-term experience, we will have to evaluate whether this forecast quality is sufficient for each respective application. However, from the perspective of risk communication research,<sup>1</sup> the first important step – and the focus of this paper – is communicating probabilistic information and successfully integrating it into decision processes, regardless of the underlying models and lead times (provided these are well-calibrated).

## 2.1 | Probabilistic forecasts for emergency management

Providing weather forecasts and warnings to emergency managers is one of the DWD's main duties. This diverse user group includes fire brigade control and dispatch centres, volunteer fire fighters, police, paramedics, and emergency managers from other technical organisations (Kox *et al.*, 2015; 2018b). The largest subgroup consists of control and dispatch centres as well as commanding fire fighters, who are responsible for coordinating emergency services and deploying vehicles and personnel. However, their educational backgrounds can differ considerably, ranging from practitioners to users with an academic background.

Although this subgroup's main duties are carried out as a reaction during or after an event, weather forecasts and early warnings are relevant for planning and preparation to maintain the ability to quickly respond to emergencies during severe weather events (Kox *et al.*, 2018b). For instance, weather forecasts can inform decisions to call in off-duty

<sup>1</sup>The concept of risk has different meanings in different disciplines (Kozyreva and Hertwig, 2019; Kozyreva *et al.*, 2019). For instance, in psychology and judgement and decision making, "risk" typically refers to situations in which the probabilities of potential outcomes are known (Luce and Raiffa, 1957) whereas in economics, "risk" often refers to the variability in outcomes (Weber *et al.*, 2004). Finally, laypeople often focus on other characteristics of hazardous events, such as their catastrophic potential or controllability (Slovic, 1987).

staff, relocate staff, or prepare vehicles and equipment. Currently, emergency managers receive deterministic warnings and forecasts through an online information system (FeWIS) operated by the DWD. It provides weather information tailored to user regions on various temporal scales, from general early warning information a week ahead to nowcasting information (Kox *et al.*, 2018a). Official weather warnings are issued up to 12 hrs in advance on a municipal level and typically serve as a signal for a range of decisions by emergency managers, such as whether to declare an emergency.

This points to a major challenge for deterministic warnings: to determine the optimal probability threshold for issuing a warning, forecasters need to know how important it is for the decision maker to prevent or ensure particular possible outcomes, that is, the relative utility of the different possible consequences (Luce and Raiffa, 1957; Murphy, 1966). The varying requirements among users further amplify this challenge: a study of emergency managers revealed that their preferred lead times differ (Kox *et al.*, 2015), which may reflect varying institutional requirements (Demeritt *et al.*, 2010), capacities, population densities, or geographical areas. The same study (Kox *et al.*, 2015) finds that the probability thresholds at which emergency managers would start preparations also vary. When presented with a hypothetical storm forecast and asked to indicate at which probability they would start preparatory measures, about 30% of managers indicated a probability of 50% and about 35% a probability of 70%. This implies that there is no single probability threshold for issuing a warning that could account for every user's needs – particularly in large user groups. Deterministic warnings therefore actually require a close interaction between forecasters and users that is rarely feasible in practice.

Probabilistic forecasts offer a solution: they allow users to apply their own decision thresholds as they see fit (Murphy, 1977) and thus increase the value of their decisions (Palmer, 2002). To date, however, even the severe weather watch (pre-warning) that covers longer lead times of up to 48 hrs consists merely of verbal descriptions of a possible weather development and its likelihood (e.g. wind speeds above 100 km/h are “likely” or local thunderstorms “may” occur). Yet verbal probability expressions are ill-suited to inform emergency managers decisions because they are ambiguous: they are interpreted differently by different people and the interpretation depends on the base rate and the severity of the target event (Weber and Hilton, 1990; Budescu *et al.*, 2014; Kox *et al.*, 2015; Pardowitz *et al.*, 2015).

Numerical probabilistic forecasts, in contrast, can convey the same quantitative information to everyone while enabling users to start preparations at a self-determined probability threshold that fits their individual needs (Murphy, 1966; Palmer, 2002). Probabilistic forecasts bridge the gap between more certain short-term warnings and earlier but less certain weather watches (expressed in ambiguous verbal probability expressions) by providing continuous information from 48 hrs to 1 hr in advance. However, this solution may or may not

fit current institutional practice or the needs of emergency managers, who often have to make decisions as a reaction to official warnings rather than in anticipation of likely upcoming weather conditions (Kox *et al.*, 2018b).

In order to evaluate the potential of probabilistic forecasts, we used a new approach to investigate two questions: Do emergency managers consider probabilistic information useful under operational constraints? Which representations do they prefer, and when? This case-study was part of the interdisciplinary research project WEXICOM funded by the Hans Ertel Centre for Weather Research (Simmer *et al.*, 2016). WEXICOM investigates how to advance the communication, understanding and use of uncertainty of weather warnings and weather risks.

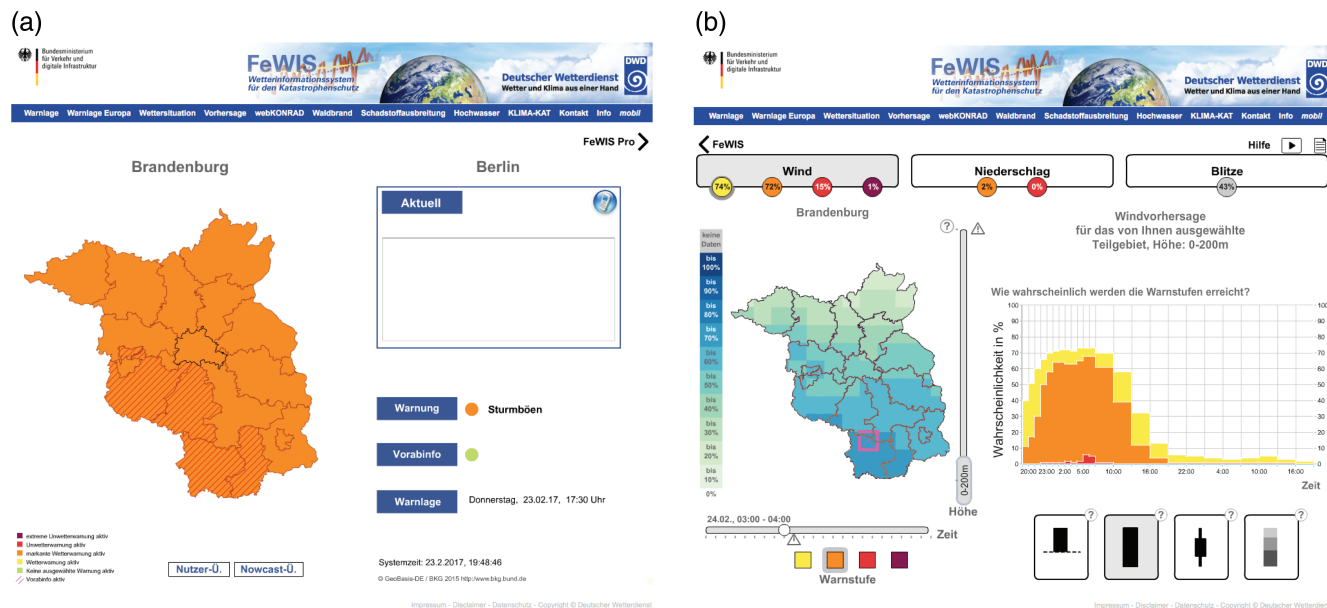
### 2.1.1 | Introducing forecast uncertainty

To test which information emergency managers might find useful, we developed and implemented five different representations of probabilistic forecasts for the most relevant weather conditions for emergency managers in Germany: wind, precipitation, and thunderstorms (the latter defined as the occurrence of lightning; Figure 1b; Kox *et al.*, 2015). Using the online system FeWIS Pro, users can choose between different representations of forecast uncertainty (Figure 2). The system makes it possible to analyse users' online behaviour by tracking which representation was selected and when from a collective user ID (usually shared by multiple managers within one control centre). This procedure was ethically approved by an Institutional Review Board.

The approach has several methodological advantages. First, observing behaviour allows for quantifying emergency managers' revealed preferences under real operational constraints. In contrast, self-reported preferences do not necessarily match people's actual preferences in real situations: What we say is not necessarily what we do (Frey *et al.*, 2017). We observed this phenomenon in our second case-study (Section 2.2). Furthermore, in self-reports users do not always prefer the information they understand best (Hildon *et al.*, 2011; Okan *et al.*, 2015). Second, users can gain experience with probabilistic information and their various representations before their understanding is tested or before they evaluate the representations (Hogarth and Soyer, 2015; Lejarraga *et al.*, 2016; Wulff *et al.*, 2018, discuss the importance of experience in decision making). Third, different information and representations are available in parallel, so that the needs of different user groups, in different situations or under different weather conditions, can be identified.

The design of the representations (Figure 2) was informed by empirically validated recommendations from research in other domains (e.g. medical risk communication) to facilitate the comprehension of uncertainty. Conceptually, the representations display forecast uncertainty in two ways: either as probabilities for binary events (in FeWIS Pro for exceeding warning thresholds or the occurrence of a





**FIGURE 1** (a) shows an official deterministic warning in FeWIS, and (b) the corresponding probabilistic forecast in the new system, FeWIS Pro. The probabilistic system allows the user to select one of three variables of interest (wind, precipitation, or thunderstorms) and display the forecast in five different representations. The map centres on the area of responsibility of each user (here Berlin) as well as the surrounding region and shows exceedance probabilities. The colours convey the probability of exceeding a selected warning threshold (here orange) for wind at a particular time, which can be adjusted using the slider below. By clicking on the buttons at the bottom right, the user can choose between four representations that give a temporal overview for a selected grid point on the map (pink square). Here, the diagram on the right shows the probabilities for exceeding the yellow, orange, and red thresholds, respectively, within the next 48 hrs; all four representations are explained in more detail in Figure 2. [High-resolution images for all applications are available at the following Open Science Framework repository: <https://osf.io/k6nb7>].

thunderstorm), or as probability thresholds for continuous physical variables (10, 25, 50, 75, and 90%iles; in FeWIS Pro for wind speed and precipitation). All representations display model output statistics from the DWD's operational WarnMOS product. The forecasts are available with a 48 hrs lead time (1 hr intervals for the first 12 hrs, then at 3 hrs intervals; spatial resolution is always 20 km × 20 km). The forecasts are updated every hour.

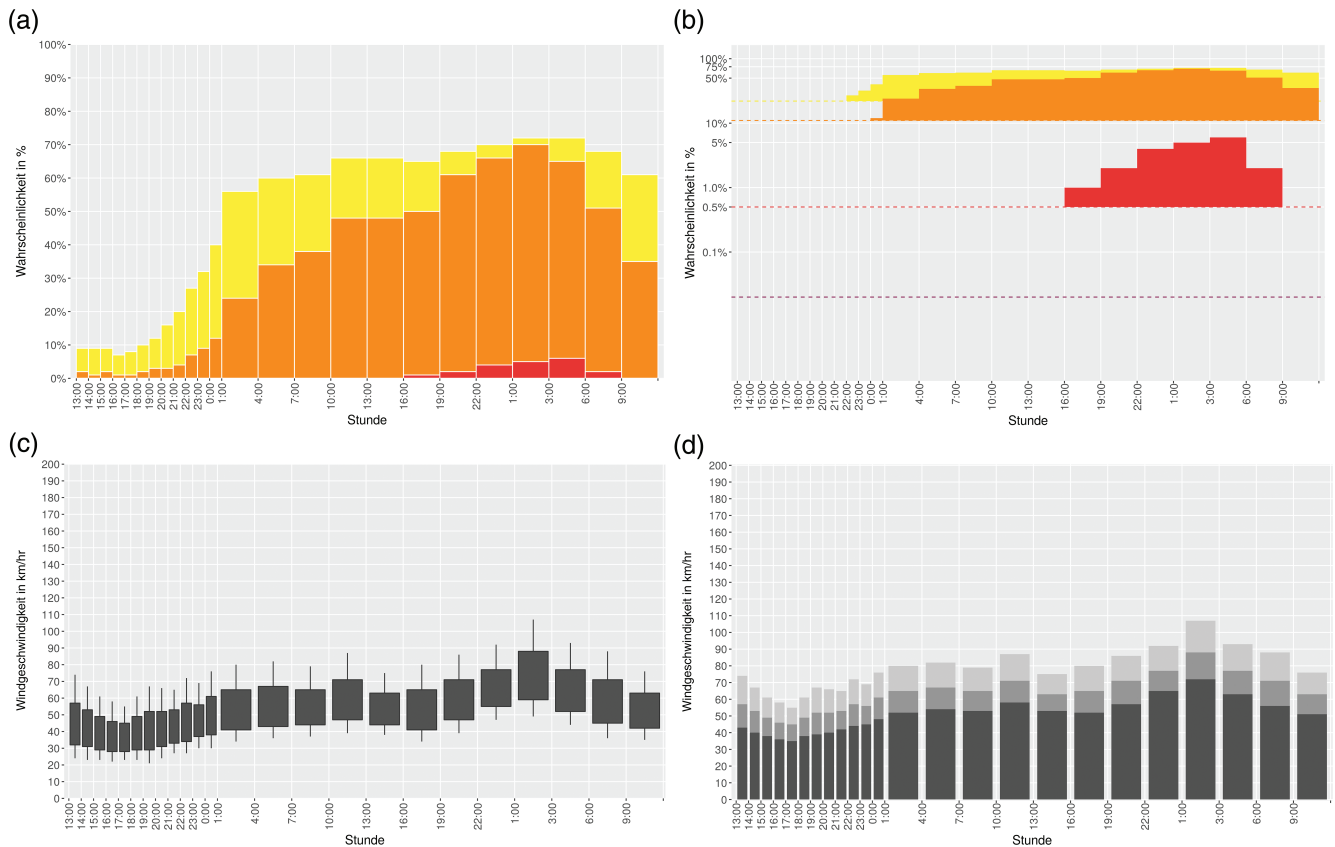
Which representations do emergency managers prefer and why? On the one hand, one might expect emergency managers to prefer probabilities for exceeding warning thresholds because they may seem similar to the deterministic warnings that underlie their current decision practice. On the other hand, a representation of the likely physical values, for instance of wind speeds, may be more familiar and easier to understand than probabilities. Including both types of information in parallel enabled us to evaluate which kind of information was useful to emergency managers under actual operational constraints.

### 2.1.2 | Visualisation of probabilistic forecasts: Recommendations from risk communication research

There is still relatively little research on how to best communicate uncertainty for continuous variables, which are typical in meteorological forecasts but less so in, say, medical risk communication (Spiegelhalter *et al.*, 2011). Therefore, we synthesised results and best practices on visualisation

from other research domains, such as medical and financial risk communication, as well as basic research on human perception. Below we discuss five best practices we considered when developing FeWIS Pro due to their relevance for communicating probabilistic forecasts in meteorology and other domains and provide broader pointers to research articles and systematic reviews.

*Encode quantitative information in a way that fosters accurate decoding.* Colour is less suited for reading exact information than bar or line graphs (Cleveland and McGill, 1985). When using colour, limit the number of shades to maintain perceptual differentiability (Kosslyn and Kosslyn, 2006). If chosen appropriately, different shades can be detected even by those who suffer from the most common forms of colour blindness (Wong, 2011). For continuous variables, like probabilities, a single-colour scale allows for an effortless interpretation of the ordering, with darker values indicating a higher probability (Harrower and Brewer, 2003; Gehlenborg and Wong, 2012; Stauffer *et al.*, 2015 give advice on colour scales). In contrast, conventional multi-colour scales, as sometimes used by meteorologists, require a legend to decipher the ordering – which means additional effort and potential mistakes (Borland and Ii, 2007). The map in FeWIS Pro uses a single-colour scale to represent the probability of exceeding a particular warning threshold (e.g. for storm >100 km/h; Figure 1). It also makes reading numerical information easier via mouseover texts that appear whenever the pointer hovers over any part of the representation. They



**FIGURE 2** Four representations of forecast uncertainty for wind during storm *Herwart*, which passed through Germany from 2100 CEST 28 October to 1400 CET 29 October 2017 (issued at the time of the first weather watch at 1300 CEST 27 October, for the next 48 hrs; x-axes). (a) and (b) display the probabilities of exceeding warning thresholds (y-axes). (b) also shows a comparison to climatology (indicated by dashed lines) on a log-scale; the height of the bars indicate how much more likely it is that the threshold will be exceeded compared to what is usual for the current season. (c) and (d) display forecast uncertainty through predefined probability thresholds (quantiles), here for wind speeds (y-axes). For (c), the box shows the interquartile range and the antennae show the 10 and 90%iles. For (d), the upper end of the bar (90%ile) highlights the worst case, i.e. the wind speed unlikely to be exceeded (10%). The upper end of the medium grey bar shows the 75%ile and the dark grey bar the 50%ile. [High-resolution images for all applications are available at the following Open Science Framework repository: <https://osf.io/k6nb7>].

show as text the time and value for the particular part of the representation, as well as how to interpret the forecast uncertainty.

*Explain probabilities.* Conveying probabilities as relative frequencies instead of single-event frequencies has been shown to help people comprehend uncertainty (Gigerenzer and Edwards, 2003; McDowell and Jacobs, 2017; but see Joslyn and Nichols, 2009 for the opposite finding in the weather domain). Descriptions in relative frequencies force the communicator to specify the reference class: “In 30 out of 100 days with a forecast like this, ...” rather than “The probability of X is 30%” (Murphy *et al.*, 1980; Gigerenzer *et al.*, 2005). However, single numbers like percentages can be better suited than relative frequencies for comparing different values, especially with variable denominators (for overviews of best practices for communicating probability information see e.g. Gigerenzer *et al.*, 2007; Lipkus, 2007; Visschers *et al.*, 2009; Trevena *et al.*, 2013). Given these previous findings, we choose a complementary approach and display probabilities as percentages across all representations, but use mouseovers to explain their meaning in terms of relative frequencies (“In X out of 100 situations with a forecast like this...”). We offer

further explanation in the help files using spatial frequency representations (using icon arrays; Ancker *et al.*, 2006).

*Prevent deterministic misinterpretations of forecast uncertainty.* People tend to misinterpret probabilities in a deterministic way (Joslyn and LeClerc, 2013). For instance, when asked which of three descriptions is the best interpretation of a forecast of a 30% chance of rain tomorrow, most people choose the interpretation that it will rain in 30% of the forecast area or 30% of the forecast time (Murphy *et al.*, 1980; Gigerenzer *et al.*, 2005) over the correct probabilistic interpretation: it will rain on 30% of days with this forecast. Although people’s answers may indicate that they are partly aware of the spatial and temporal uncertainty inherent in any forecast (i.e. about where or when it will rain), deterministic misinterpretations are nevertheless a cause for concern. Such misinterpretations imply that if it does not rain anywhere in the forecast area or time period, the forecast constitutes a false alarm – even though a lack of rain is still reconcilable within the correct, probabilistic interpretation.

To prevent the deterministic misinterpretation of probabilities as physical values (e.g. wind speed or the amount of precipitation), we intentionally avoided any similarities

with the colour scheme of the DWD's four levels of warnings (yellow, orange, red, and violet) when developing our single-colour scale for probabilities. Emphasising particular points on a graph, such as the medians (Broad *et al.*, 2007) or the bounds of an uncertainty interval, may inappropriately withdraw attention from the uncertainty of the forecast; furthermore, in interval forecasts, for instance, people may misinterpret the interval between a lower and an upper probability quantile as a range, such as diurnal variation in temperature (Cumming, 2007; Savelli and Joslyn, 2013). Therefore, FeWIS Pro's boxplots (Figure 2c) display neither medians nor end caps on whiskers, which could suggest an upper bound. When reading barplots, people tend to believe that points within the bar are more likely than outside the bar – even if they are equidistant relative to the top of the bar (Newman and Scholl, 2012; Okan *et al.*, 2018). FeWIS Pro's barplots (Figure 2d) factor this tendency in by displaying the 90%ile as the full height of the stacked bars, which implies that values outside the bar are in fact less likely than those within the bar.

*Put rare but severe events into perspective.* High-impact weather is rare. The challenge is to communicate that even small probabilities are important when evaluating severe risks. One approach is to communicate the relative increase in probability compared to a baseline. If the baseline provides a useful comparison, this can help laypeople to put numbers into perspective (Peters *et al.*, 2009; Barrio *et al.*, 2016). However, to avoid misunderstandings, both the baseline risk itself and the absolute level of the target risk should be included (Bodemer *et al.*, 2014); otherwise, people may overestimate the increase in risk because relative increases in risk from a low baseline tend to look large. This can result in potential overreaction or loss of trust (Gigerenzer *et al.*, 2007). Displaying values using logarithmic axes enlarges the visual difference between small values and can thus highlight small probabilities (Lipkus, 2007). However, at the same time, logarithmic axes may impair people's ability to accurately read off the absolute risks (Lipkus, 2007). In one representation in FeWIS Pro, we chose the climatology as the baseline, which shows the probability on a “normal” day for that season, and only plot forecasts that are above the baseline (using a logarithmic scale; Figure 2b). Mouseovers display the increase relative to the climatology (e.g. “five times more likely than usual at this time of year”), the climatological probability and the absolute probability of exceeding the warning threshold.

*Choose the level of detail in accordance with what can be reasonably predicted.* Despite the common desire of users to receive ever more detailed and certain information, communicators should resist conveying more precise predictions than are available (Stephens *et al.*, 2012). Communicating forecasts and their uncertainty in more detail than the underlying models can provide may raise false expectations about the predictability of weather events – especially in lay audiences. In order to make the spatial resolution of the forecasts transparent, we coloured each 20 km×20 km area of the map according to the forecast probability without smoothing their borders

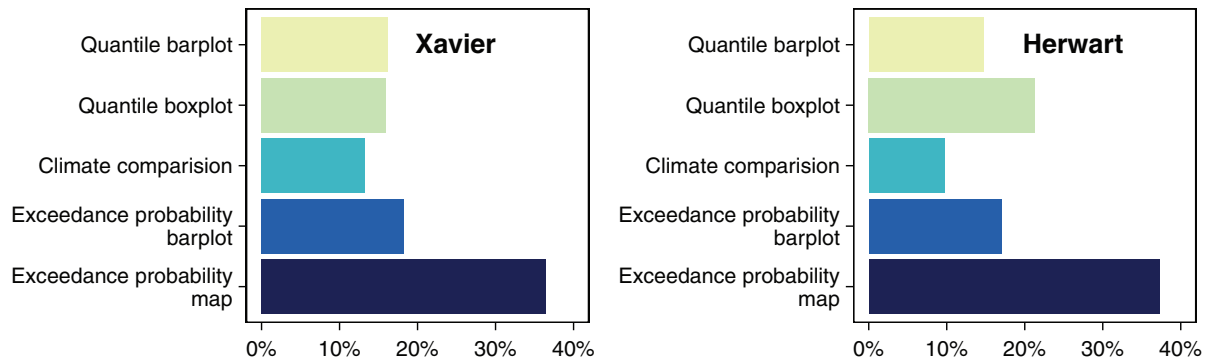
(Figure 1b). For the same reason, the width of the intervals in the diagrams (Figure 2) matched the temporal resolution of the underlying forecasts (i.e. forecasts for each hour for the first 12 hrs and then forecasts for blocks of 3 hrs onwards).

### 2.1.3 | Evaluating the usefulness of the forecasts: Lessons learned

We compared the representations users preferred and the lead times at which they started searching for information for two severe storms in Germany in October 2017: *Xavier* and *Herwart*. This case-study allowed us to analyse user preferences under real operational constraints during severe weather.

As the probabilistic forecasts have a lead time of 48 hrs, we analysed users' behaviour from 48 hrs before to 24 hrs after the storm. *Xavier* passed through Germany on 5 October 2017 from 0800 to 2000 CEST, moving from the northwest to the southeast of the country (CEST=Central European Summer Time=UTC+0200). The first weather watch for the northwest was issued at 1208 CEST on 4 October, followed by a warning at 2204 CEST. *Herwart* started on 28 October 2017 at 2100 CEST and lasted until 1400 CET on 29 October, moving from the northwest of Germany towards the east (CET=Central European Time=UTC+0100; CEST ended on 29 October 2017). The first weather watch for the northwest was issued at 1300 CEST on 27 October, followed by a warning at 1545 CEST. Depending on a storm's path, users may start looking for information at different points in time. For each storm, we therefore only considered users who had reason to expect being affected first (i.e. for *Xavier* and *Herwart*, the northwest of Germany). We defined this group as all users for whose areas the probabilistic forecasts in FeWIS Pro predicted wind speeds with a 90%ile of  $\geq 110$  km/h during the first third of each storm. This procedure resulted in 93 collective users for *Xavier* and 114 for *Herwart*.

For both severe storms, the selected representations (aggregated over wind, precipitation, and thunderstorm) showed a clear preference for the spatial representation of the map (Figure 3), which displays the probability of exceeding warning thresholds (Figure 1b). When comparing the four representations that give a temporal overview, the representation showing the probability of exceeding warning thresholds was slightly preferred over boxplots or barplots during *Xavier*, but overall differences were small. During *Herwart*, in contrast, the boxplot was the second most-used information after the map. Thus, the preference for the map likely reflects a preference for a spatial overview rather than a more general preference for probabilities for exceeding a warning threshold over quantile information. The spatial overview may be particularly useful for emergency managers to coordinate people and resources within their area because it shows which areas are most likely to be affected. At the same time, information about the likely range of wind speeds, as provided in boxplots, was also clearly used. Figure 4 shows that quantile



**FIGURE 3** Which representations were preferred during storms *Xavier* and *Herwart*? Relative proportion of times users selected a representation during the storms. In total, users selected a representation 439 times during *Xavier* and 722 times during *Herwart*. [High-resolution images for all applications are available at the following Open Science Framework repository: <https://osf.io/k6nb7>].

information was consulted especially frequently about 12 hrs before and even more during the event.

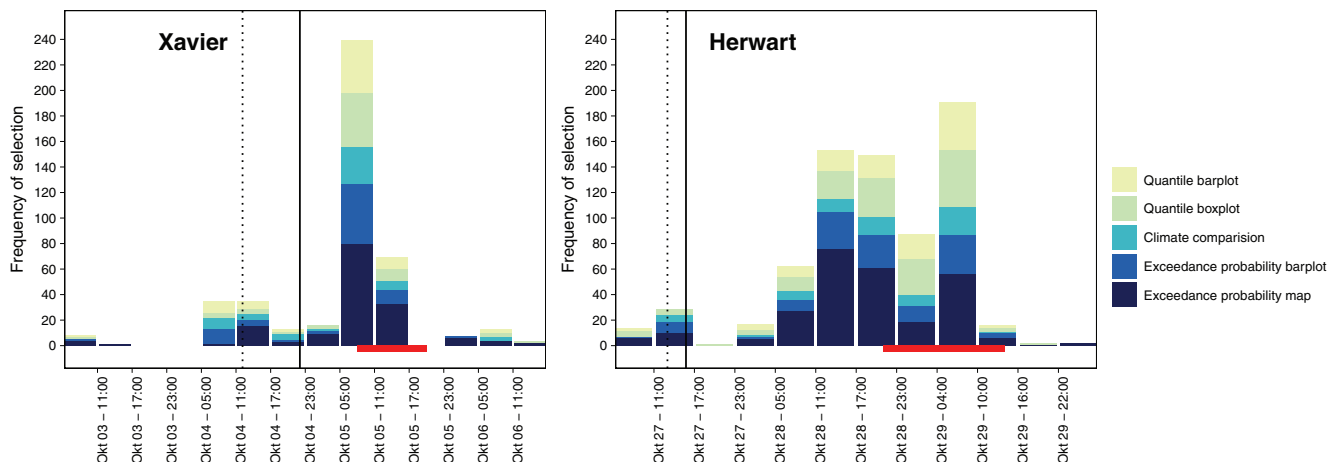
The analysis of users' behaviour also allowed us to examine the lead times at which emergency managers looked for information. For *Xavier*, users started checking the probabilistic forecasts around 24 hrs before the event and even before the first weather watch was issued (Figure 4). Peak activity was observed shortly before and during the event. For *Herwart*, in contrast, user activity before the storm was almost as high as during the storm. The earlier search for information may reflect the fact that *Herwart* occurred just three weeks after *Xavier* had caused significant damage throughout Germany. The awareness of potential impacts was thus likely much higher before *Herwart* than before *Xavier*, and this may have resulted in more anticipatory search and preparation by emergency managers. Importantly, users' behaviour shows that probabilistic forecasts were consulted at least as much during the storms as before they hit.

This latter result is also supported by a survey of 100 FeWIS Pro users. Emergency managers indicated that the probabilistic forecasts were useful not only for preparatory action

and planning, but also for managing resources and personnel during the event: 57% indicated that they considered the probabilistic forecasts useful in advance (when events are still uncertain), 66% found them useful when they became more concrete, 43% considered probabilistic forecasts useful during severe weather events, and 27% even indicated using them on a regular basis. Overall, users were satisfied with the system ( $M = 3.8$  on a 5-point rating scale), with ratings for ease of use ( $M = 3.2$ ) and understanding ( $M = 3.4$ ) being a little lower.

FeWIS Pro currently offers an online manual, video explanations, and pop-up graphic explanations; the latter two are used quite frequently. Providing further explanations in the form of videos or interactive learning through online games, simulations (Hogarth and Soyer, 2015), or quizzes may be a promising way to direct attention to important aspects of the representations and facilitate comprehension. Almost all respondents (88%) indicated that they would like to continue receiving probabilistic forecasts in the future.

In sum, the frequent use of FeWIS Pro and the positive response of survey participants shows that the probabilistic



**FIGURE 4** When did users start searching information before, during, and after storms *Xavier* and *Herwart*? Total number of selected representations by time (per 6-hour-interval). The colour indicates which representations were selected. The vertical dotted and solid lines show the times when the first weather watch and warning were issued, respectively. The horizontal red lines below the bars mark the period when the storms passed through Germany.

[High-resolution images for all applications are available at the following Open Science Framework repository: <https://osf.io/k6nb7>].



forecasts were clearly informative for emergency managers. The map and its spatial overview was the most popular representation, but boxplots, which show the likely range of wind speeds or precipitation, were also frequently selected; combining a map and quantile information into a spatial representation of quantiles may thus fit the users' needs even better than the representations we have implemented so far. The fact that users looked for forecasts more than 12 hrs before the event indicates that probabilistic forecasts can help to bridge the gap between short-term warnings (up to 12 hrs lead time) and the earlier, uncertain weather watch (up to 48 hrs lead time). As suggested by previous research, emergency managers would in fact use probabilistic forecasts even during a storm (Kox *et al.*, 2018b). Here, emergency managers preferred quantile information, such as boxplots, about the range of possible wind speeds, which may have helped them to maintain their ability to respond during the event and to prepare for daily requirements and the operations happening directly afterwards.

How probabilistic forecast information can best be combined with deterministic warnings given the current institutional practices of emergency managers in Germany is still an open question (Kox *et al.*, 2018b). Nevertheless, our results are encouraging and indicate that probabilistic forecasts are indeed useful for a number of different decisions of emergency managers under real operational constraints.

## 2.2 | Probabilistic forecasts for road authorities

Road authorities and operation centres have a long tradition of collaborating with the DWD; supporting them is a crucial part of the DWD's public services. Members of this user group share the common goal of preventing accidents and traffic jams, but differ in terms of personnel at their maintenance depot. For instance, some maintenance depots are staffed 24 hrs a day, seven days a week, while others cover weekends and nights with standby staffing and external service providers. Yet in terms of educational background, the user group is quite homogeneous; road workers complete a specialised apprenticeship and can take advanced training in management duties.

In wintertime, on-duty road workers decide on two main weather-dependent questions:

- (i) Is a preventive application of gritting salt necessary?
- (ii) Is it necessary to prepare for snow clearance and gritting?

The head of the road operation centre has a further weather-dependent duty:

- (iii) Scheduling standby duty on weekends and public holidays.

All decisions require predictive planning in order to prepare equipment and adapt service schedules. The first two road

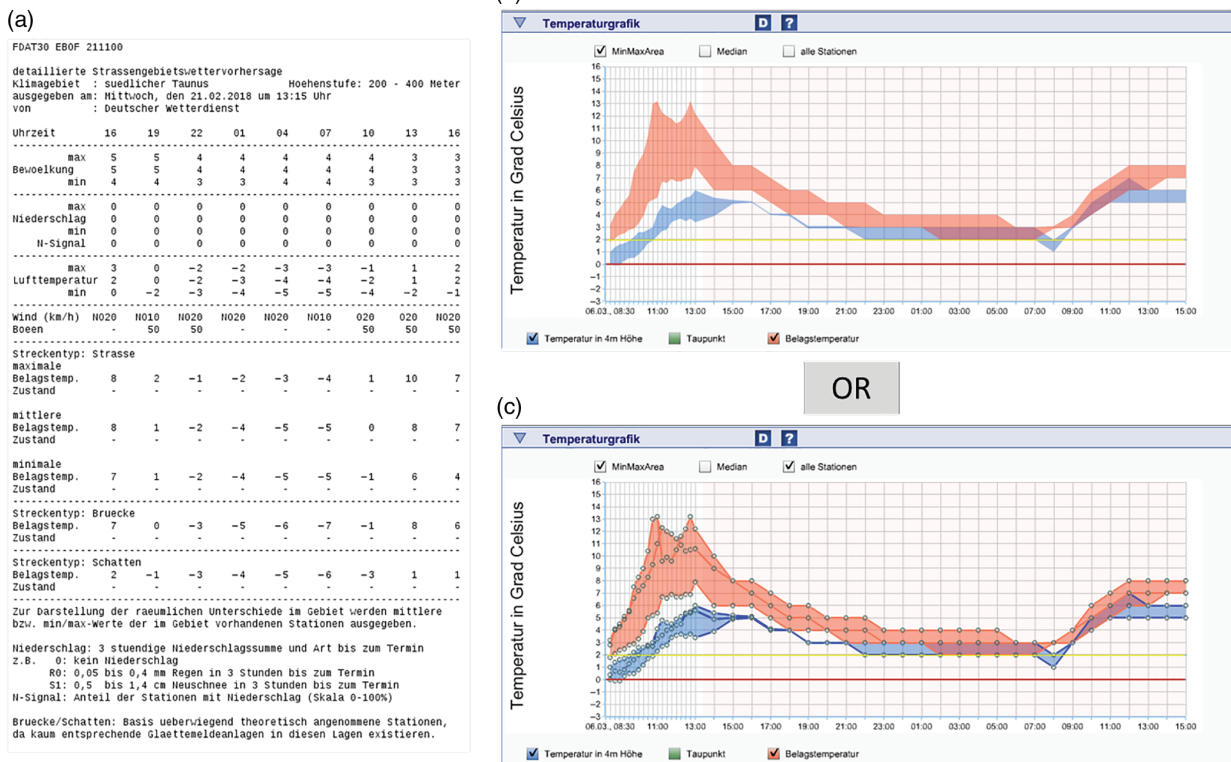
operation decisions rely on SWIS, a web-based interactive tool that the DWD developed to supply this user group with tailored information. Besides general weather information such as radar and satellite maps and weather reports, SWIS provides observations and associated road weather forecasts at around 1,500 German road weather stations, with a 27 hrs lead time. The third decision requires forecasts with a longer lead time – up to seven days into the future. This longer forecast horizon was initially not included in SWIS for the users' local area as they did not request it and mentioned relying on the regional, text-based weather report within SWIS or free websites for this information.

The development of SWIS relies on well-established interactions between DWD and road authorities. A user workshop is held every two years at the DWD headquarters to encourage the dialogue between meteorologists and users. The aim is to jointly identify strengths and weaknesses of SWIS and to schedule new developments. During these workshops, members of the DWD present verification results of road weather forecasts, as well as new developments. In turn, users ask questions, highlight new requirements, and give feedback on the past winter season and the ability of SWIS to support their work. In addition to the user workshop, DWD staff members offer regular tutorials on SWIS and are in close contact with road operation centres during the winter in order to advise on the actual weather situation.

For the first two road operation decisions, users focus on short-term forecasts that cover the next 10 hrs. The reason for this is twofold: first, the preparation time required for a road operation is relatively short (1–3 hrs), and second, the interest and confidence in numerical road weather forecasts beyond a short lead time has traditionally been limited. For time horizons beyond 10 hrs, for example, to schedule standby duty over several days into the future, users preferred to rely on their own observations of the roads in their catchment, text-based regional weather reports and their experiences of the local weather conditions and likely developments. However, the confidence in road weather forecasts has increased over time and the interest shown in a more graphical presentation is generally higher in the new generation of road workers than in their older colleagues. As a result, SWIS has evolved in recent years from a text-based information platform to an interactive tool with maps and charts covering forecast horizons of up to 27 hrs.

Road workers previously did not have the opportunity to experience probabilistic forecasts in SWIS. However, prior to the introduction of probabilistic forecasts, we had already introduced a graphical tool within SWIS for combining forecasts for a variable of interest for several road weather stations in one graph, for example, road temperature (Figure 5). The tool allowed them to get an overview of road weather conditions as well as of the spatial variability in their catchment. Working with these graphs familiarised users with variability and may have helped them to develop an informal understanding of forecast uncertainty.

## Weather forecasts as table → Visualisation of spatial variability in short-term forecasts



**FIGURE 5** Transition from text-based weather forecasts (a) to graphical presentations of temperature (blue) and road temperature (red) forecasts (b, c) within the graphical tool of the road weather information system SWIS. With point forecasts as a function of time (b, c) rather than forecasts for predefined regions in a table (a), users can define groups of road weather stations that they want displayed in the same chart. The default graph shows the coloured envelope around the point forecasts for the selected stations (b); users can also display the individual forecasts (c). The combined presentation of the forecasts familiarises users with variability. [High-resolution images for all applications are available at the following Open Science Framework repository: <https://osf.io/k6nb7/>].

### 2.2.1 | Introducing forecast uncertainty

As users' confidence in road weather forecasts and interest in graphical information increased, an active discussion between users and DWD staff began on how best to use probabilistic forecasts for winter services. This was a far cry from early discussions, in which users objected to probabilistic forecasts for short-term decisions; with the great burden of responsibility for preventing accidents, they were wary of basing their decisions on "uncertain" information.

Therefore, in collaboration with the road authorities, we identified the long-term planning of personnel resources as a suitable trial case. In our view, this weather-dependent decision was particularly suitable to explore how to use probabilistic forecasts for three reasons. First, users need forecasts for up to seven days into the future to schedule stand-by duties. At the time, not even deterministic forecasts for their catchment and the upcoming week were available to users within SWIS, and thus no established routines with familiar products existed. Second, users already understood that forecasts for the long-term future tend to be more uncertain than those for the short-term horizon, and were thus prepared to accept a probabilistic, uncertain forecast in this context. Third, the stakes were low; there was no risk to human lives or property. Taking preventive action and

preparing staff for a road mission incurs higher personnel costs – regardless of whether or not that activity proves to have been necessary. Not taking action means that additional personnel might have to be called at short notice and at a higher price. The risk was thus purely financial.

The users' hesitation and concerns around probabilistic forecasts indicated the need for a participatory approach that allowed for a close dialogue with the meteorological service in order to overcome reservations (Palmer, 2002) and engage users in order to promote the understanding and potential benefits of probabilistic forecasts for their work. A close interaction was feasible due to the already well-established interactions with road authorities and the similarity of users' weather-dependent decisions.

To introduce probabilistic forecasts, we organised a test user workshop for the new products called "trend forecasts". The aim was twofold: first, to explain probabilistic forecasts and demonstrate their potential benefit for users' work, and second, to jointly develop possible visualisations for their implementation in SWIS. The road authorities identified 40 test users, all of whom were responsible for planning personnel, to be invited to the workshop. Each federal state and kind of road operation centre was represented within this sample. The programme included a tutorial about forecast uncertainty

in numerical weather predictions and how weather services try to address it with ensemble prediction systems (EPS). The interactive presentation illustrated the practical implications for users and the information content of probabilistic forecasts using realistic weather scenarios. The aim was to highlight the differences between the familiar, deterministic short-term forecasts and the new long-term, probabilistic forecasts.

To demonstrate the potential benefits of probabilistic forecasts, we explained the cost–loss concept (Murphy, 1977; Richardson, 2001) with the help of the trial case, and discussed the quantification of costs and losses for this particular decision. In addition, we explained how to understand a probabilistic forecast and what kind of information can be drawn from it by showing examples inspired by the users' own daily tasks.

We opened the dialogue with simple quiz questions about the interpretation of probabilistic forecasts in order to engage the users after the presentation and to assess how the concept of EPS, derived products, and their information was received and understood. The answers revealed that the large majority of users understood the idea of probabilistic forecasts.

To demonstrate the implications of probabilistic forecasts for users' decisions, we then relied on a simple economic game. The users had to decide whether or not to plan for stand-by road workers for the upcoming weekend (3–4 days lead time) based on a probability for snow – knowing that additional personnel were associated with additional costs even without an on-site operation.

Finally, we set aside 2 hrs to discuss possible visualisations of probabilistic forecasts and their implementation in SWIS. We used a World Café format (Schieffer *et al.*, 2004; Fouché and Light, 2011; Kox *et al.*, 2018a) for which we prepared different possible visualisations as suggestions to be discussed, changed, and refined.

For the following winter season, we asked workshop participants to test the new trend forecasts and complete a survey at the end of the test phase.

### 2.2.2 | Visualisation of probabilistic forecasts: building on what is familiar to users

Based on the discussion during the user workshop, we implemented the new probabilistic forecasts in SWIS as an extension of the existing graphical tool providing local forecasts described above. Our aim was to build on the presentation the users were familiar with and to introduce users to forecast uncertainty in an informal way.

For temperature, the trend forecasts show only one uncertainty envelope, covering the most likely, middle 68% of the probability distribution for each road weather station (Figure 6). If the user combines several stations within one graph, the default setting displays an overall uncertainty envelope, showing the range between the minimum and maximum of all lower and upper quantiles across the selected stations. Users can also add the median of the

temperature forecast; by default, however, the median is not shown in order to focus attention on the uncertainty of the forecasts (Broad *et al.*, 2007). A next step would be to tailor the uncertainty envelope directly to users' informational needs once they have gained experience with the forecasts. An interesting way to do this would be to experimentally elicit how wide users prefer the envelope to be (that is, how "certain" they want to be) while still finding it informative (Yaniv and Foster, 1995).

In earlier discussions, users had identified two temperature thresholds (2 °C and 0 °C) that are critical to their decisions and are therefore highlighted in the graph in yellow and red, respectively. A second important step for future development is also to display the probabilities that temperature falls below these two thresholds. This would provide users with probability information about the risk of road ice and slickness that matches the overall task goal and their decision thresholds (Joslyn *et al.*, 2009).

For the precipitation forecasts for rain, snow, and sleet, the plots of the short-term precipitation forecast served as a model for the representation of the probabilistic forecast (Figure 7). We maintained the colour coding for the states of precipitation and the presentation of road weather stations as separate rows. However, in the short-term forecast view, the user sees quantitative precipitation forecasts, whereas in the medium-range forecast view the user sees the probability of any rain, snow, or sleet. Probabilities of the different precipitation states are shown as stacked bars. To visually highlight high probabilities yet avoid a deterministic misinterpretation (Joslyn and LeClerc, 2013) as amount of precipitation, the bars are by default centred around a baseline. The numerical probabilities are not shown on the y-axes, but are displayed on mouseover (Figure 7). Users can also choose a non-centred bar chart if they prefer.

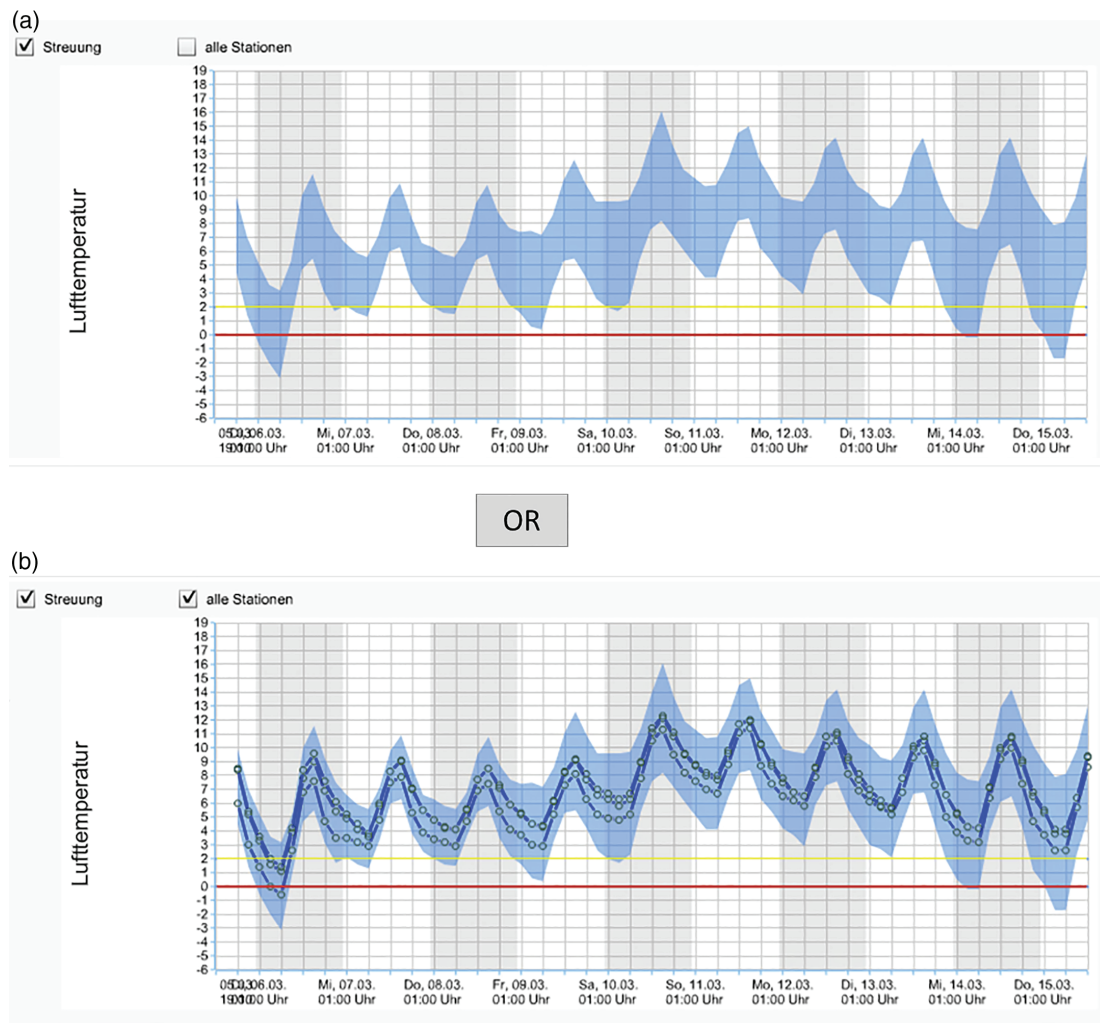
In the future, our aim is to consistently show quantitative probabilistic forecasts for the entire forecast horizon from short- to long-term. Currently, deterministic and probabilistic forecasts are available in parallel because users have not agreed to entirely replace the existing short-term deterministic forecast with the quantitative probabilistic forecasts.

### 2.2.3 | Evaluating the usefulness of the forecasts: Lessons learned

There are tentative, qualitative conclusions we can draw from the discussions during the workshop, the feedback users sent during the subsequent winter season, and from the follow-up survey.

Users clearly appreciated the collaborative nature of the development process and the opportunity to help design new products. They valued our meteorological presentation about forecast uncertainty, as well as the extensive time for questions and discussions during the workshop.

The workshop was useful for preventing misunderstandings of probabilistic forecasts and addressing user objections



**FIGURE 6** Example of probabilistic long-term temperature forecasts within SWIS. The transition from the short-term temperature graph in Figure 5 to uncertainty forecasts (as illustrated here) is only a minor visual change. (a) shows the default setting, which displays one envelope showing the range between the minimum and maximum of all lower and upper bounds of the inner 68%iles of the selected stations. In (b), the forecast mean temperature for each station is also shown (dark blue). The two user-specified, critical temperature thresholds (2 °C and 0 °C) are highlighted in yellow and red, respectively. [High-resolution images for all applications are available at the following Open Science Framework repository: <https://osf.io/k6nb7/>].

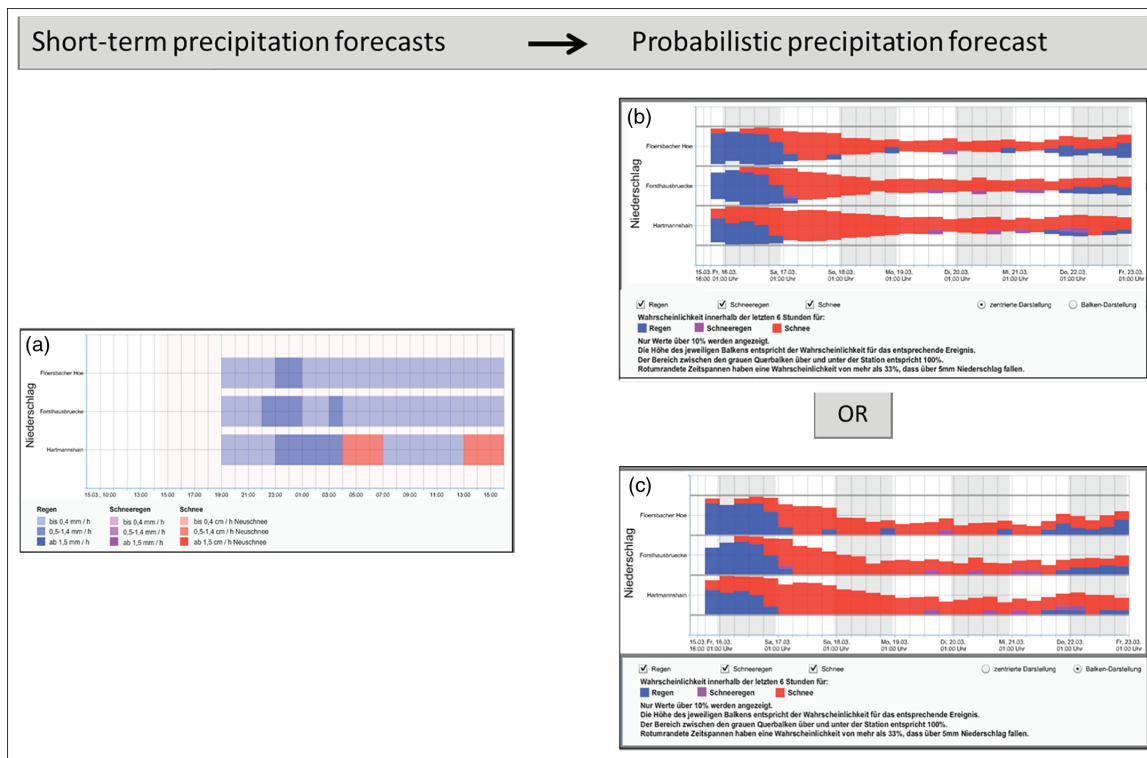
directly. For instance, the users' main concern with probabilistic forecasts was how to get from a range of possible outcomes to a deterministic decision. Here, the economic games using the jointly defined trial case helped get even reluctant users on board and to collaboratively develop a visual representation of the probabilistic forecasts that support their decision processes. Moreover, during the discussion users reacted more positively towards probabilistic information when it was framed as being about the "confidence" in the forecast rather than about the "uncertainty" of the forecast.

The games and the accompanying discussions suggested that users had a good basic understanding of probabilistic forecasts. However, they chose to take action at fairly low thresholds. When asked for a threshold directly, users agreed that a probability of 10% for snow was sufficient to schedule stand-by personnel. From what we know about the cost-loss ratio for this particular decision, we would have anticipated a higher threshold as the costs for waiting and calling additional staff at short notice are still not that high, according to earlier

bilateral conversations with some users. One potential explanation, suggested by user comments, is that they are generally used to act in a risk-averse manner in order not to miss events (e.g. snowfall or ice) and thus endanger the lives of road users. An alternative explanation may be that decisions were hypothetical without real costs or consequences. Moreover, users had no opportunity yet to experience the probabilistic forecasts together with the corresponding weather conditions when making their decisions in reality.

Interestingly, users' thresholds after testing the probabilistic forecasts back at their depots over the winter season markedly differed from the thresholds they selected during the workshop. In the follow-up survey ( $N = 20$ ), no single user reported having taken action (i.e. increasing stand-by weekend personnel) at a 10% probability of snowfall. The reported thresholds at the end of the winter season actually ranged, fairly uniformly, from 20% to 80%. Despite the variance, the higher thresholds at least suggest that users adjusted the thresholds based on their day-to-day experience.





**FIGURE 7** The evolution from short-term deterministic precipitation forecasts towards probabilistic precipitation forecasts in SWIS. Short-term deterministic forecasts for several user-defined locations (y-axis) train the user to consider different forecast scenarios (a). Colours indicate different states of the expected precipitation (rain, snowfall, sleet); the intensities of the colours indicate the respective amounts. Medium-range probabilistic forecasts are shown as centred around a baseline (b) or as a bar chart starting at the bottom of the y-axes (c). [High-resolution images for all applications are available at the following Open Science Framework repository: <https://osf.io/k6nb7>].

The greatest challenge that emerged was a strong pushback against the cost-loss concept (Murphy, 1977; Richardson, 2001) for most of the decisions taken by road workers beyond the trial case. Road workers work daily to prevent accidents. They are reluctant to assign numbers to costs and losses for the associated decisions, thereby ultimately putting a price on a human life (cf. “taboo trade-offs”; Tetlock, 2003). This is presumably why we were not able to put values to the costs and losses of the trial case decision problem during the workshop. We will address this issue in upcoming regular meetings in order to better understand the users’ perspective and to then jointly develop a strategy on how to include probabilistic forecast information into those decisions.

Overall, our main success with this user group was the positive feedback on the new forecast products from a large number of users. They use the probabilistic forecasts daily, and have even recently requested that the system show probabilistic forecasts for more road weather parameters, in particular for road temperature and humidity. We take this as a clear sign that this user group has accepted probabilistic forecasts and found them useful for their work.

### 2.3 | Probabilistic forecasts for transmission system operators

Germany is a leader in producing renewable energy from solar and wind power. The shift towards renewable energy

means that weather increasingly affects electrical power generation and transmission in Germany and beyond. First, weather variability implies variability in solar and wind energy production. As a result, the total electricity demand needs to be covered by a variable mix of conventional and renewable energy sources. Second, the weather influences the functionality of the energy grid system itself (e.g. transport capacity in overhead power lines). These weather dependencies create a number of new challenges for transmission system operators (TSOs) in particular and the entire electricity market in general. Wind and solar power forecasts are now key to anticipating and regulating a variety of power generation and transmission processes in order to secure both a constant energy supply and the stability of the energy grid.

As a result, the energy sector has new and growing needs for information from weather services. In order to specify those needs and optimise weather and climate products, the DWD is engaged in several interdisciplinary research projects that aim at optimising DWD’s numerical weather prediction models and developing forecast products tailored to user needs. As the DWD’s duty is to support operators of critical infrastructures, we focus in these projects on the requirements of the four German TSOs.

All these research projects are organised in consortia bringing together the DWD, power forecast service providers and experts from the TSOs. Both power forecast service providers

and TSOs are highly educated, mostly in physics, mathematics, or engineering. Power forecast providers are trained to develop numerical and statistical models for energy applications. TSOs are used to analyse complex models and data and to make far-reaching decisions under time pressure.

The task of power forecast service providers is to transform meteorological forecasts into power production forecasts, providing wind and solar power forecasts on different spatio-temporal resolutions and horizons and offering tools and products to support TSOs and the entire electricity market in their decisions. They input weather information into their statistical and physical power models and follow-up applications (e.g. virtual power plants). In our research projects, their task is twofold: they develop techniques that correctly propagate weather-dependent uncertainties through the power models and follow-up applications and they quantify other sources of uncertainty within their applications. They also collaborate with TSOs and the DWD to develop solutions for implementing probabilistic information in applications for TSOs.

TSOs feed the power forecasts from the service providers into their (currently deterministic) grid models and forecast tools to inform decisions on energy trading, plant operation, and grid operations that ensure grid stability and system security. TSOs also use weather information directly to dynamically adapt the transmission capacity in overhead power lines to weather conditions (“dynamic line rating”), in high-impact weather situations, and as additional information when power forecasts differ significantly across providers.

As soon as updated forecasts are available, TSOs can select from a number of procedures to correct for power forecast errors and adjust their decisions online. As a result, grid stability and security have to date not been jeopardised, even in the case of large forecast errors. However, each correction incurs additional costs, such as the costs of procuring expensive balancing energy or compensating wind and solar power plant operators when TSOs had to down-regulate the energy that plant operators feed into the grid. Moreover, with the planned increase of installed wind and solar power in Germany, the vulnerability to forecast errors will rise. As the economic pressure on TSOs and the potential for forecast errors to endanger network stability increases in the near future, users see quantifying the uncertainty associated with weather and power forecasts and propagating them through downstream models as the right way forward.

### 2.3.1 | Introducing forecast uncertainty

Our starting point for introducing probabilistic forecasts was quite favourable, as TSOs were already aware of forecast error and intended to address it with new forecast information, ultimately adapting their decision procedures. For the energy sector as a whole, the research community already provides a wealth of methods and ideas for using forecast uncertainty to increase economic benefit and the reliability of the power system (e.g. Pinson *et al.*, 2007; Matos and Bessa, 2011; Zhou

*et al.*, 2013). Although several tasks have been identified that can profit from probabilistic forecast information, they are so far predominantly at the power plant level and address plant operators and power traders (e.g. Bessa *et al.*, 2017).

The duties of TSOs, in contrast, are more manifold and require forecasts on a variety of spatial scales. Moreover, the vulnerability to forecast errors and mitigation possibilities differs across tasks. Thus, a successful integration of new probabilistic weather forecasts into existing regulation processes requires a close dialogue with both power forecast service providers and TSOs. While the tasks and associated decision processes are numerous and complex, the total TSO user group in Germany is rather small, which made it possible for us to collaborate closely with experts from all four German TSOs. During regular meetings and workshops with all project partners, we first gathered the weather-dependent decision processes that are necessary to ensure grid stability and energy supply. One challenge that emerged is the high number of continuously updated decision processes for TSOs. These processes are currently designed for deterministic forecasts; without automatic decision support, manually converting probabilistic forecasts into categorical decisions would overburden the users. Although the TSOs recognised the need for a paradigm shift, this requires new software solutions to optimally process probabilistic and uncertainty input.

Before introducing probabilistic forecasts to grid operations, we therefore needed to first analyse the vulnerability to forecast errors and the potential benefit of probabilistic forecasts separately for each task and work towards an optimal integration into existing (automatic decision) processes. As most of this work is still in progress, we report here on ongoing activities and our approach for advancing the use of probabilistic information. We divide the tasks identified by TSOs into three groups: power forecast applications, weather-dependent grid operation, and critical weather conditions for the transmission grid. For the first two applications, forecasts need to be directly fed into automatic applications. For the third, forecasts are provided as weather warnings and additional weather information.

*Power forecast applications.* TSOs must specify the expected generation of renewable energy that will be fed into their system. Each TSO needs a spatially aggregated forecast for their control area and for the whole of Germany to balance generation and consumption, identify power procurement, anticipate and prevent congestion, and trade renewable power at day-ahead and intraday energy exchange markets. Along with uncertainties in the weather and power prediction models, further uncertainties arise from incomplete knowledge of installed capacities of wind and solar power plants and from unknown regulating activities of plant operators and distribution system operators upstream. Ideas on how to integrate probabilistic forecasts into trading strategies are currently only available for power producers acting on a power plant scale (e.g. Pinson *et al.*, 2007; Bessa *et al.*, 2017). For trading activities, TSOs therefore use forecasts from several service

providers that are spatially aggregated to reduce these uncertainties through averaging effects. To this end, TSOs also asked for improved weather and power forecasts rather than for probabilistic information. The main problem is that for critical weather situations (e.g. low stratus, timing and localisation of cyclones) the predictions of all weather models suffer from similar large systematic errors. The challenge for the meteorological community is to reduce these biases in order to provide sufficiently reliable forecasts for these weather situations.

Power forecasts are also needed to simulate expected load flow, congestion, and contingencies in the transmission grid. These grid simulations aim at finding an optimal and secure network topology given the expected feed-in and load fluxes. Thus, they require spatially explicit power forecasts to simulate the conditions at grid nodes and transformer stations and, ultimately, for the entire grid. At these spatial scales, it becomes important to consider uncertainties in the weather and resulting power forecasts. TSOs are particularly interested in integrating possible weather and power scenarios into their simulations and operational security analyses. For this application, the challenge is to generate useful power scenarios derived from ensemble weather predictions as input for all feeding points in the grid and, even more importantly, to develop strategies for converting these scenarios into preventive measures if necessary.

*Weather-dependent grid operation processes.* The intermittent, decentralised, and spatially distributed power generation results in changing transmission grid requirements. For instance, the power generated from wind turbines in the north and east of Germany has to be transported to the industrial plants in the south. In order to complement the expensive, long-term grid expansion necessary to meet these requirements, TSOs are obligated by governmental authorities to maximise the transmission capacities of overhead power lines. One solution is a dynamic line rating, where load is increased under ambient weather conditions that are favourable for transmission capacities, and reduced under unfavourable conditions (Michiorri *et al.*, 2015). To calculate transmission and conductors' ampere capacity, weather forecasts and associated uncertainties are necessary for each span of the line. Low winds are particularly important, as they do not cool down the lines as much as higher winds do and thus reduce the lines' capacity. Given the estimated capacities, a risk policy then defines the maximum allowed probability for a line's capacity to be exceeded. From the probabilities and the overhead line model results, the grid operator derives the maximum current that ensures that conductors will not be damaged and that the lines do not sag below a critical point. For this application, the challenge for the weather service is to provide high-quality probabilistic weather forecasts at very small spatial scales.

*Critical weather conditions for grid operators.* To further understand how vulnerable TSOs are to weather forecast errors in their trading activities and to identify critical weather

conditions for which TSOs could benefit from probabilistic forecasts, we established an on-demand messaging system for TSOs to report cases where forecast errors resulted in high costs for correction procedures. Our project staff then characterised the weather situations at the time of each incident and evaluated the performance of the forecasts. With a large number of cases collected over the entire project duration of four years, we classified the critical weather conditions and developed tailored probabilistic forecast products for visualisation. TSOs use these weather forecasts as an additional guideline for their trading decisions, yet (so far) without a standardised procedure on how to incorporate the uncertainty information.

In order to address weather-related hazards to the transmission grid, TSOs throughout Europe have jointly specified weather conditions and criteria that could result in multiple losses of grid elements ("exceptional contingencies"; European Commission, 2017). When these weather conditions are sufficiently likely, TSOs perform operational security analyses that take these potential losses into account. Critical weather conditions include high wind speeds or conditions under which ice could accumulate on overhead lines.

Unlike the first two application areas, these forecasts and weather warnings are not directly fed into automatic applications. Rather, they aim to support the TSOs in a number of decisions concerning the trading activities, the power procurement or the plant schedule. To provide users with direct and easy access to this additional weather information, we developed a demonstration tool to enhance the developer–user dialogue. It displays real-time forecasts and allows grid operators to test and experience the new forecasts under critical weather conditions.

### 2.3.2 | Visualisation of probabilistic forecasts: develop a demonstration tool as testbed

This newly developed demonstration tool gives users hands-on experience with new forecasts and allows them to rapidly test the forecasts under real conditions. All collaboration partners in our research projects can access this tool and learn to use it during workshops and tutorials.

The tool presents visualisations of four types of forecasts from the limited-area COSMO-D2 ensemble prediction system (Gebhardt *et al.*, 2011; Peralta *et al.*, 2012): probabilistic wind forecasts at typical hub heights, probabilistic radiation forecasts, and – as a result of the analysis of critical weather situations – forecasts for low stratus risk (Koehler *et al.*, 2017), as well as forecasts from a cyclone detection algorithm (Steiner *et al.*, 2017). These forecasts aim at providing TSOs with uncertainty information that are as yet not included in their automatic applications. Although the uncertainty in weather forecasts is only one part of the total uncertainty these applications ultimately need to account for, we hope to sharpen the awareness for the possible consequences of significant uncertainties. For example, an uncertain weather

forecast in a specific area could result in a more conservative power procurement strategy than it would from using a deterministic forecast.

The tool also displays the DWD's official weather warnings. The next important development step is to implement the decision thresholds at which the exceptional contingencies security analyses have to be performed. With these forecasts, our goal is to directly inform the decision process for when to perform the special, fail-safe security analyses in order to prepare the grid for weather-induced losses.

Figure 8 shows an example of how the mean and spread of the radiation forecast is represented simultaneously by colour and hatching, respectively, in order to call the user's attention to areas for which forecasts are particularly uncertain.

To reduce the cognitive load evoked by multidimensional data, we developed a traffic light system, displayed in the upper-right corner, that spatially aggregates the forecasts shown on the map, reducing the values to three predefined categories. Our aim here is to provide a quick overview of potential weather-related problems, which was requested in the iterative development process with the users. In the case of a critical forecast, a user can zoom in for more detailed information from maps for the area of interest and for the entire forecast horizon.

As a starting point, we used discussions with the users as a basis for setting probability thresholds and the values for defining the warning levels for the traffic light system. As grid operators become more familiar with the new forecasts, we will encourage them to develop standardised strategies for using the probabilistic forecasts for specified decisions (e.g. necessary power procurement). By doing so, the users will then be able to evaluate how beneficial the forecasts are to their decisions. We hope this will also allow us to refine the implemented thresholds for the traffic light system based on user feedback and suggestions collected in future project meetings and discussions.

### 2.3.3 | Evaluating the usefulness of the forecasts: Future challenges

Our work with the TSOs and power forecast service providers is still in progress and the developments outlined here will need to mature before we are able to draw practical conclusions about our approaches to introducing probabilistic forecasts. Nevertheless, we can already summarise some key points characterising this user group and its requirements, as well as identify the most important next research steps and the main future challenges.

First, most of the decisions the TSOs identified have purely financial implications. Some decisions are associated with very high financial risks, such as compromised grid security or a deficient electricity supply. Yet the correction measures TSOs can use to adapt their decisions also imply significant costs, which raises the economic pressure to optimise the decision processes and with this the underlying information and decision tools. Here, the challenge is to quantify the costs

and losses for each decision in the face of a large number of different possible loss scenarios in order to calculate useful probability thresholds.

Second, the reason probabilistic forecasts are not already being implemented for all suitable decisions lies in the complexity of the applications and the resulting high demands on the software needed. This issue is reflected by the large number of current research projects addressing energy–meteorological problems. TSOs are aware of the limitations of their decision processes and are seeking appropriate solutions; for example, first attempts exist to extend the operational security analyses by a number of weather and power scenarios. The challenge here for meteorologists and power service providers is to contribute to the paradigm shift taking place at TSOs by developing models and tools that use probabilistic forecasts and by optimising decisions based on probabilistic forecasts.

Third, new automatic applications will have to be tested prior to their introduction in operational practice. This will require a simulation environment and software that includes both the existing, deterministic application and the new application that is based on probabilistic forecasts. Given the multitude of potential applications and their interdependencies, the challenge for TSOs will be to develop suitable evaluation strategies and a robust migration plan.

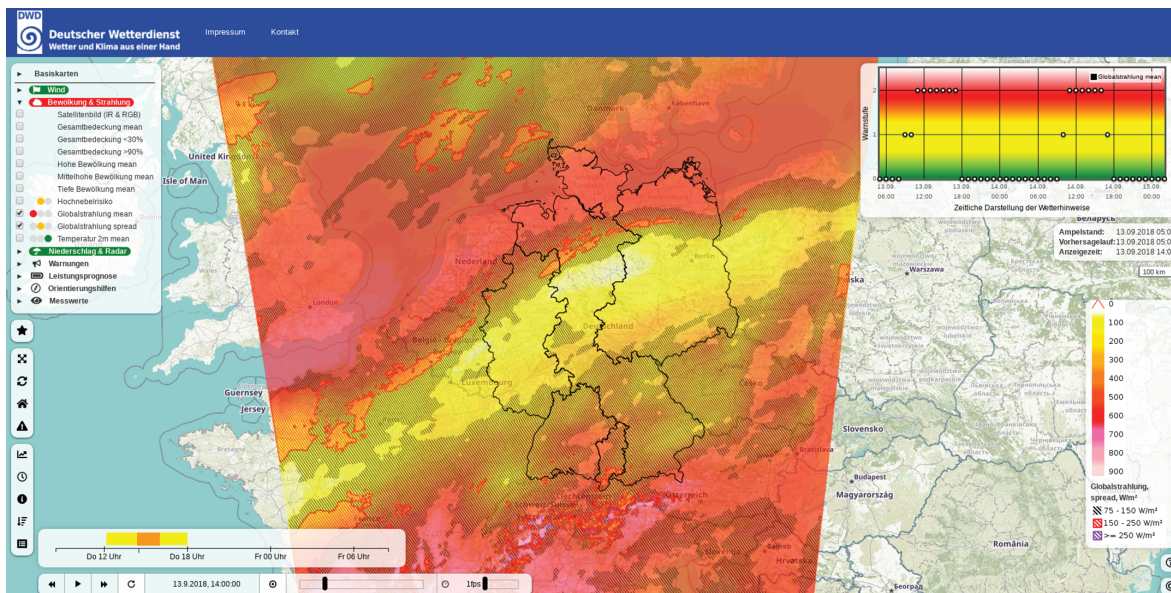
Fourth, our demonstration tool for visualisations of forecasts that are critical to grid stability will improve as we implement more user requirements, in particular the criteria for exceptional contingencies. In order to then evaluate the benefit of these new warnings, we need to develop ways to compare day-to-day practice with and without the visualisation tool. The challenge here will lie in compiling a robust database that includes the rather rare critical weather conditions upon which the decisions are based.

## 3 | DISCUSSION

How can users deal with the uncertainty inherent in weather forecasts? They can rely on deterministic forecasts and try to make decisions based on their own experience of forecast uncertainty. Because a single person's experience is limited, they will often have to take corrective actions as soon as updated forecasts are available. This approach usually results in additional costs (e.g. for redispaching fire engines, last-minute increases in on-duty personnel, or adjusting the feed-in to the power grid). Rising economic pressures in the public and private sectors make such costs increasingly problematic. Users are therefore looking for alternative strategies to proactively deal with the uncertainty in weather forecasts.

One such strategy is for users to make decisions based on probabilistic forecasts produced by ensemble prediction systems or statistical models. The main aim of this article was





**FIGURE 8** Simultaneous visualisation of mean global radiation and associated spread of the forecasts on our interactive demonstration webpage, based on the COSMO-D2-EPS model. The operational implementation of a calibration process for these is under way (Ben Bouallègue *et al.*, 2016; Ben Bouallègue, 2017). The coloured areas on the map indicate the expected radiation level; the shaded areas indicate where radiation forecasts are associated with high uncertainty. In the upper right corner, a traffic-light system shows the spatially aggregated time series of warning levels. The user can interact with the map, browse through the forecast horizon, or zoom onto the area of interest. The time series in the traffic-light system is coupled with the spatial radiation forecasts and enables the user to directly jump to a critical time step in the map. [High-resolution images for all applications are available at the following Open Science Framework repository: <https://osf.io/k6nb7/>].

to discuss our experience with different approaches to introducing and communicating probabilistic forecasts, to note the strengths and limitations of these approaches, and to highlight theoretical and practical challenges that emerged.

### 3.1 | How to introduce uncertainty?

Our three practical examples illustrate that there is no one-size-fits-all solution. Each user group and each practical application must be considered individually when introducing probabilistic forecast information.

The approach that is most appropriate in a given case will depend on a number of factors, such as users' prior experience with and attitude towards probabilistic forecasts, or the kind and number of decisions that may benefit from probabilistic forecasts. Moreover, the heterogeneity and size of a user group constrains how far results from one sample of the group may generalise towards the whole group.

In our experience, users without prior experience with probabilistic forecast, such as emergency managers and road workers, may view probabilistic forecasts with both interest and hesitation. As the case of road workers illustrates, this hesitation does not necessarily reflect a comprehension problem, but rather a concern around how to translate a probabilistic forecast into a decision. This concern becomes stronger if lives depend on the users' decisions, as in the example of the road workers or emergency managers.

The first two examples explored two very different approaches to overcoming this hesitation and introducing probabilistic forecasts. The approach for emergency managers

was to provide different representations of probabilistic forecasts in parallel without tailoring them closely to a particular decision. The strength of this approach is that it creates a large observational database of user behaviour under real operational conditions, making it possible to observe and objectively quantify users' actual revealed preference instead of relying on self-reports. This is particularly useful with large, heterogeneous user groups such as the emergency managers. By collecting data continuously on a long-term basis, we can study how preferences for various representations of forecast uncertainty differ between regular and severe weather conditions, and how preferences evolve more generally over time as users become more familiar with the way forecast uncertainty is communicated. As most users in this type of group are not introduced to the new forecasts in person, this approach must include carefully designed explanations and follow-up tests to ensure comprehension and elicit the decisions made based on the information. Moreover, it calls for some prior knowledge about the users' requirements, such as appropriate lead times, to ensure the information they receive is relevant. In this case, this was already available from previous studies (Kox *et al.*, 2018a; 2018b) and DWD staff's interactions with users; in other cases, surveys or qualitative research methods may be necessary as a first step to understand the context in which decisions are made (e.g. Creswell and Creswell, 2017 give information about qualitative research methods).

The approach for road workers, in contrast, relied on close, personal developer–user contact and a heavily guided introduction to the new probabilistic forecast products. This

participatory approach proved helpful for understanding users' decision processes and constraints, and for developing a common language, mutual understanding, and a successful, sustainable cooperation. Users' critical questions could readily be answered and misconceptions could be clarified in a timely fashion. Engaging users and jointly defining trial cases in the workshop helped them integrate new forecast products into their decision processes in the test phase afterwards; tailoring the forecast to a particular decision jointly with users can open the door to a wider adoption of probabilistic forecasts by the whole user group. Typically this type of approach is only feasible with smaller samples of users. Yet the preferences identified in a small sample of users will only represent the preferences of other users if the members of the whole group and their tasks are sufficiently homogeneous. It is also important to note that in such close dialogues between researchers, developers and end-users, all sides necessarily – and by intention – mutually influence each other. This means that the researchers involved are not “objective” observers detached from their object of study (Slocum-Bradley, 2005; Fouché and Light, 2011; Ritchie *et al.*, 2013 give more information about participatory qualitative research).

The challenges that arose in the case of the transmission system operators were quite different from the other two applications. There were no major challenges associated with users' knowledge or willingness to use probabilistic forecasts. Users had a highly technical background and were familiar with forecast uncertainty from other forecast models. Furthermore, they were aware that research suggests that probabilistic forecasts lead to more reliable power systems and economic benefits, and were therefore already looking for these forecasts. Moreover, the stakes, while high, were mostly purely financial. Instead, challenges arose due to the highly automated, complex systems and the multiple tasks and processes that may profit from probabilistic forecasts. To address these issues, we relied on a close collaboration of expert teams from DWD, power forecast providers and grid operators, and a demonstration tool for all newly developed applications so they can be reviewed under real conditions by the end-users.

### 3.2 | How are probabilistic forecasts received by the users?

A key insight that emerged across all user groups and approaches was that users need to experience probabilistic forecasts firsthand. This seems true even – indeed, especially – if users are initially sceptical about the potential benefits of probabilistic forecasts, such as road workers. Probabilistic forecasts must be ready to use under real operational conditions in order for users to gain experience with the products themselves, integrate them into their decisions, and evaluate which information is useful for a particular decision or situation (Hogarth and Soyer, 2015; Hertwig *et al.*, 2019; Lejarraga *et al.*, 2016; Wulff *et al.*, 2018 discuss the importance of experience in decision making).

The discrepancies we noticed between self-reported needs and how users actually deal with uncertainty in their day-to-day work are a strong case in point. For instance, the emergency managers initially reported in a survey that they, on average, can work with lead times of only 6 hrs (Kox *et al.*, 2015), but in practice they used probabilistic forecasts even earlier than 12 hrs before an event. This indicates that if longer-term probabilistic forecasts are made available, users may find them useful despite their initial belief to the contrary, and even adapt their decision routines. Moreover, the probabilistic forecasts may be used for decisions that might not have been suggested otherwise; for example, emergency managers used the probabilistic forecast during a severe weather event to plan post-event duties.

Similarly, the road workers in our second example stated during the introductory workshop that they would call for additional personnel as soon as the probability for snowfall reached 10%, but in the survey they completed after the winter season they indicated much higher probability thresholds. When asked to indicate a probability threshold before the probabilistic forecast product is introduced, users can only resort to whatever prior considerations about the decision problem they bring to the table. Once probabilistic forecasts are available in practice, however, users can adapt their probability thresholds based on their actual experience with the forecasts.

Whether the thresholds will converge on an appropriate value for a given user is an empirical question that is difficult to answer without knowing users' utilities for different outcomes, i.e. how important it is for users to prevent or ensure particular outcomes. However, the unwillingness to assign values to human life points to a general limitation of the cost–loss model for evaluating the benefit of probabilistic forecasts (Lazo, 2010), as it requires putting all outcomes on the same monetary or utility scale. Although economics and psychology offer methods for systematically eliciting the utility of outcomes that do not have a market price, these typically measure decision makers “willingness to pay” (cf. “contingent valuation” methods; Carson, 2012) to prevent or ensure a certain outcome. Yet for taboo trade-offs (Tetlock, 2003), people may refuse to name a price, or the price they give may not reflect a stable preference (Hausman, 2012). For road workers, for instance, ultimately lives may be at stake when they decide whether or not to salt a road. Moreover, such non-economic stakes are not necessarily reflected in decision experiments used to understand the benefits of probabilistic forecast (e.g. using a road salting task; Joslyn and LeClerc, 2013).

An alternative for evaluating the benefit of probabilistic forecasts are field experiments (Levitt and List, 2009) that compare objective outcome measures before and after the introduction of new forecasts; or that compare a group using these forecasts to the outcomes of a comparable control group not using them. However, these approaches require access to actual data (e.g. about costs and accidents)

on a larger scale, and due to their methodological challenges may be most feasible in interdisciplinary research collaborations.

### 3.3 | How to communicate forecast uncertainty?

As with the approach to integrating probabilistic forecasts into decision-making processes, there is no single best way to communicate probabilistic information (Fischhoff *et al.*, 2012). Users differ in their background and decision processes, which require distinct types and amounts of information (Raftery, 2014).

Thus the first important point to consider is which kind of information best fits the decisions that have to be made. For instance, does a decision depend on the particular probability that a threshold is exceeded, such as the probability of (any) snow in the case of thresholds for road workers, or does it depend on the possible range, as in the case of emergency workers who used boxplots during a storm as information when the wind may decrease to plan post-event duties? A second important point is that less can be more. Users may appreciate products that simplify and summarise – for instance via a traffic light system, as in the case of the grid operators, or by providing an overall view as in the case of the road workers – yet allow the user to access more detail when needed. Today, interactive platforms can permit users to choose between different representations of a forecast to inform different decisions (as in case of the emergency managers), rather than combining everything in one complex representation or offering only one representation for all decisions. As we currently lack research testing the benefit of interactive representations for the comprehension of uncertainty, future studies in this area would be particularly valuable (Spiegelhalter *et al.*, 2011). As a third point, we found it useful to build on what is familiar to users, especially if they lacked experience with probabilistic forecasts. In the case of the road workers, we found that displaying uncertainty qualitatively at first can pique users' interest in numerical uncertainties. Online information systems make it easy to provide users with both the quantitative forecast and its interpretation simultaneously (e.g. using mouseovers as in the application for emergency managers). Presenting quantitative information and its correct interpretation alongside familiar expressions, for instance by combining graphs and verbal probabilities, may make it easier for users to make the transition to the less-familiar quantitative information (Budescu *et al.*, 2014). Similarly, building on familiar representations, for instance of variability (for road workers) or probabilities for conventional thresholds (for emergency managers) may be useful as well.

A positive framing, such as referring to “confidence forecasts” instead of “uncertainty forecasts,” could also help users to overcome their hesitation; this was suggested by the case of the road workers. An interesting question for future research would be to test which framing is closest to the ways laypeople

talk about uncertainty in daily life (e.g. how sure something is, what the chances of something are, or how precisely one knows something), and whether the framing would foster lay users' comprehension of different ways to represent uncertainty.

Despite the need to tailor how forecast uncertainty is communicated to each user group and to each application, research on risk communication and information design offers examples of best practice that are broadly relevant and applicable to communicating forecast uncertainty in meteorology and other domains in general. In Section 2.1.2 we discussed five relevant recommendations as a first guide for developers:

1. Encode quantitative information in a way that fosters accurate decoding.
2. Explain probabilities, for instance through relative frequencies.
3. Prevent deterministic misinterpretations of forecast uncertainty.
4. Put rare but severe events into perspective.
5. Choose the level of detail in accordance with what can be reasonably predicted.

In addition, explanatory video clips were a promising way to direct attention to important aspects of the representations. Other tools, such as online games and quizzes, may also be effective; here, more research is needed to understand how interactive learning may help users grasp concepts (Spiegelhalter *et al.*, 2011).

### 3.4 | Interdisciplinary cooperation to meet future challenges

The paradigm shift from deterministic thinking towards working with probabilities in decision making is an ongoing but auspicious process. Although the dialogue between meteorologists and end-users requires considerable time and effort, it is necessary if practical applications are to profit from the valuable information in ensemble-based weather products. To this end, interdisciplinary exchange between meteorologists and social, behavioural, and cognitive scientists is essential. Meteorologists and developers will benefit from using the tools and methods available in social, behavioural and cognitive science to systematically test and develop practical applications. In turn, social, behavioural and cognitive research will benefit from study designs that reflect real-world settings and their challenges, such as the reluctance to put a price on life, or from opportunities for field experiments. We strongly encourage all sides to seek out such collaborations and provide sufficient resources to support them. Combining interdisciplinary expertise will not only improve practical applications but also advance the scientific understanding of how to communicate probabilistic forecasts and the increasing uncertainty across short-, medium- and long-term forecasts.



## 4 | CONCLUSIONS

The extensive research and development effort of the past 25 years in ensemble weather forecasting has resulted in valuable quantitative information about forecast uncertainty. Probabilistic information promises to be beneficial for a variety of decision-making processes.

The presented applications illustrate three key points:

1. Without probabilistic information, people can only guess the underlying uncertainty of forecasts. Communicating probabilistic forecasts is thus critical to support informed decisions by users with varying needs.
2. Probabilistic information can be understood by laypeople if representations are well-developed, tailored to user needs, and tested (Fischhoff *et al.*, 2012).
3. Users need the opportunity to use probabilistic forecasts in their actual work in order to experience potential benefits and develop probabilistic thresholds for their decisions.

With our discussion of three cases and the practical pointers to research insights from various fields, we hope to foster interest in interdisciplinary and evidence-based approaches, and in the broader use of the research tools and methods available across disciplines. A final challenge for promoting the use of probabilistic weather forecasts remains: how can we quantify the benefit of forecast uncertainty for users' real-life decisions in an objective way? Based on our experience, we are confident that the joint effort of interdisciplinary groups of scientists, developers, and practitioners can successfully meet this challenge in the future.

## ACKNOWLEDGEMENTS

VJF, NF and SMH equally contributed to the entire paper. NF, SMH and MG were responsible for the first application (FeWIS Pro; emergency managers) and VJF was responsible for the collaboration with the road workers and the transmission system operators. MG provided critical input and valuable feedback, and RH coordinated the joint effort and significantly added to the basis of this entire work.

We thank the users, the DWD's project partners from the Fraunhofer Institute for Energy Economics and Energy System Technology and from energy & meteosystems GmbH, and various members of DWD for the great collaboration, Jörg van Hoorn for programming (FeWIS Pro and SWIS) and Ralph Hertwig for his support. We also would like to thank Tim Palmer, Michelle McDowell, Felix Rebitscheck, Björn Meder and an anonymous reviewer for their constructive comments, and Deb Ain for editing the manuscript and her valuable suggestions.

## ORCID

Vanessa J. Fundel  <https://orcid.org/0000-0003-0619-6473>

Nadine Fleischhut  <https://orcid.org/0000-0002-6807-3455>

Stefan M. Herzog  <https://orcid.org/0000-0003-2329-6433>

## REFERENCES

- Ancker, J.S., Senathirajah, Y., Kukafka, R. and Starren, J.B. (2006) Design features of graphs in health risk communication: a systematic review. *Journal of the American Medical Informatics Association*, 13, 608–618.
- Barrio, P.J., Goldstein, D.G. and Hofman, J.M. (2016) Improving comprehension of numbers in the news. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 2729–2739. ACM Press, New York, NY.
- Bauer, P., Thorpe, A.J. and Brunet, G. (2015) The quiet revolution of numerical weather prediction. *Nature*, 525, 47–55.
- Ben Bouallègue, Z. (2017) Statistical postprocessing of ensemble global radiation forecasts with penalized quantile regression. *Meteorologische Zeitschrift*, 26, 253–264.
- Ben Bouallègue, Z., Heppelmann, T., Theis, S.E. and Pinson, P. (2016) Generation of scenarios from calibrated ensemble forecasts with a dual-ensemble copula-coupling approach. *Monthly Weather Review*, 144, 4737–4750.
- Bessa, R., Möhrle, C., Fundel, V., Siefert, M., Browell, J., El Gaidi, S.H., Hodge, B.-M., Cali, U. and Kariniotakis, G. (2017) Towards improved understanding of the applicability of uncertainty forecasts in the electric power industry. *Energies*, 10, 1402.
- Bodemer, N., Meder, B. and Gigerenzer, G. (2014) Communicating relative risk changes with baseline risk: presentation format and numeracy matter. *Medical Decision Making*, 34, 615–626.
- Borland, D. and Li, R.M.T. (2007) Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Applications*, 27, 14–17.
- Branković, C., Palmer, T.N., Molteni, F., Tibaldi, S. and Cubasch, U. (1990) Extended-range predictions with ECMWF models: time-lagged ensemble forecasting. *Quarterly Journal of the Royal Meteorological Society*, 116, 867–912.
- Broad, K., Leiserowitz, A., Weinkle, J. and Steketee, M. (2007) Misinterpretations of the “cone of uncertainty” in Florida during the 2004 hurricane season. *Bulletin of the American Meteorological Society*, 88, 651–668.
- Budescu, D.V., Por, H.-H., Broomell, S.B. and Smithson, M. (2014) The interpretation of IPCC probabilistic statements around the world. *Nature Climate Change*, 4, 508–512.
- Buizza, R., Houtekamer, P.L., Pellerin, G., Toth, Z., Zhu, Y. and Wei, M. (2005) A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review*, 133, 1076–1097.
- Carson, R.T. (2012) Contingent valuation: a practical alternative when prices aren't available. *Journal of Economic Perspectives*, 26, 27–42.
- Cleveland, W.S. and McGill, R. (1985) Graphical perception and graphical methods for analyzing scientific data. *Science*, 229, 828–833.
- Creswell, J.W. and Creswell, J.D. (2017) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publishing, Los Angeles, CA.
- Cumming, G. (2007) Inference by eye: Pictures of confidence intervals and thinking about levels of confidence. *Teaching Statistics*, 29, 89–93.
- Demeritt, D., Nobert, S., Cloke, H. and Pappenberger, F. (2010) Challenges in communicating and using ensembles in operational flood forecasting. *Meteorological Applications*, 17, 209–222.
- Ebert, E.E. (2001) Ability of a poor-man's ensemble to predict the probability and distribution of precipitation. *Monthly Weather Review*, 129, 2461–2480.
- European Commission (2017) Commission regulation (EU) 2017/1485 of 2 August 2017 establishing a guideline on electricity transmission system operation (Text with EEA relevance); Directorate-General for Energy, Brussels. <http://data.europa.eu/eli/reg/2017/1485/oj>; accessed 11 February 2019
- Fischhoff, B., Brewer, N.T. and Downs, J.S. (2012) *Communicating Risks and Benefits: An Evidence-Based User's Guide*. Government Printing Office, Silver Spring, MA.
- Fouché, C. and Light, G. (2011) An invitation to dialogue: “The World Café” in social work research. *Qualitative Social Work*, 10, 28–48.
- Frey, R., Pedroni, A., Mata, R., Rieskamp, J. and Hertwig, R. (2017) Risk preference shares the psychometric structure of major psychological traits. *Science Advances*, 3(10), e1701381.



- Frick, J. and Hegg, C. (2011) Can end-users' flood management decision making be improved by information about forecast uncertainty? *Atmospheric Research*, 100, 296–303.
- Gebhardt, C., Theis, S., Paulat, M. and Ben Bouallegue, Z. (2011) Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmospheric Research*, 100, 168–177.
- Gehlenborg, N. and Wong, B. (2012) Mapping quantitative data to color. *Nature Methods*, 9, 769.
- Gigerenzer, G. and Edwards, A. (2003) Simple tools for understanding risks: from innuery to insight. *British Medical Journal*, 327, 741–744.
- Gigerenzer, G. and Muir Gray, J.A. (Eds.) (2011) *Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020*. MIT Press, Cambridge, MA.
- Gigerenzer, G., Hertwig, R., Van Den Broek, E., Fasolo, B. and Katsikopoulos, K.V. (2005) A 30% chance of rain tomorrow: How does the public understand probabilistic weather forecasts? *Risk Analysis*, 25, 623–629.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L.M. and Woloshin, S. (2007) Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8, 53–96.
- Gill, J. (2008) Communicating forecast uncertainty for service providers. *WMO Bulletin*, 57, 237–243.
- Gneiting, T., Raftery, A.E., Westveld, A.H. and Goldman, T. (2005) Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 133, 1098–1118.
- Hamill, T.M., Whitaker, J.S. and Wei, X. (2004) Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Monthly Weather Review*, 132, 1434–1447.
- Harrower, M. and Brewer, C.A. (2003) Colorbrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40, 27–37.
- Hausman, J. (2012) Contingent valuation: from dubious to hopeless. *Journal of Economic Perspectives*, 26, 43–56.
- Hertwig, R., Pleskac, T.J., Pachur, T. and The Center for Adaptive Rationality (2019) *Taming uncertainty*. MIT Press, Cambridge, MA.
- Hildon, Z., Allwood, D. and Black, N. (2011) Impact of format and content of visual display of data on comprehension, choice and preference: A systematic review. *International Journal for Quality in Health Care*, 24, 55–64.
- Hoffrage, U., Lindsey, S., Hertwig, R. and Gigerenzer, G. (2000) Communicating statistical information. *Science*, 290, 2261–2262.
- Hogarth, R.M. and Soyer, E. (2015) Communicating forecasts: The simplicity of simulated experience. *Journal of Business Research*, 68, 1800–1809.
- Houtekamer, P.L., Lefaiivre, L., Derome, J., Ritchie, H. and Mitchell, H.L. (1996) A system simulation approach to ensemble prediction. *Monthly Weather Review*, 124, 1225–1242.
- Joslyn, S. and LeClerc, J. (2013) Decisions with uncertainty: The glass half full. *Current Directions in Psychological Science*, 22, 308–315.
- Joslyn, S.L. and Nichols, R.M. (2009) Probability or frequency? Expressing forecast uncertainty in public weather forecasts. *Meteorological Applications*, 16, 309–314.
- Joslyn, S.L., Nadav-Greenberg, L., Taing, M.U. and Nichols, R.M. (2009) The effects of wording on the understanding and use of uncertainty information in a threshold forecasting decision. *Applied Cognitive Psychology*, 23, 55–72.
- Koehler, C., Steiner, A., Saint-Drenan, Y.-M., Ernst, D., Bergmann-Dick, A., Zirkelbach, M., Ben Bouallegue, Z., Metzinger, I. and Ritter, B. (2017) Critical weather situations for renewable energies – Part B: Low stratus risk for solar power. *Renewable Energy*, 101, 794–803.
- Kosslyn, S.M. and Kosslyn, S.M. (2006) *Graph Design for the Eye and Mind*. Oxford University Press, Oxford, UK.
- Kox, T., Gerhold, L. and Ulbrich, U. (2015) Perception and use of uncertainty in severe weather warnings by emergency services in Germany. *Atmospheric Research*, 158–159, 292–301.
- Kox, T., Kempf, H., Lüder, C., Hagedorn, R. and Gerhold, L. (2018a) Towards user-orientated weather warnings. *International Journal of Disaster Risk Reduction*, 30, 74–80.
- Kox, T., Lüder, C. and Gerhold, L. (2018b) Anticipation and response: emergency services in severe weather situations in Germany. *International Journal of Disaster Risk Science*, 9, 116–128.
- Kozyreva, A. and Hertwig, R. (2019) The interpretation of uncertainty in ecological rationality. *Synthese*, 1–31.
- Kozyreva, A., Pleskac, T.J., Pachur, T. and Hertwig, R. (2019) Interpreting uncertainty: A brief history of not knowing, In *Taming uncertainty*. Hertwig R., Pleskac T.J., Pachur T. and The Center for Adaptive Rationality, MIT Press, Cambridge, MA, pp. 343–362.
- Lazo, J.K. (2010) The costs and losses of integrating social sciences and meteorology. *Weather, Climate, and Society*, 2, 171–173.
- LeClerc, J. and Joslyn, S. (2015) The cry wolf effect and weather-related decision making. *Risk Analysis*, 35, 385–395.
- Lejarraga, T., Woike, J.K. and Hertwig, R. (2016) Description and experience: How experimental investors learn about booms and busts affects their financial risk taking. *Cognition*, 157, 365–383.
- Levitt, S.D. and List, J.A. (2009) Field experiments in economics: The past, the present, and the future. *European Economic Review*, 53, 1–18.
- Lipkus, I.M. (2007) Numeric, verbal, and visual formats of conveying health risk: Suggested best practices and future recommendations. *Medical Decision Making*, 27, 696–713.
- Luce, R.D. and Raiffa, H. (1957) *Games and Decisions: Introduction and Critical Survey*. Wiley, New York, NY. Reprinted 1989 by Dover Books, Mineola, NY.
- Matos, M.A. and Bessa, R.J. (2011) Setting the operating reserve using probabilistic wind power forecasts. *IEEE Transactions on Power Systems*, 26, 594–603.
- McDowell, M. and Jacobs, P. (2017) Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin*, 143, 1273–1312.
- Michiorri, A., Nguyen, H.-M., Alessandrini, S., Bremnes, J.B., Dierer, S., Ferrero, E., Nygaard, B.-E., Pinson, P., Thomaidis, N. and Uski, S. (2015) Forecasting for dynamic line rating. *Renewable and Sustainable Energy Reviews*, 52, 1713–1730.
- Molteni, F., Buizza, R., Palmer, T.N. and Petroliagis, T. (1996) The ECMWF ensemble prediction system: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, 122, 73–119.
- Morss, R.E., Demuth, J.L. and Lazo, J.K. (2008) Communicating uncertainty in weather forecasts: A survey of the U.S. public. *Weather and Forecasting*, 23, 974–991.
- Murphy, A.H. (1966) A note on the utility of probabilistic predictions and the probability score in the cost–loss ratio decision situation. *Journal of Applied Meteorology*, 5, 534–537.
- Murphy, A.H. (1977) The value of climatological, categorical and probabilistic forecasts in the cost–loss ratio situation. *Monthly Weather Review*, 105, 803–816.
- Murphy, A.H. (1998) The early history of probability forecasts: some extensions and clarifications. *Weather and Forecasting*, 13, 5–15.
- Murphy, A.H., Lichtenstein, S. and Fischhoff, B. (1980) Misinterpretations of precipitation probability forecasts. *Bulletin of the American Meteorological Society*, 61, 695–701.
- Newman, G.E. and Scholl, B.J. (2012) Bar graphs depicting averages are perceptually misinterpreted: the within-the-bar bias. *Psychonomic Bulletin & Review*, 19, 601–607.
- Okan, Y., Garcia-Retamero, R., Cokely, E.T. and Maldonado, A. (2015) Improving risk understanding across ability levels: Encouraging active processing with dynamic icon arrays. *Journal of Experimental Psychology: Applied*, 21, 178–194.
- Okan, Y., Garcia-Retamero, R., Cokely, E.T. and Maldonado, A. (2018) Biasing and debiasing health decisions with bar graphs: Costs and benefits of graph literacy. *Quarterly Journal of Experimental Psychology*, 71, 2506–2519.
- 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, 747–774.
- Palmer, T.N. (2018) The ECMWF ensemble prediction system: looking back (more than) 25 years and projecting forward 25 years. *Quarterly Journal of the Royal Meteorological Society*.
- Palmer, T.N., Molteni, F., Mureau, R., Buizza, R., Chapelet, P. and Tribbia, J. (1992) Ensemble Prediction. Technical Memorandum 188, ECMWF, Reading, UK.
- Pardowitz, T., Kox, T., Göber, M. and Bütow, A. (2015) Human estimates of warning uncertainty: numerical and verbal descriptions. *Mausam*, 66, 625–634.
- Peralta, C., Ben Bouallegue, Z., Theis, S., Gebhardt, C. and Buchhold, M. (2012) Accounting for initial condition uncertainties in COSMO-DE-EPS. *Journal of Geophysical Research; Atmospheres*, 117(D7).
- Peters, E., Dieckmann, N.F., Västfjäll, D., Mertz, C.K., Slovic, P. and Hibbard, J.H. (2009) Bringing meaning to numbers: The impact of evaluative categories on decisions. *Journal of Experimental Psychology: Applied*, 15, 213–227.

- Pinson, P., Chevallier, C. and Kariniotakis, G.N. (2007) Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Transactions on Power Systems*, 22, 1148–1156.
- Raftery, A.E. (2014) Use and communication of probabilistic forecasts. *Statistical Analysis and Data Mining*, 9, 397–410.
- Raftery, A.E., Gneiting, T., Balabdaoui, F. and Polakowski, M. (2005) Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133, 1155–1174.
- Richardson, D.S. (2001) Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quarterly Journal of the Royal Meteorological Society*, 127, 2473–2489.
- Ritchie, J., Lewis, J., Nicholls, C.M. and Ormston, R. (2013) *Qualitative Research Practice*. Sage Publications, Los Angeles, CA.
- Roulston, M.S. and Smith, L.A. (2003) Combining dynamical and statistical ensembles. *Tellus A*, 55, 16–30.
- Roulston, M.S., Bolton, G.E. and Kleit, A.N. (2006) A laboratory study of the benefits of including uncertainty information in weather forecasts. *Weather and Forecasting*, 21, 116–122.
- Savelli, S. and Joslyn, S. (2013) The advantages of predictive interval forecasts for non-expert users and the impact of visualizations. *Applied Cognitive Psychology*, 27, 527–541.
- Schieffer, A., Isaacs, D. and Gyllenpalm, B. (2004) The world café: Part one. *Transformation*, 18, 1–9.
- Simmer, C., Adrian, G., Jones, S., Wirth, V., Göber, M., Hohenegger, C., Janjić, T., Keller, J., Ohlwein, C., Seifert, A., Trömel, S., Ulbrich, T., Wapler, K., Weissmann, M., Keller, J.K., Masbou, M., Meilinger, S., Riss, N., Schomburg, A., Vormann, A. and Weingärtner, C. (2016) HERZ: the German Hans-Ertel Centre for Weather Research. *Bulletin of the American Meteorological Society*, 97, 1057–1068.
- Slocum-Bradley, N. (2005) *Participatory Methods Toolkit. A Practitioner's Manual*. King Baudouin Foundation and the Flemish Institute for Science and Technology Assessment, Brussels.
- Slovic, P. (1987) Perception of risk. *Science*, 236, 280–285.
- Spiegelhalter, D., Pearson, M. and Short, I. (2011) Visualizing uncertainty about the future. *Science*, 333, 1393–1400.
- Stauffer, R., Mayr, G.J., Dabernig, M. and Zeileis, A. (2015) Somewhere over the rainbow: How to make effective use of colors in meteorological visualizations. *Bulletin of the American Meteorological Society*, 96, 203–216.
- Steiner, A., Koehler, C., Metzinger, I., Braun, A., Zirkelbach, M., Ernst, D., Tran, P. and Ritter, B. (2017) Critical weather situations for renewable energies – Part A: cyclone detection for wind power. *Renewable Energy*, 101, 41–50.
- Stephens, E.M., Edwards, T.L. and Demeritt, D. (2012) Communicating probabilistic information from climate model ensembles – Lessons from numerical weather prediction. *Wiley Interdisciplinary Reviews: Climate Change*, 3, 409–426.
- Taylor, A., Kox, T. and Johnston, D. (2018) Communicating high-impact weather: Improving warnings and decision making processes. *International Journal of Disaster Risk Reduction*, 30, 1–4.
- Tetlock, P.E. (2003) Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, 7, 320–324.
- Toth, Z. and Kalnay, E. (1993) Ensemble forecasting at NMC: The generation of perturbations. *Bulletin of the American Meteorological Society*, 74, 2317–2330.
- Toth, Z. and Kalnay, E. (1997) Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review*, 125, 3297–3319.
- Trevena, L.J., Zikmund-Fisher, B.J., Edwards, A., Gaissmaier, W., Galesic, M., Han, P.K., King, J., Lawson, M.L., Linder, S.K., Lipkus, I., Ozanne, E., Peters, E., Timmermans, D. and Woloshin, S. (2013) Presenting quantitative information about decision outcomes: A risk communication primer for patient decision aid developers. *BMC Medical Informatics and Decision Making*, 13(S7).
- Visschers, V.H.M., Meertens, R.M., Passchier, W.W.F. and De Vries, N.N.K. (2009) Probability information in risk communication: A review of the research literature. *Risk Analysis*, 29, 267–287.
- Weber, E.U. and Hilton, D.J. (1990) Contextual effects in the interpretations of probability words: Perceived base rate and severity of events. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 781–789.
- Weber, E.U., Shafir, S. and Blais, A.-R. (2004) Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review*, 111, 430–445.
- Wong, B. (2011) Color blindness. *Nature Methods*, 8, 441.
- Wulff, D.U., Mergenthaler-Canseco, M. and Hertwig, R. (2018) A meta-analytic review of two modes of learning and the description–experience gap. *Psychological Bulletin*, 144, 140–176.
- Yaniv, I. and Foster, D.P. (1995) Graininess of judgment under uncertainty: An accuracy–informativeness trade-off. *Journal of Experimental Psychology: General*, 124, 424–432.
- Zhou, Z., Botterud, A., Wang, J., Bessa, R.J., Keko, H., Sumaili, J. and Miranda, V. (2013) Application of probabilistic wind power forecasting in electricity markets. *Wind Energy*, 16, 321–338.

**How to cite this article:** Fundel VJ, Fleischhut N, Herzog SM, Göber M, Hagedorn R. Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers and end-users. *QJR Meteorol Soc.* 2019;145 (Suppl. 1):210–231. <https://doi.org/10.1002/qj.3482>