THE WEATHER ROULETTE

A Game to Communicate the Usefulness of Probabilistic Climate Predictions

Marta Terrado, Llorenç Lledó, Dragana Bojovic, Asun Lera St. Clair, Albert Soret, Francisco J. Doblas-Reyes, Rodrigo Manzanas, Daniel San-Martín, and Isadora Christel

We apply a game to communicate the usefulness of climate predictions to users, showing that in skillful areas economic benefits are obtained in the long term.

Seasonal-to-decadal climate predictions try to anticipate the most likely climate conditions for the next few months up to a decade into the future (Doblas-Reyes et al. 2013; Meehl et al. 2014). Sitting between weather forecasts (from the next few hours up to a few days ahead) and climate change projections (from a few decades up to centuries), climate predictions have the potential to inform different climate-sensitive sectors (e.g., energy, agriculture, water management, health, insurance, tourism) in adapting their short- to medium-term practices and plans to climate variability and change (Thomson et al. 2006; Jewson et al. 2009; Torralba et al. 2017; Turco et al. 2017). Climate-sensitive sectors can benefit from understanding climate predictions and how they can be used to make better informed decisions and thus gain strategic advance toward other competitors. However, despite their potential advantages and the recent efforts to develop underpinning science for climate predictions, so far there has been limited uptake of these tools by users (McNie 2007; Feldman and Ingram 2009).

Main barriers hindering users' uptake of climate predictions include (i) the lack of a common and widely accepted terminology between climate scientists and user communities, (ii) the difficulty to deal with probabilistic rather than deterministic outcomes, (iii) their lower skill (i.e., the quality of the prediction based on its performance in the past) compared to the skill of weather forecasts, and (iv) the need to move from a short- to a long-term approach for the assessment of benefits in the business sector, since the benefits from adopting climate predictions can only be perceived in the long term. Adding to these barriers, there is also little evidence of the use of climate predictions for operational applications (Coelho and Costa 2010), often ascribed

AFFILIATIONS: Terrado, Lledó, Bojovic, St. Clair, Soret, and Christel—Department of Earth Sciences, Barcelona Supercomputing Center, Barcelona, Spain; Doblas-Reyes—Department of Earth Sciences, Barcelona Supercomputing Center, and Institut Català de Recerca i Estudis Avançats, Barcelona, Spain; Manzanas and San-Martín—Predictia Intelligent Data Solutions SL, Santander, Spain
CORRESPONDING AUTHOR: Marta Terrado, marta.terrado@bsc.es

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/BAMS-D-18-0214.1

In final form 3 July 2019

to the users' difficulty to integrate predictions into existing decision support systems. In this sense, there is a need to improve the way in which actionable climate information is made salient and relevant to different users.

An important step toward encouraging the adoption of climate predictions for supporting decision-making consists in quantifying and communicating the potential economic value, either in terms of increased outcome, avoided cost, or vulnerability reduction. Different methods can be found in the literature, ranging from quantitative studies that focus on the technical aspects of forecasts to more qualitative and user-centered approaches (Bruno-Soares et al. 2018).

Games are a powerful way to facilitate a more thorough analysis of complex issues, transferring scientific information into understandable and tailored knowledge that is tacitly connected to the target audience (van Pelt et al. 2015). Therefore, game-based learning has seen promising results in different areas, including the field of climate science (Ramos et al. 2013; Vincent et al. 2017; van Pelt et al. 2015; Arnal et al. 2016; Crochemore et al. 2016). These works have used games to demonstrate the potential utility of probabilistic forecasts for taking better risk-based decisions, while also suggesting that greater attention needs to be paid to the communication of uncertainties. Although uncertainties constitute an added value of probabilistic over deterministic forecasts, they also present challenges for both forecasters and users of forecasts (Arnal et al. 2016). Indeed, forecast skill is one of these sources of uncertainty that needs additional communication efforts to be presented in a way that is well understood by users (Taylor et al. 2015). In this line Hagedorn and Smith (2009) developed the Weather Roulette (WR) conceptual framework, a simple and easy to understand approach for communicating the value of probabilistic weather forecasts.

The WR was first applied to communicate the value of probabilistic weather forecasts for the next few days (Palmer et al. 2005; Hagedorn and Smith 2009). However, the approach has recently been extended to multiannual hurricane predictions (Caron et al. 2018), and to seasonal predictions of temperature extremes (Lazenby et al. 2014). The WR approach interprets probabilistic predictions in terms of economic value, translating metrics commonly used by the scientific community (e.g., ignorance skill score) to other metrics more easily understood in the private sector such as return on investment.

In this paper, we provide an example of how gamification can overcome communication and understanding barriers for the uptake of probabilistic climate predictions. In a simplified context, we use a betting game based on the WR approach to demonstrate the efficiency of climate predictions compared to climatology (past observations). This is supported through the translation of skill scores into economic terms, which provides a different approach to communicate forecast uncertainty to users and allows comparing the potential economic value of climate predictions in skillful regions with regions of limited skill. Understanding the usefulness of climate predictions could provide the basis for a better integration of knowledge about climate anomalies into operational and managerial processes.

The Weather Roulette approach: From theory to practical implementation. To illustrate the practical application of the game, in the context of the European Provision of Regional Impacts Assessments on Seasonal and Decadal Time Scales (EUPORIAS) and the Subseasonal to Seasonal Climate Forecasting for Energy (S2S4E) projects, we developed the WR mobile app (Predictia 2019), which is specifically addressed to the wind energy sector—an industry with an increasing awareness of the need to understand climate risk (Lledó et al. 2018). Traditionally, climatology has been used for wind resource assessment on the ever less reliable assumption that what happened in the past will be representative of future conditions (Carta et al. 2013; Lledó 2017). However, recent advances in climate prediction have already shown that probabilistic forecasting, once tailored to the specific needs of users, could provide opportunities for better informed investments, improving risk assessments, and indicating the climate exposure of energy assets. Therefore, there are a number of wind energy companies that could become early adopters of climate predictions (Terrado et al. 2017).

The WR mobile app has been implemented as a communication and engagement tool that shows the potential benefits of climate predictions over climatology in the long term. The app has the aim to engage relevant users within energy companies to consider the uptake of climate predictions and foster interaction with climate scientists to create more complex and customized climate services that inform their decision-making (European Commission 2015; see sidebar).

THE WEATHER ROULETTE GAME FOR CLIMATE PREDICTION.

Data for the Weather Roulette. The WR is defined here as a game where a player chooses between two forecast options, aiming to select the one that predicts better in which category (i.e., a range of values) the observed value will fall. There are different possible categorizations. For the WR mobile app described in this paper, we have used tercile categories based on the historical climatology (above normal, normal, and below normal), which is a standard categorization for presenting seasonal predictions (Jupp et al. 2012). Other category settings more relevant to specific user decisions could be defined instead, such as quintile categories or even asymmetrical categories (i.e., below the 10th percentile, a central category between the 10th and 90th percentile, and above the 90th percentile).

The two forecast options considered in the WR game are option 1, which corresponds to the use of seasonal climate predictions, and option 2, corresponding to the use of climatological predictions (see description below). Observational data have also been used for comparison with predicted data.

Option 1—Climate predictions (calibrated ECMWF System 4 prediction system). Global information on seasonal variations of the wind resource are obtained from the RESILIENCE prototype (http://resilience.bsc.es), an interactive climate service interface for wind industry users developed as part of the European funded projects EUPORIAS (FP7; http://euporias.eu/) and CLIM4ENERGY [Copernicus Climate Change Service (C3S); http://clim4energy.climate.copernicus.eu/]. RESILIENCE uses the calibrated predictions coming from the 51-member ensemble version of the ECMWF System 4 seasonal forecasting system (Molteni et al. 2011). Winter wind predictions have both higher skill and variability in the Northern Hemisphere, and provide a good test case. Therefore, we focus exclusively on winter (DJF) predictions of surface (10 m) wind speed, initialized on 1 November for a period of 33 past years, from 1981 to 2013, at those locations with installed wind power capacity (n = 2,023) obtained from the windpower.net database. The technique of variance inflation (von Storch and Zwiers 2001) is selected for calibration and applied as in Doblas-Reyes et al. (2005); the reader is referred to Manzanas et al. (2019) for further details on the effect of calibration of seasonal forecasts. The percentage of probability for the different categories to occur is computed as the percentage of ensemble members falling within each category.

CLIMATE PREDICTION: WHAT WE SHOULD KNOW

The atmosphere is chaotic in nature and therefore becomes unpredictable after a few days. This is why weather forecasts only provide useful information up to a few days ahead. However, the atmosphere is forced by other components of the Earth system, namely, the ocean, land, and sea ice components that evolve much slower and are predictable at longer time scales. Climate predictions, which take into account these forcings, can be used to compute the likelihood of a certain outcome (e.g., having above-normal, normal, or below-normal wind speed conditions for the next season). This probabilistic nature often does not align with the expectations of users, who are more interested in a yes/no answer to whether they should implement or not a particular action. Therefore, the integration of probabilistic predictions into actionable decision-making constitutes an important challenge.

Besides its probabilistic nature, there are other aspects of climate predictions that should be considered and that, potentially, further limit their usability. Any probabilistic prediction should be accompanied by an estimate of its past performance, known as forecast verification, which can guide users about the expected performance of future predictions (Weisheimer and Palmer 2014). Forecast verification should address the accuracy—how close the forecast probabilities are to the observed frequencies; the utility—the economic or other advantages of the probabilistic forecasts; and the skill—how the probabilistic forecasts compare with a reference forecast (Jolliffe and Stephenson 2012). However, as the predictability of weather forecasts comes from initial atmospheric conditions, their skill is normally high at the beginning of the forecast period and experiences a fast decrease after a few days, whereas the skill of climate predictions is lower than that of weather forecasts and is kept more stable and decreases at a slower pace as lead time increases (White et al. 2017). The generally low skill exhibited by climate predictions in extratropical regions such as Europe has resulted in their limited practical applications (Doblas-Reyes et al. 2013; Manzanas et al. 2014). Apart from the region, useful skill also depends on the time of the year (e.g., the season), which further lowers the perceived reliability in these predictions (Bruno-Soares and Dessai 2016). It is also paramount to understand that a single prediction is not representative of the long-term performance of climate predictions, even in an area where the model has skill for the period of interest.

Option 2—Climatological predictions (probabilities derived from historical observations). The observed frequencies of occurrence of different categories in the historical records (ERA-Interim reanalysis) have been used as forecast probabilities. As already mentioned, climatology has been traditionally the preferred choice for the wind energy sector when assessing risks, and therefore has been set as the baseline to bet against.

Observational data. Reanalysis data from ERA-Interim (Dee et al. 2011) have been employed as truth for comparison with predicted data (forecast verification).

Skill of climate predictions. Different quality metrics are available for this task, often quantifying different characteristics of forecast performance. Here, two skill scores are calculated taking climatology as reference: the ignorance skill score (ISS; or logarithmic score; Good 1952) and the ranked probability skill score (RPSS; Wilks 2011). Both scores have been computed at the selected locations using the retrospective climate predictions and observational data described in the "Data for the Weather Roulette" section. The ISS has been considered because it possesses geometric symmetry and a correspondence with the WR approach. Therefore, it has a clear interpretation in terms of gambling returns, being easily communicated as an effective interest rate. On the other hand, RPSS is a widely used skill score in atmospheric science (Jolliffe and Stephenson 2012), and therefore, there is an increasing interest in bringing it closer to the user community. ISS is defined as an average of logarithms of the probabilities assigned to the observed outcome. It only takes into account the probabilities assigned to the observed or winning category; hence, it is technically defined as a local score (Mason 2008; Jolliffe and Stephenson 2012). Conversely, RPSS is not local; rather, it uses the probabilities assigned to all categories and the distance to the observed category to compute the verification, taking into account how big the probabilities predicted for the nonobserved categories are. Both ISS and RPSS range from 1 to minus infinity. Values above zero indicate that the verified seasonal forecasts perform better than a simple, constant prediction based on climatology.

The Weather Roulette game. The WR game [see Hagedorn and Smith (2009) for an extensive description of the method] is defined as a bet between two different forecast options: seasonal climate predictions and climatological predictions. The roulette slots represent the possible outcome categories that can contain the observation. An initial capital (c_0) is set, and every time all the capital is reinvested in the next round, with one round for each year. To start, the initial capital is spread in the different slots proportionally to the percentage probabilities predicted by each of the options (climate predictions and climatological predictions). The winning slot is then determined as the slot where the real observations fall. Then, for each option, payments are received proportionally to the bets in the winning category. The odds (i.e., the payoff to stake ratio) are inverse to the climatological probabilities for that category. The bet invested in the other categories is lost.

The code used to apply the WR game to the locations and years selected in this work was developed in R language (R Core Team 2015). The data described in the "Data for the Weather Roulette" section were used to run the code.

Translating skill scores into economic value. The WR can be played both for individual years (one round) and for the 33-yr period considered (33 rounds). Results are expressed in well-known economic measures (Table 1): (i) the return ratio for each individual round (r_t) calculated as the ratio between the capital obtained after and before playing the WR; (ii) the average or overall return ratio (R) for the 33 rounds, corresponding to the geometric average of r_t; (iii) the effective interest rate obtained for the full period played (IR, in %), which gives the annualized proportion of money earned each year over a given time period; and (iv) the return on investment (ROI), also for the full 33-yr period.

The added value of using climate predictions corresponds to the difference between the gains resulting from using climate predictions and the gains resulting from directly using a climatological constant prediction. Return ratios (r_t and R) larger than 1 indicate

Table 1. Definition and calculation of economic metrics used in the Weather Roulette approach.

Economic metric	Calculation
Initial capital (c_0)	Arbitrary value to be defined by the player
Number of rounds (n)	n = 33 (the number of DJF seasons in 1981–2013 period)
Final capital (c_n)	$c_n(\mathrm{EUR}) = c_0 r_1 r_2 r_3 \ldots r_n = c_0(R)^n$
Return ratio for each individual round (r_t)	$r_t = c_t/c_{t-1}$
Average return ratio for the whole period played (R)	$R = (r_1 r_2 r_3 \ldots r_n)^{1/n}$
Effective interest rate obtained for the full period played (IR)	IR (%) = (R – 1) × 100
Return on investment for the full period played (ROI)	ROI = $(c_n - c_0)/c_0$

gains. For instance, a value of 1.5 corresponds to a return of half the bet on top of that bet. A value of 1 indicates a neutral return (no gain and no loss) and values below one indicate losses. Note that for climatology, the return ratio is always 1, as the invested amount in the winning category is proportional to the climatological probabilities while the odds are inversely proportional to it. A positive IR indicates a net gain over the years, whereas net losses are indicated by negative IR values. The ROI indicates the net gains associated to an initial investment (c_0). These economic measures allow the immediate comparison of different prediction systems, and show which of the systems produce higher net gains after a certain period of time.

The Weather Roulette app. The WR game can be played from an interactive interface where the app simulates how much a player would have won or lost using either seasonal climate predictions or climatological predictions for decision-making. This allows for a comparison of the performance of both forecasts considering tercile categories for wind speed based on the historical climatology. At the beginning of the game, the user is presented with a global map with the distribution of the skill (ISS) and can select a particular location to start playing (Fig. 1a). For the selected location, the player can decide either to play for a single year among the period 1981–2013, or for all years (the game runs a forecast consecutively for each of the 33 years of the period). Together with the level of skill, the player is shown the probabilities predicted by the seasonal climate prediction for each of the tercile categories (Fig. 1b). Based on this information the player decides its preferred option to play the game. When playing for all years the skill value is shown and the preferred initial capital can be set by the player (Fig. 1c). Note that in the case of climatology all three categories are equally probable, with a probability of 1/3 each. After playing for a single year, the app returns the value of the return ratio (r_t) for that particular year. When the game is played for all years, the app returns the value of the effective IR and the ROI after the 33 years. The winning forecast option (seasonal forecasts or climatological forecasts) is reported at the end of the game and results obtained for both options are compared.

RESULTS. *Skill of seasonal wind predictions.* Global maps of ISS and RPSS have been calculated to assess the skill of winter wind speed predictions (Fig. 2a). Although the skill of seasonal predictions is in general low at extratropical latitudes (Manzanas et al. 2014), some positive skill is found in certain regions

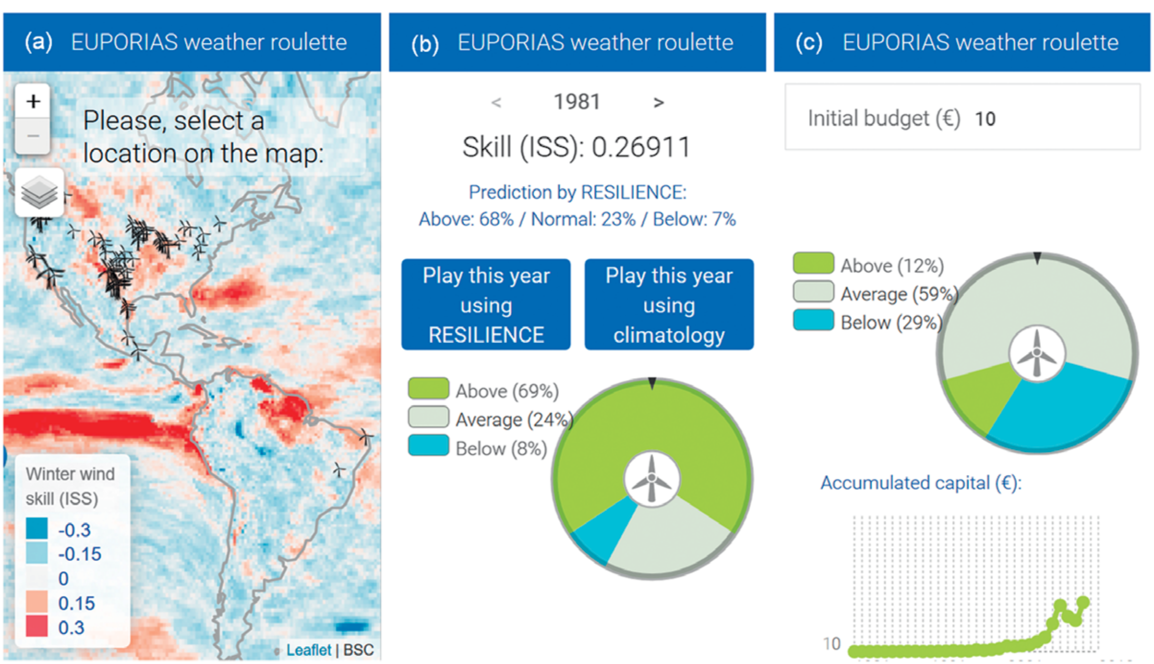

Fig. 1. The WR app: (a) global map of skill (ISS) with the option to select a particular location, (b) option to play for a single year choosing either the seasonal climate prediction or the climatology, and (c) result of the game after playing all years with the seasonal climate prediction. After playing the roulette for a selected year (b) and for all years (c), the screen displays a message informing the player of the winning option, and the return ratio or the return on investment. The black triangle in the roulette shows the tercile where the observation falls.

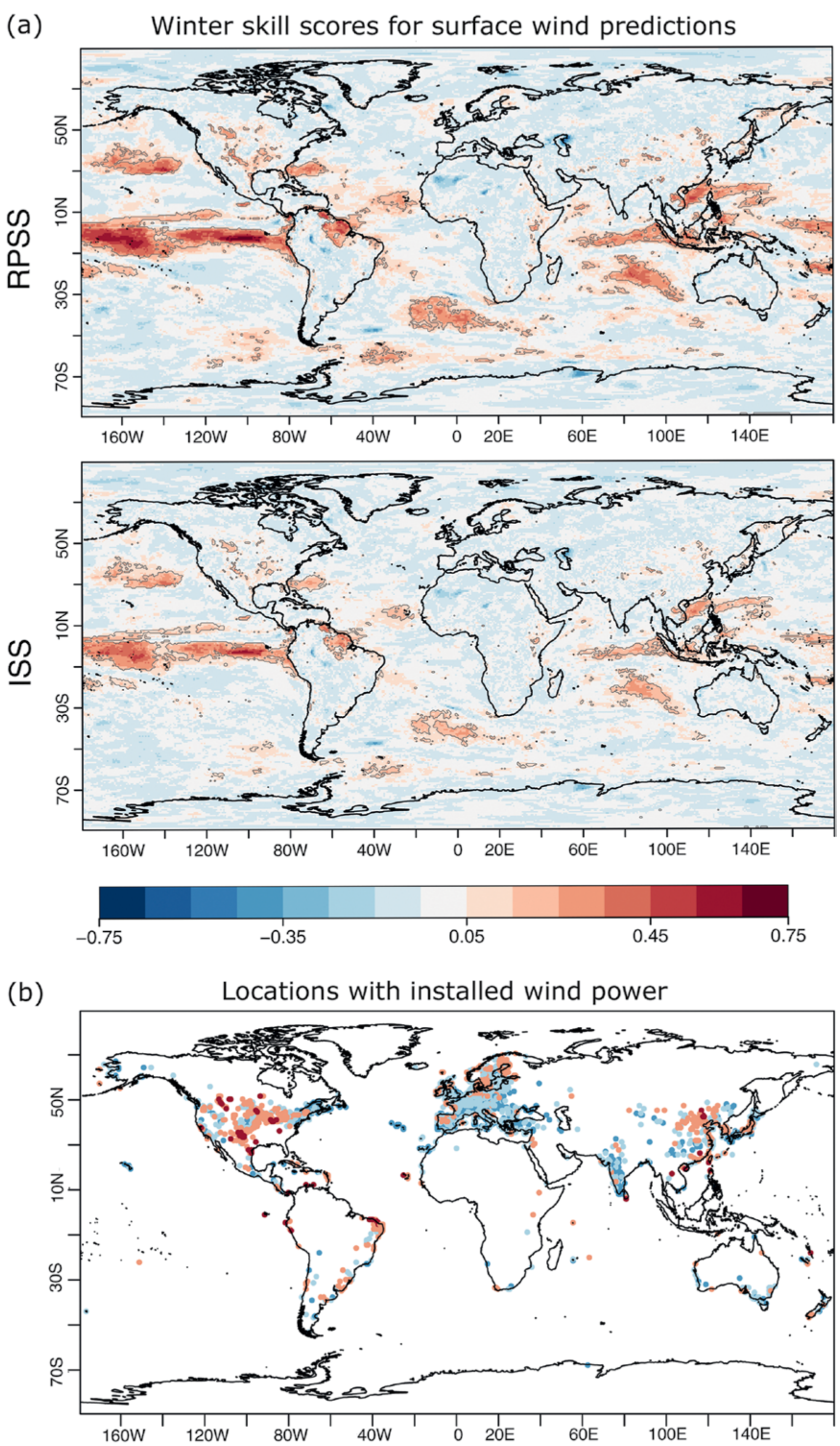

Fig. 2. (a) Skill scores (RPSS and ISS) for tercile categories of winter (DJF) surface (10 m) wind speed, as given by the calibrated seasonal forecasts from ECMWF S4 for the period 1981–2013 (ERA-Interim has been considered as reference). Red (blue) areas show higher (lower) performance than a climatological prediction. Gray contours enclose statistically significant values with a 95% of confidence level. (b) Considered locations with installed wind power divided in locations with negative RPSS (blue) and locations with positive RPSS (red). Light colors indicate nonsignificant RPSS values.

of Europe such as the North Sea or Scandinavia. However, there is a significant number of wind farms in Europe (Fig. 2b), with a nonnegligible amount of installed wind power in regions with low or negative skill such as in southern Europe (Torralba et al. 2017). Wind speed has positive skill in some North American regions. From the 2,023 locations with installed wind power, 473 were located in skillful areas (RPSS > 0). RPSS values tend to be higher than ISS values in most locations. This is due to the nature of the metrics themselves (see Fig. 3 for more detail). Statistical significance of skill score values has been assessed according to Bradley et al. (2008), employing a confidence level of 95%. Skill scores lower than 0.15 are nonsignificant at this confidence level.

Relationship between skill scores and economic indices. Hagedorn and Smith (2009) showed that the average return ratio (R) is a mathematical transformation of the ignorance score (IS) used to calculate the ISS. For the particular case presented here, $R = 3 \times 2^{(-\mathrm{IS})}$. Wind farms with $R > 1$, which is equivalent to the condition IS < 1.58, will produce a return superior to the climatology, meaning that a player choosing climate predictions will win in the WR game.

Although RPSS and ISS do not measure exactly the same thing, they are highly correlated in the case of the wind farms selected in this work (R^2 = 0.840). Thus, in skillful areas (with RPSS > 0 and ISS > 0), higher RPSS and ISS values lead to higher gains in

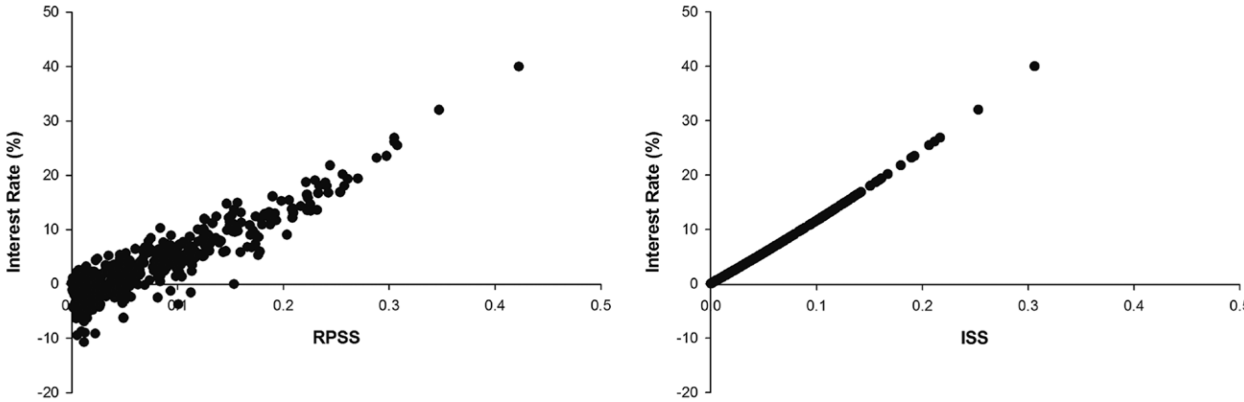

Fig. 3. Interest rate after 33 years at locations in skillful areas: (left) RPSS > 0 and (right) ISS > 0.

the WR (Fig. 3). Unlike ISS, for which positive values are always associated to long-term benefits, for some locations with RPSS values close to zero (0–0.15), either gains or losses can be experienced in terms of IR and ROI. This means that in this RPSS range some climate predictions are not better than climatology despite the positive RPSS value.

Translating skill scores into economic value. The application of the WR to each of the selected locations allows calculating the added value of using seasonal climate predictions compared to using the climatology. Figure 4 shows both the return ratios for each year (r_t, represented by black dots) as well as the average return ratio after 33 years (R, represented by a solid line) at nine different locations. The dashed line in Fig. 4 plots corresponds to $R = 1$. Above this line, predictions perform better than climatology and below they perform worse. A sample of locations from areas with different levels of skill is shown in Fig. 4: locations where RPSS and ISS are both negative (upper row), locations where RPSS is nonsignificant (i.e., ranging from 0 to 0.15) but ISS can have either positive or negative values (middle row), and locations where RPSS and ISS are both positive, with RPSS values above 0.15 (lower row).

In all cases, r_t values can be found indistinctly above or below the dashed line, indicating that a better performance of either climate predictions or climatology depends on the particular year of interest (Fig. 4). However, for those locations with negative values for both RPSS and ISS, the solid line is majorly found below the dashed line. The R values for the three selected locations with negative skill are below 1 (0.80–0.83), and both the IR and ROI report economic losses, meaning that at these locations using climate predictions does not provide any added value over using climatology (Fig. 4, upper row). For locations with RPSS between 0 and 0.15, the solid line can indistinctly appear above or below the dashed line, with a trend to approach the dashed line at the end of the 33-yr period. The R values for these selected locations are around 1 (0.99–1.05) and both economic losses and gains are reported depending on the sign of the ISS value (Fig. 4, middle row). Note that negative IR and ROI values are obtained when ISS < 0. Finally, for locations with positive ISS values and RPSS above 0.15, the solid line is majorly found above the dashed line. These locations have R values above 1 (1.25–1.40) and report economic gains, shown by the positive IR and ROI values (Fig. 4, lower row).

Results in Fig. 4 show how the ROI at the location in Greece (X37407; upper row), which has no skill, is −0.99. This means that the initial bet decreases by almost 100% after 33 years of playing the WR. The case of the locations in Denmark and eastern United States (X24592 and X36231; middle row) illustrates situations where, although the skill is nonsignificant (RPSS < 0.15), using climate predictions is still better than using climatology. However, in other locations with similar skill, such as the one in Canada (X28027; middle row), the WR reports losses at the end of the 33-yr period, meaning that in this case it would have been better to use the climatology. The location in southern United States (X39788; lower row), which exhibits a good skill, has an ROI of 66,049. This means that the initial bet increases by 6,604,900% after 33 years. Figure 5 shows the ROI at skillful locations (where RPSS > 0). Whereas this figure shows some locations with losses (corresponding to 0 < RPSS < 0.15), benefits are obtained in many locations, the highest being in North America and around the tropics.

DISCUSSION. Potential users of climate predictions are far from being a homogeneous group: they belong to different socioeconomic sectors and have

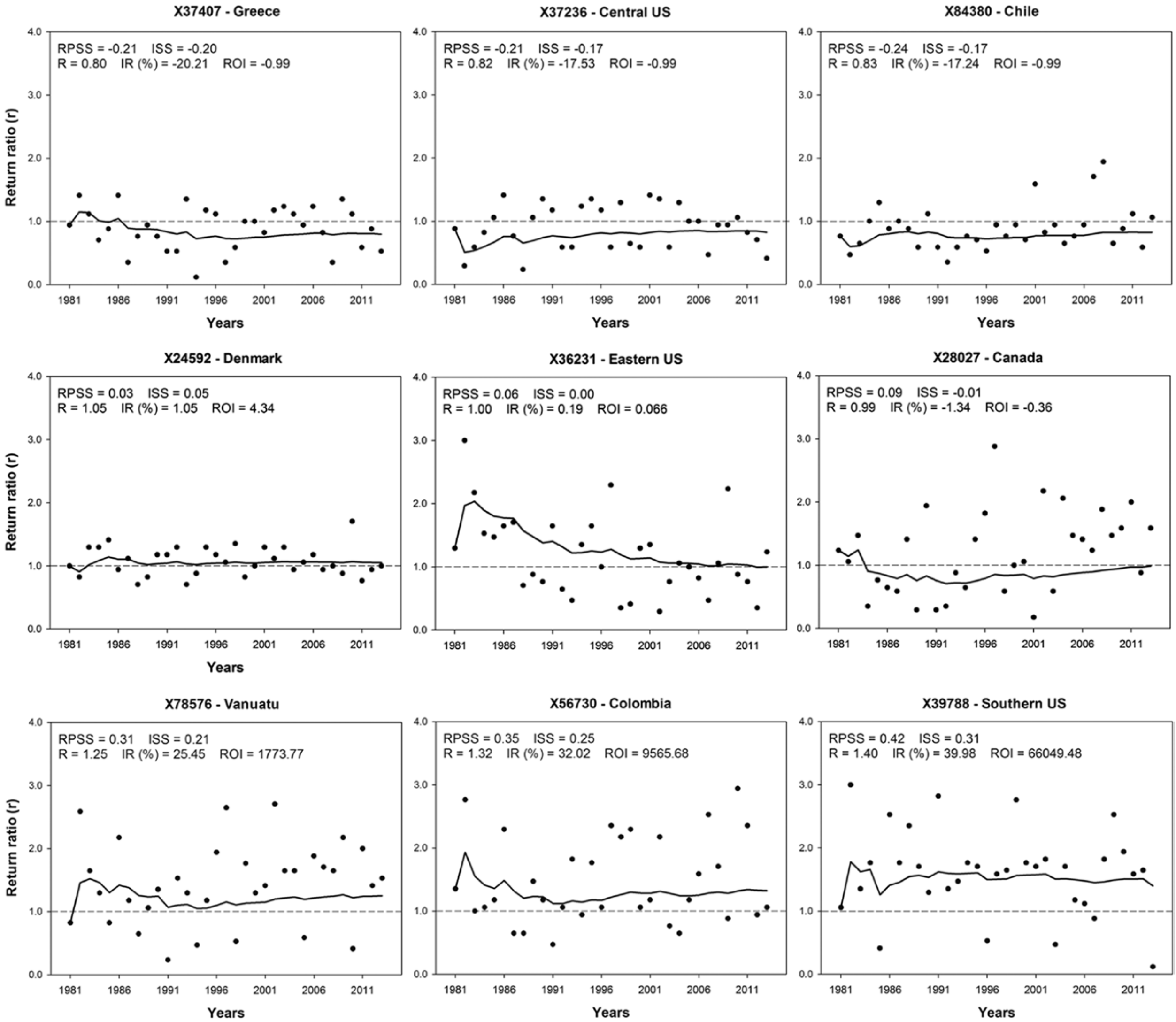

Fig. 4. Example of results from the WR at locations with different levels of skill: (top) RPSS < 0 (Greece–central United States–Chile), (middle) 0 < RPSS < 0.15 (Denmark–eastern United States–Canada), and (bottom) RPSS > 0.15 (Vanuatu–Colombia–southern United States). Black dots show the return ratio (r_t) for each of the 33 years (1981–2013). The solid line is the evolution of the average return ratio (*R*) [i.e., the geometric average of all the previous individual return ratios (r_t)]. The final *R* value is used to calculate the effective interest rate (IR) and the return on investment (ROI) with an initial investment of EUR 10. Over the dashed line (*R* > 1), climate predictions outperform climatology, whereas climatology performs better below the dashed line.

different backgrounds ranging from highly technical users to those with a business background. Therefore, the communication of climate predictions and their associated uncertainties to different audiences requires a transdisciplinary approach able to illustrate the benefits of using climate predictions through alternative approaches such as games. This communication task is normally undertaken by climate knowledge brokers and science communicators working at the interface between the science and user communities. They work to improve coherence and smooth the collaboration between providers and users of climate predictions, which is essential to build trust in such predictions (Bruno-Soares and Dessai 2016). By using the WR approach, we address some of the barriers that have been identified to the uptake of climate predictions.

One of the barriers is related to the uncertainty of an event happening according to a particular forecast, also known as first-order uncertainty (Taylor et al. 2015). There is a mismatch between model outcomes (probabilistic) and users' decision-making approaches (deterministic). From our experience in user engagement, the predicted probability of the most likely category is highly relevant to many users, who often associate higher predicted probabilities to more

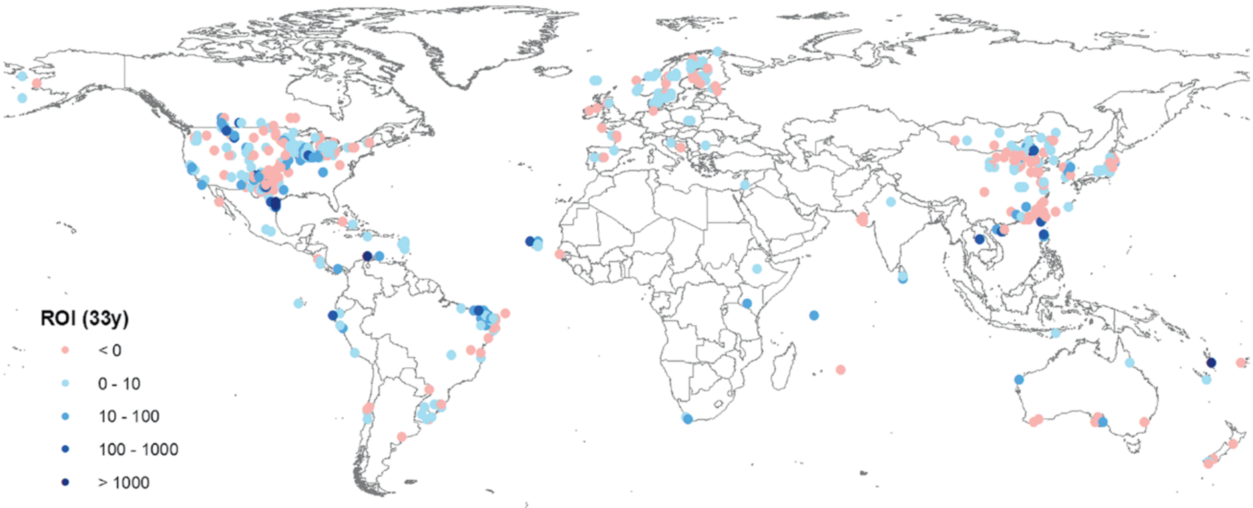

Fig. 5. Return on investment (ROI) at skillful locations (RPSS > 0) for DJF after 33 years.

trustworthy predictions or willingness to take action. Therefore, requests for high predicted probabilities as a method to reduce uncertainty are often found in the descriptions of user needs. However, establishing an appropriate threshold for those probabilities is not straightforward. This requires that users know which are the probability thresholds that maximize their benefits for each relationship between the costs of implementing an action and the losses that users would have incurred if no action had been taken (MacLeod et al. 2015). In addition, although some users may feel comfortable with a predicted probability for the most likely category above 50%, this only occurs occasionally. Indeed, for the locations selected for this study, the predicted most likely tercile probability was above 50% only the 23% of the times (Fig. 6).

Besides the first-order uncertainty, there is a second-order uncertainty related to the quality of the forecast that is more complex to convey to users. The scientific community deals with this uncertainty through the calculation of various metrics, such as skill scores (Taylor et al. 2015). By translating the skill of climate predictions into economic value, we illustrate how the application of climate predictions in areas with skill brings accumulated benefits in the long term. However, for particular years, predictions based on climatology can perform better than climate predictions, even if they are for an area with skill (see Fig. 4). It is important that potential users are aware of this and understand that one single prediction is not representative of the general performance of climate predictions. Early adopters of climate predictions have to accept that they might need to wait a few years to see the benefits of adopting such predictions for decision-making. This is an important point, especially given the short-term thinking of many companies as well

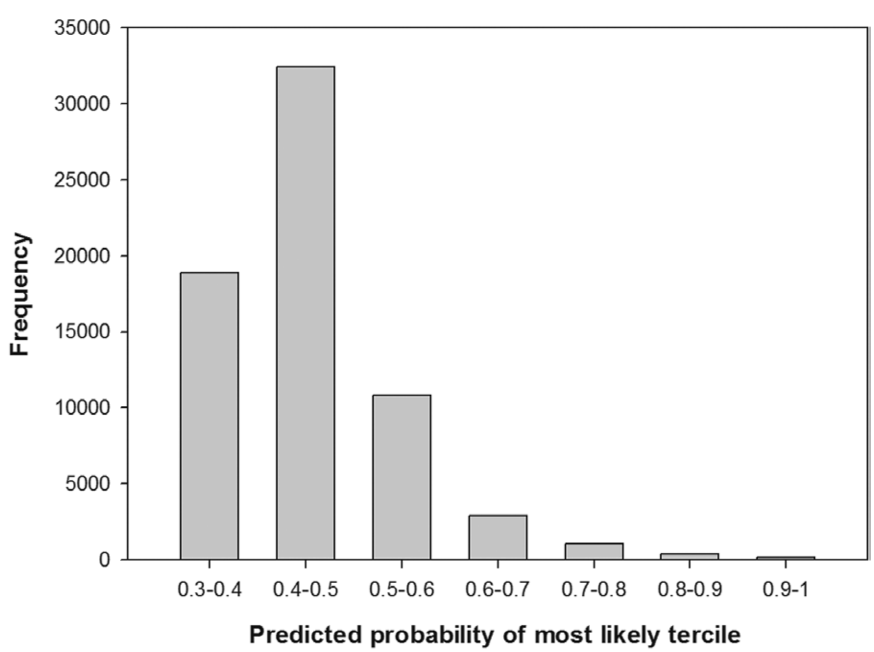

Fig. 6. Frequency of the predicted probability of the most likely category of wind speed for the locations with installed wind power.

as their incentives to avoid risks. In practice, it is unlikely that a user continues using probabilistic forecasts after two years in a row of losses, unless the user understands and has confidence in obtaining benefits in the long term. In this sense, the fact that the WR presents results for several subsequent years helps users to switch from the traditional gain and fail perspective for particular events to a long-term strategy where adjusted probabilities are included in the risk assessments, as a source of information for decision-making.

The evaluation of forecast performance plays a central role in the interpretation and use of forecast systems. Thus, an appropriate communication is needed to make the user aware of the spatiotemporal dependence of skill, and also of its dependence on the type of variable considered (wind speed, temperature, precipitation, etc.). Creating an effective communication strategy requires handling user expectations and looking for windows of opportunity to the application of climate predictions. In this regard, potential users playing the WR app in a location in the south of Europe would easily get the impression that winter wind predictions do not provide any added value over climatology, unless they are aware this is an area of limited skill for this variable. Conversely, the high positive skill in some North American regions can have important implications for the uptake of seasonal climate predictions by the wind energy sector, since the region is characterized by a high installed wind power capacity.

In this study, we show how ISS and RPSS skill scores can be easily explained through the use of the economic metrics such as interest rate and return on investment. This translation into economic terms addresses the terminology gap between climate scientists and users regarding second-order uncertainty. Results of the WR game show that positive ISS and RPSS values are generally associated to obtaining economic benefits in the long term (Figs. 3 and 4). However, for RPSS, there is a range between 0 and 0.15 where either gains or losses are possible (as shown in Fig. 4, middle row). Despite the uncertainty of obtaining long-term economic benefits with low RPSS values, this score deserves special attention, since it is widely used among the climate community (Torralba et al. 2017; Lledó et al. 2018; Manzanas et al. 2019).

An advantage of using RPSS is that the score does not only take into consideration whether or not the prediction system is able to predict a higher probability for the winning (observed) category, but also how big the probabilities are for the nonobserved categories. This is important for real-world applications, when losses and costs of any response action depend on the observed category, and highlights the importance of selecting a verification metric that is relevant to the user's gains and losses. For instance, economic implications will be different if a high probability for above-normal winds is forecasted and normal winds are observed than when below-normal winds are observed. The reason is that the protecting actions that the user takes might still work with normal winds but might be damaging when below-normal winds occur. Moreover, failing to predict the observed category in the wind energy sector would usually cause higher damages than the benefits of succeeding to predict it (Vigo et al. 2018).

We illustrate that the RPSS standard forecast quality measure has a slightly different relation to long-term user benefit than ISS. In all cases, the results highlight that statistically significant skill is not absolutely necessary for a user to obtain a long-term gain. This agrees with the broader discussion that reliance on thresholds of statistical significance can be misleading (Amrhein et al. 2019). Actually, it should be communicated to users that statistical significance, while hugely valuable in a scientific context, tries to respond to questions on specific aspects of the forecast that are not directly linked to the user benefit (Mason 2008; Amrhein et al. 2019). As a result, users should not base their decisions exclusively on the basis of the statistical significance of the results.

CONCLUSIONS. The WR mobile app conveys with an interactive game the different aspects presented in this paper as barriers to the adoption of climate predictions. The terminology gap is overcome through the translation of technical concepts into economic concepts that users are more familiar with. The difficulty to understand the uncertainty of probabilistic outcomes is dealt by allowing players to choose between the climate prediction or climatology for single years after showing them the predicted probabilities. This helps players understand that it is not only the predicted probabilities that matter, but also other factors related to the quality of the climate prediction. Finally, limitations to the understanding of the concept of skill, which needs a long-term perspective, are overcome by informing players on the skill of climate predictions at the selected location and allowing them to play the WR for the entire period. This enables players to see the long-term benefits of integrating climate predictions in their decision-making in skillful areas.

The app has been designed as a simple interface with a limited number of decisions left for the player (selection of a geographical location, selection of the preferred forecast option and definition of the initial bet). More complexity could be added to make the game more interactive (e.g., add data for other seasons, allow the possibility to reinvest only a part of the bet), but it would make user interaction also more complex, especially for those unfamiliar with the type of concepts that are communicated. Future efforts should include some experimental designs to assess users' understanding of the concepts conveyed by the game before and after playing it. This would allow us to quantify the users' learning curve.

The way the WR approach has been applied in this study (setting a random initial capital that is fully reinvested, or considering three tercile categories) is a simplification, and the calculated ROIs cannot be directly translated into real ROIs for a particular company, unless the company agrees in taking the challenge of carrying out a real exercise of including climate predictions in their regular decision-making. This is unlikely due to the high sensitivity of real data on gains and losses. However, even if nonreal economic values are used, it still provides a more intuitive translation of climate-based skill scores into potential economic benefits. We expect that this game could encourage energy users to adopt climate predictions in skillful areas, since revenues will be higher than using the climatology. These predictions, after being tailored to specific decision-making contexts, can be integrated in many decision-making processes, including operations and management strategies, resource allocation for optimum task scheduling, or grid management taking into account renewable energy supply and demand.

Although the WR app is primarily directed to improve the communication of climate services based on seasonal predictions for the wind energy sector, it is a tool that can be useful to illustrate the potential value of using climate predictions in other socioeconomic sectors. A transdisciplinary approach, which implies transcending the disciplinary boundaries and involving actors from outside academia, requires the use of a common language between the climate science and user communities that is necessary to achieve a real coproduction of climate services. In this sense, the WR constitutes a transdisciplinary effort to communicate the usefulness and value of climate predictions in economic terms to different types of users.

The outcomes of this study can be interesting not only in the context of the many projects and initiatives working in the field of climate services and the interface between science and applications, but also for climate scientists that aim to transfer the knowledge arising from their research to potential communities of users.

ACKNOWLEDGMENTS. The research leading to these results has received funding from the EU FP7 Programme under Grant Agreement 308291 (EUPORIAS), the EU H2020 Programme under Grant Agreements 776787 (S2S4E) and 776613 (EUCP), and the Ministerio de Economía y Competitividad (MINECO) as part of the CLINSA Project CGL2017-85791-R. It is also part of the Copernicus Climate Change Service (C3S) (Framework Agreement C3S_441_Lot2_CEA), a program being implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF) on behalf of the European Commission. The authors want to thank Amaia Aizpurua for her valuable technical help.

REFERENCES

Amrhein, V., S. Greenland, and B. McShane, 2019: Scientists rise up against statistical significance. *Nature*, **567**, 305–307, https://doi.org/10.1038/d41586-019-00857-9.

Arnal, L., M.-H. Ramos, E. Coughlan, H. L. Cloke, E. Stephens, F. Wetterhall, S.-J. van Andel, and F. Pappenberger, 2016: Willingness-to-pay for a probabilistic flood forecast: A risk-based decision-making game. *Hydrol. Earth Syst. Sci.*, **20**, 3109–3128, https://doi.org/10.5194/hess-20-3109-2016.

Bradley, A. A., S. S. Schwartz, and T. Hashino, 2008: Sampling uncertainty and confidence intervals for the Brier score and Brier skill score. *Wea. Forecasting*, **23**, 992–1006, https://doi.org/10.1175/2007WAF2007049.1.

Bruno-Soares, M., and S. Dessai, 2016: Barriers and enablers to the use of seasonal climate forecasts amongst organisations in Europe. *Climatic Change*, **137**, 89–103, https://doi.org/10.1007/s10584-016-1671-8.

——, M. Daly, and S. Dessai, 2018: Assessing the value of seasonal climate forecasts for decision-making. *Wiley Interdiscip. Rev.: Climate Change*, **9**, e523, https://doi.org/10.1002/wcc.523.

Caron, L. P., L. Hermanson, A. Dobbin, J. Imbers, L. Lledó, and G. A. Vecchi, 2018: How skillful are the multiannual forecasts of Atlantic hurricane activity? *Bull. Amer. Meteor. Soc.*, **99**, 403–413, https://doi.org/10.1175/BAMS-D-17-0025.1.

Carta, J. A., S. Velázquez, and P. Cabrera, 2013: A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site. *Renewable Sustainable Energy Rev.*, **27**, 362–400, https://doi.org/10.1016/j.rser.2013.07.004.

Coelho, C. A. S., and S. M. S. Costa, 2010: Challenges for integrating seasonal climate forecasts in user applications. *Curr. Opin. Environ. Sustainability*, **2**, 317–325, https://doi.org/10.1016/j.cosust.2010.09.002.

Crochemore, L., M.-H. Ramos, F. Pappenberger, S.-J. van Andel, and A. W. Wood, 2016: An experiment on risk-based decision-making in water management using monthly probabilistic forecasts. *Bull. Amer. Meteor. Soc.*, **97**, 541–551, https://doi.org/10.1175/BAMS-D-14-00270.1.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Doblas-Reyes, F. J., R. Hagedorn, and R. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting—II. Calibration and combination. *Tellus*, **57A**, 234–252, https://doi.org/10.3402/tellusa.v57i3.14658.

——, J. García-Serrano, F. Lienert, A. P. Biescas, and L. R. Rodrigues, 2013: Seasonal climate predictability and forecasting: Status and prospects. *Wiley Interdiscip. Rev.: Climate Change*, **4**, 245–268, https://doi.org/10.1002/wcc.217.

European Commission, 2015: A European research and innovation roadmap for climate services. European Commission Directorate-General for Research and Innovation Rep., 56 pp., https://doi.org/10.2777/702151.

Feldman, D. L., and H. M. Ingram, 2009: Making science useful to decision makers: Climate forecasts, water management, and knowledge networks. *Wea. Climate Soc.*, **1**, 9–21, https://doi.org/10.1175/2009WCAS1007.1.

Good, I. J., 1952: Rational decisions. *J. Roy. Stat. Soc.*, **14A**, 107–114.

Hagedorn, R., and L. A. Smith, 2009: Communicating the value of probabilistic forecasts with Weather Roulette. *Meteor. Appl.*, **16**, 143–155, https://doi.org/10.1002/met.92.

Jewson, S., E. Bellone, S. Khare, T. Lepple, M. Lonfat, A. O. Shay, J. Penzer, and K. Coughlin, 2009: Five year prediction of the number of hurricanes that make United States landfall. *Hurricanes and Climate Change*, J. B. Elsner and T. H. Jagger, Eds., Springer, 73–99, https://doi.org/10.1007/978-0-387-09410-6_5.

Jolliffe, I. T., and D. B. Stephenson, 2012: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons, 288 pp.

Jupp, T. E., R. Lowe, C. Coelho, and D. B. Stephenson, 2012: On the visualization, verification and recalibration of ternary probabilistic forecasts. *Philos. Trans. Roy. Soc.*, **370A**, 20110350, https://doi.org/10.1098/rsta.2011.0350.

Lazenby, M. J., W. A. Landman, R. M. Garland, and D. G. DeWitt, 2014: Seasonal temperature prediction skill over southern Africa and human health. *Meteor. Appl.*, **21**, 963–974, https://doi.org/10.1002/met.1449.

Lledó, L., 2017: Seasonal predictions for wind energy: Anticipating quarterly winds and revenues one month ahead. WindTech International, www.windtech-international.com/editorial-features/seasonal-predictions-for-wind-energy.

——, O. Bellprat, F. J. Doblas-Reyes, and A. Soret, 2018: Investigating the effects of Pacific sea surface temperatures on the wind drought of 2015 over the United States. *J. Geophys. Res. Atmos.*, **123**, 4837–4849, https://doi.org/10.1029/2017JD028019.

MacLeod, D. A., A. Jones, F. Di Giuseppe, C. Caminade, and A. P. Morse, 2015: Demonstration of successful malaria forecasts for Botswana using an operational seasonal climate model. *Environ. Res. Lett.*, **10**, 044005, https://doi.org/10.1088/1748-9326/10/4/044005.

Manzanas, R., M. D. Frías, A. S. Cofiño, and J. M. Gutiérrez, 2014: Validation of 40 year multimodel seasonal precipitation forecasts: The role of ENSO on the global skill. *J. Geophys. Res. Atmos.*, **119**, 1708–1719, https://doi.org/10.1002/2013JD020680.

——, J. M. Gutiérrez, J. Bhend, S. Hemri, F. J. Doblas-Reyes, V. Torralba, E. Penabad, and A. Brookshaw, 2019: Bias adjustment and ensemble recalibration methods for seasonal forecasting: A comprehensive intercomparison using the C3S dataset. *Climate Dyn.*, **53**, 1287–1305, https://doi.org/10.1007/s00382-019-04640-4.

Mason, S. J., 2008: Understanding forecast verification statistics. *Meteor. Appl.*, **15**, 31–40, https://doi.org/10.1002/met.51.

McNie, E. C., 2007: Reconciling the supply of scientific information with user demands: An analysis of the problem and review of the literature. *Environ. Sci. Policy*, **10**, 17–38, https://doi.org/10.1016/j.envsci.2006.10.004.

Meehl, G. A., and Coauthors, 2014: Decadal climate prediction: An update from the trenches. *Bull. Amer. Meteor. Soc.*, **95**, 243–267, https://doi.org/10.1175/BAMS-D-12-00241.1.

Molteni, F., and Coauthors, 2011: The new ECMWF seasonal forecast system (system 4). ECMWF Tech. Memo. 656, 51 pp.

Predictia, 2019: Weather Roulette. Predictia, https://play.google.com/store/apps/details?id=es.predictia.weatherroulette.

Palmer, T. N., G. J. Shutts, R. Hagedorn, F. J. Doblas-Reyes, T. Jung, and M. Leutbecher, 2005: Representing model uncertainty in weather and climate prediction.

Annu. Rev. Earth Planet. Sci., **33**, 163–193, https://doi.org/10.1146/annurev.earth.33.092203.122552.

Ramos, M. H., S. J. van Andel, and F. Pappenberger, 2013: Do probabilistic forecasts lead to better decisions? *Hydrol. Earth Syst. Sci.*, **17**, 2219–2232, https://doi.org/10.5194/hess-17-2219-2013.

R Core Team, 2015: R: A language and environment for statistical computing. R Foundation, www.r-project.org/.

Taylor, A. L., S. Dessai, and W. Bruine de Bruin, 2015: Communicating uncertainty in seasonal and interannual climate forecasts in Europe. *Philos. Trans. Roy. Soc.*, **373A**, 20140454, https://doi.org/10.1098/rsta.2014.0454.

Terrado, M., N. González-Reviriego, L. Lledó, V. Torralba, A. Soret, and F. J. Doblas-Reyes, 2017: Climate services for affordable wind energy. *WMO Bull.*, **66**, 48–53.

Thomson, M. C., F. J. Doblas-Reyes, S. J. Mason, R. Hagedorn, S. J. Connor, T. Phindela, and T. N. Palmer, 2006: Malaria early warnings based on seasonal climate forecasts from multi model ensembles. *Nature*, **439**, 576–579, https://doi.org/10.1038/nature04503.

Torralba, V., F. J. Doblas-Reyes, D. MacLeod, I. Christel, and M. Davis, 2017: Seasonal climate prediction: A new source of information for the management of wind energy resources. *J. Appl. Meteor. Climatol.*, **56**, 1231–1247, https://doi.org/10.1175/JAMC-D-16-0204.1.

Turco, M., A. Ceglar, C. Prodhomme, A. Soret, A. Toreti, and F. J. Doblas-Reyes, 2017: Summer drought predictability over Europe: Empirical versus dynamical forecasts. *Environ. Res. Lett.*, **12**, 084006, https://doi.org/10.1088/1748-9326/aa7859.

van Pelt, S. C., M. Haasnoot, B. Arts, F. Ludwig, R. Swart, and R. Biesbroek, 2015: Communicating climate (change) uncertainties: Simulation games as boundary objects. *Environ. Sci. Policy*, **45**, 41–52, https://doi.org/10.1016/j.envsci.2014.09.004.

Vigo, I., and Coauthors, 2018: User needs and decision-making processes that can benefit from S2S forecasts. S2S4E Tech. Rep. D2.2, 92 pp.

Vincent, K., A. J. Dougill, J. L. Dixon, L. C. Stringer, and T. Cull, 2017: Identifying climate services needs for national planning: Insights from Malawi. *Climate Policy*, **17**, 189–202, https://doi.org/10.1080/14693062.2015.1075374.

von Storch, H., and F. W. Zwiers, 2001: *Statistical Analysis in Climate Research*. Cambridge University Press, 496 pp.

Weisheimer, A., and T. N. Palmer, 2014: On the reliability of seasonal climate forecasts. *J. Roy. Soc. Interface*, **11**, 20131162, https://doi.org/10.1098/rsif.2013.1162.

White, C. J., and Coauthors, 2017: Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteor. Appl.*, **24**, 315–325, https://doi.org/10.1002/met.1654.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 676 pp.