Introduction to Probability Forecasts

Summary

World Climate Service seasonal and sub-seasonal weather forecasts emphasize probabilistic information, which is less intuitive but more powerful than the traditional style of short-term weather forecasts. Probability forecasts are able to empower quantitative decision systems, because they allow the user to calculate the financial consequences of action or inaction based on the forecasts. In this way, World Climate Service forecasts enable users to make precisely correct decisions in the face of uncertainty and thereby to take control of weather risk.

1. Introduction

The world is full of uncertainty about future outcomes, ranging from unexpected storms to accidents, elections, or unruly financial markets. Modeling this uncertainty through the use of probabilities and statistical inference provides the key to making better and more useful predictions, leading to more successful decisions. The World Climate Service relies heavily on a probability framework to convey sub-seasonal to seasonal (S2S) forecast information.

A significant challenge associated with probability forecasts is that first-time users often encounter difficulty in interpreting and using the forecast information. The difficulty arises because probability forecasts represent a fundamentally different kind of information from the short-range weather forecasts that are common in modern society. Traditional short-range weather forecasts are generally "deterministic", meaning that the forecast shows a specific outcome (e.g. temperature or wind speed) for each day or hour.[1] However, probability forecasts appear to show a wide range of possibilities and it is often not immediately clear how to use this type of information.

The rationale for presenting S2S guidance as probability forecasts is two-fold:

- Outcomes beyond about seven days in the future are inherently very uncertain, and therefore it does not make sense to show specific outcomes that are almost certain not

[1] An exception to this rule is that short-range precipitation forecasts usually show a "chance" or probability of precipitation, because it is widely understood that forecasters often cannot say with confidence whether or not rain or snow will occur in a specific day or hour.

to occur. In other words, the average error associated with deterministic forecasts is very large at longer lead times.

- More importantly, probability information can be used to calculate expected financial outcomes based on decisions made in response to the forecasts. Probability forecasts therefore translate directly into expected profit and loss, which is a primary concern for many enterprises.

The goal of this paper is to explain the main elements of probability forecasting and to highlight the power of probability forecasts for decision-making. Section 2 describes the conceptual framework for probability forecasts and the forecast performance metric used by the World Climate Service, and Section 3 outlines the mechanism for calculating financial outcomes.

2. Components of Probability Forecasts

a. Tercile Probabilities

The concept of probability applies to binary ("either/or") outcomes, and therefore to make probability statements about continuous variables like temperature, it is necessary to define categories of outcome that either will or will not occur. Once the categories are defined, then the probability of each category can be found. It is traditional in seasonal forecasting to divide the possible outcomes into three categories called "terciles": below-normal, near-normal, and above-normal. All possible outcomes fall into one of these categories, and there is no overlap between them. The terciles represent an equal division of the ranked historical data within a specific historical period such as 1981-2010, and one-third of the historical data points fall in each tercile. If the climate were unchanging, then each tercile would be equally likely in the long run, and if no forecast is made, then the probability of each tercile is 33%. The goal of probability forecasts is to show how the likelihood of each tercile differs from equal-chances; each tercile's probability can range from 0% to 100%, but the sum of the three tercile probabilities is always 100%.

Figure 1 shows an example of a probabilistic forecast that indicates the expected evolution of the El Niño – Southern Oscillation (ENSO) phenomenon in the equatorial Pacific Ocean. The forecast shows extremely high probabilities of El Niño conditions in the first several months, followed by a rising probability of neutral conditions and eventually a near-50% probability of La Niña conditions late in the forecast period.

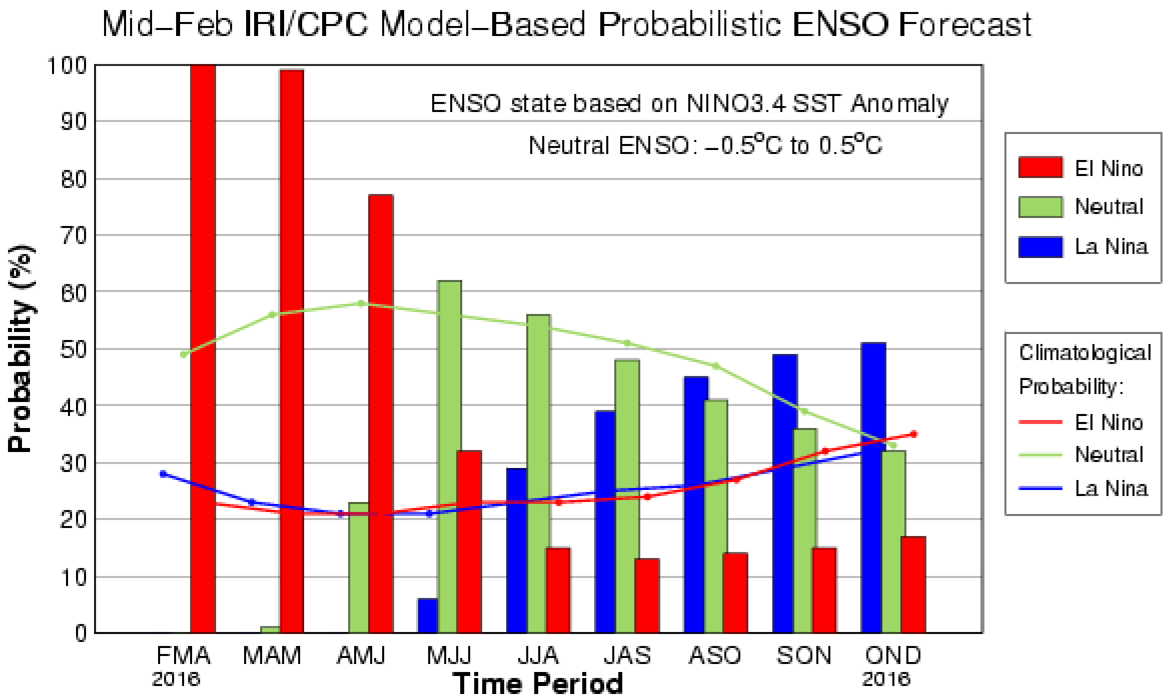

Figure 1. Probabilistic ENSO forecast issued by the International Research Institute for Climate and Society (IRI) and the NOAA Climate Prediction Center (CPC).

World Climate Service probability forecasts are presented in maps that display tercile probabilities with respect to a stated historical normal ("climatology") such as 1981-2010 (e.g. Figure 2). The shading indicates the highest of the three tercile probabilities at each point. For example, if the map shows an above-normal probability of 60% at a point, then the below-normal and near-normal probabilities are not displayed, but the user knows that they must add up to 40% so that the total probability is 100%.

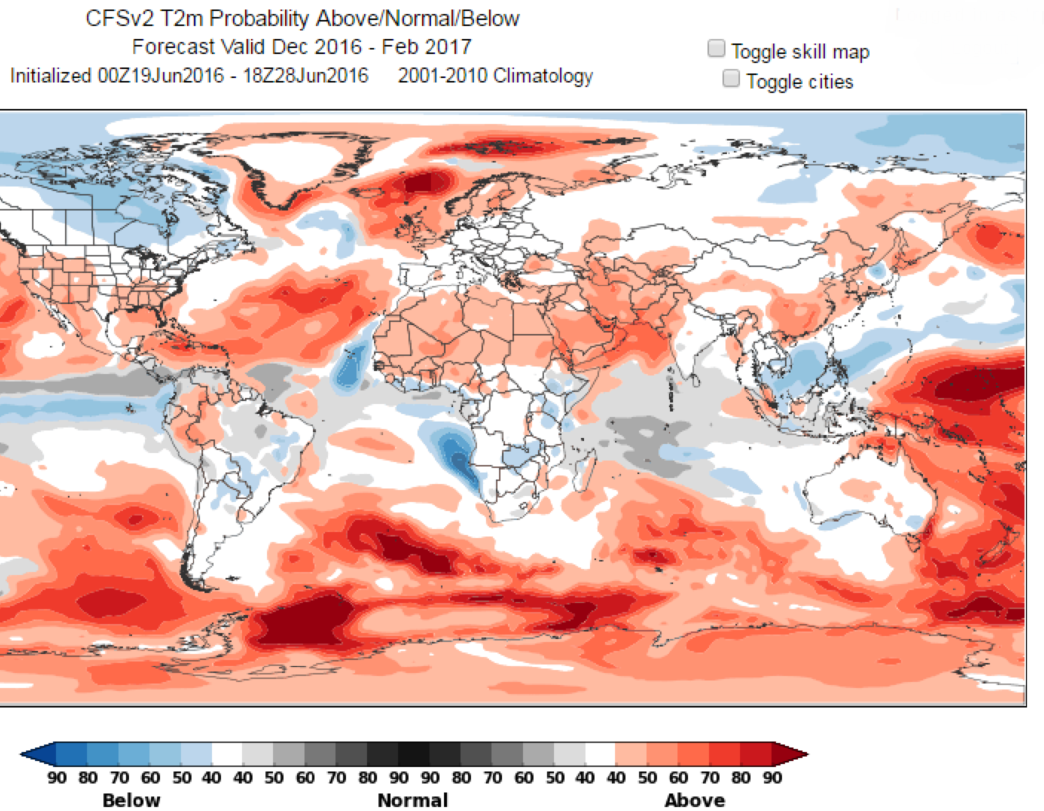

Figure 2. WCS tercile probability forecast of mean 2m temperature in December-February 2016-2017, based on the CFSv2 model ensemble forecast.

In locations where the map shows no shading (i.e. white), the tercile probabilities are nearly equal, as none of them is greater than 40%. This is a forecast for “nearly equal chances” of any outcome and ***is not the same as a forecast for near-normal conditions***; in other words, white shading does not indicate that near-normal conditions are particularly likely. If the forecast were to indicate a high chance of a near-normal outcome, then the map would show gray shading; however, it is uncommon to see a high probability of near-normal conditions, because the near-normal tercile is more difficult to predict than the below-normal and above-normal terciles.

b. Reliability

An important concept for understanding probability forecasts is that of reliability. A reliable probability forecast is one for which the frequency of the outcome matches the predicted probability over the long-term. For example, if a perfectly reliable forecast shows a 70% chance of below-normal conditions on 10 different occasions, then 7 of the 10 occasions will in fact produce below-normal conditions. A reliable forecast is neither over-confident nor under-confident, and this is an essential quality if forecasts are to be used in quantitative decision systems as described in Section 3.

It is the responsibility of a forecast provider such as the WCS to ensure that probability forecasts are reliable, and forecast calibration is necessary to achieve this goal. Forecast calibration is a complex scientific problem, and the WCS has made large investments of time and computational resources to develop robust calibration schemes that rely on long histories of model forecasts. WCS clients benefit from this multi-year research and development work and can have confidence that the WCS probability forecasts are appropriately calibrated.

Forecast reliability can be assessed with the help of a reliability diagram, which plots forecast probability against observed frequency for a large set of forecasts. If the forecasts are perfectly reliable, then the points will fall along the diagonal line with a slope of 1:1. Deviations from perfect reliability are evident in departures from the diagonal. For example, Figure 1 shows a reliability diagram for ***uncalibrated*** CFSv2 forecasts of 2m temperature during winter; these probability forecasts were obtained simply by counting model ensemble members within the model's own historical terciles. It is clear that the uncalibrated forecasts are overconfident, because the slope of the reliability curves is less than one.

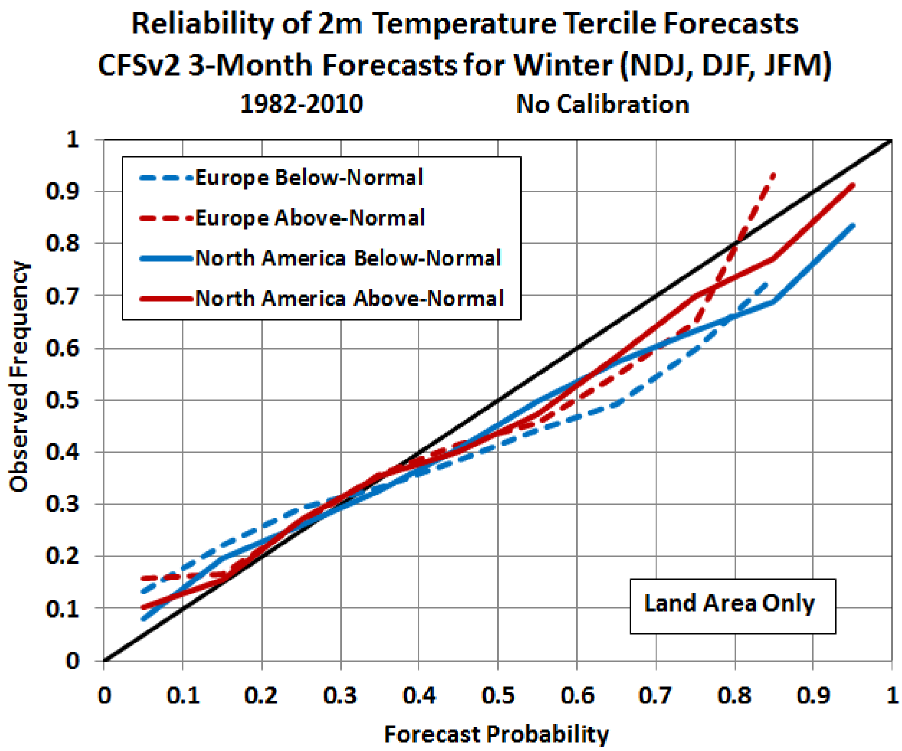

Figure 3. Reliability diagram for uncalibrated CFSv2 seasonal (3-month mean) 2m temperature tercile probability forecasts for winter over Europe and North America.

Figure 4 shows the reliability diagram for the same forecasts after applying the WCS calibration. While there are some departures from perfect calibration at the upper end of the forecast probability range, the sample size is small at these high probability values, and the reliability is very good over most of the probability range.

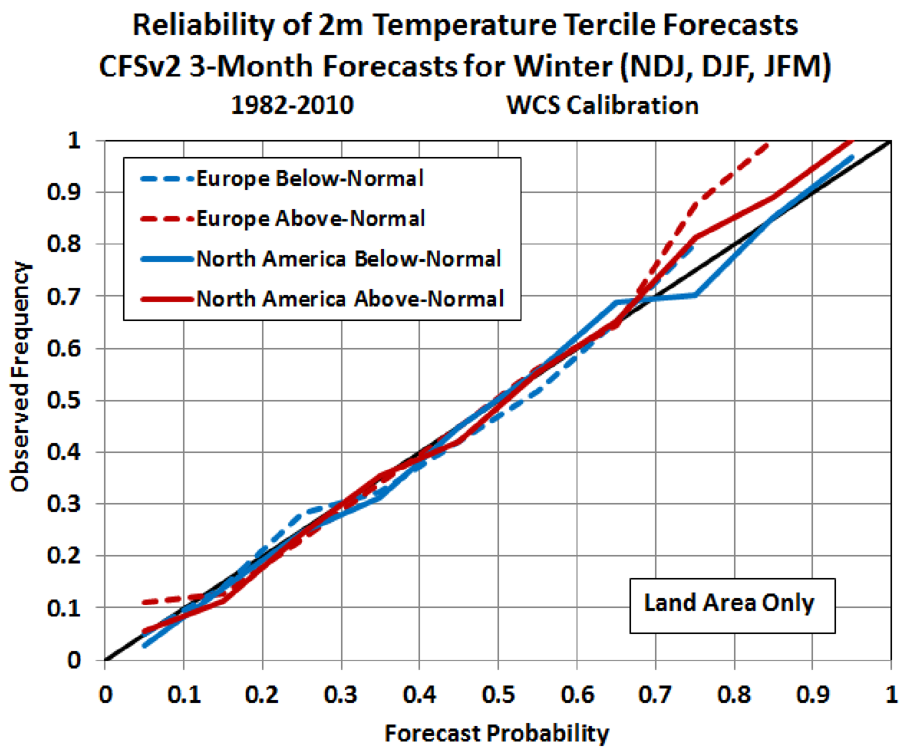

Figure 4. As in Figure 3, but for forecasts calibrated by the WCS.

Figures 5 and 6 show reliability diagrams for winter forecasts of precipitation, which are more seriously overconfident prior to calibration. Note that the reliability curves for Europe do not extend above 75% forecast probability, because there are too few high-probability forecasts over Europe to calculate the reliability at the upper end of the probability range.

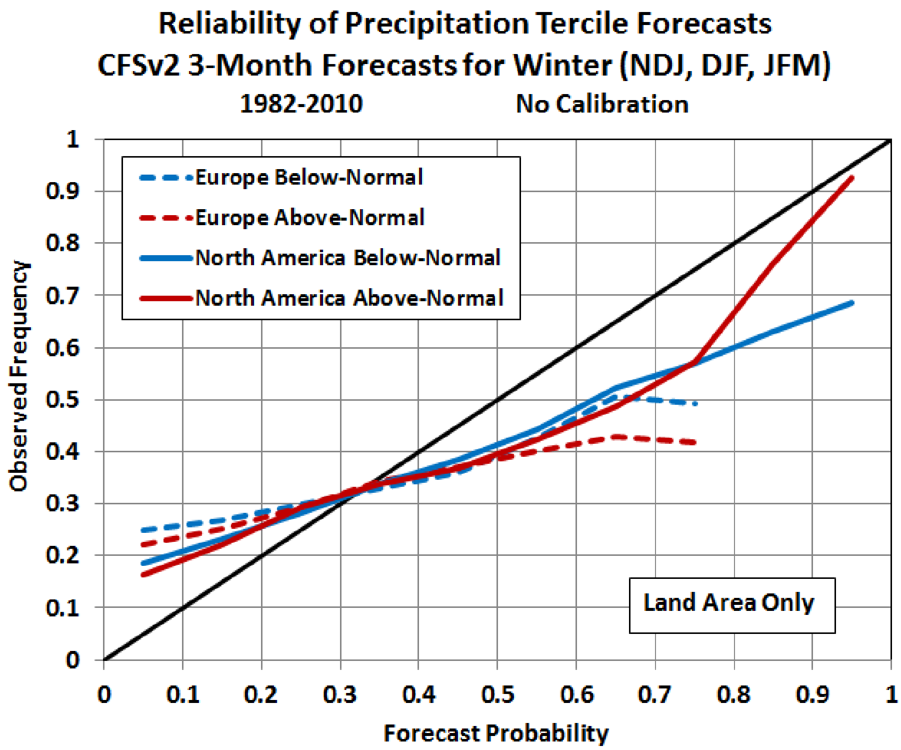

Figure 5. Reliability diagram for uncalibrated CFSv2 seasonal (3-month total) precipitation tercile probability forecasts for winter over Europe and North America.

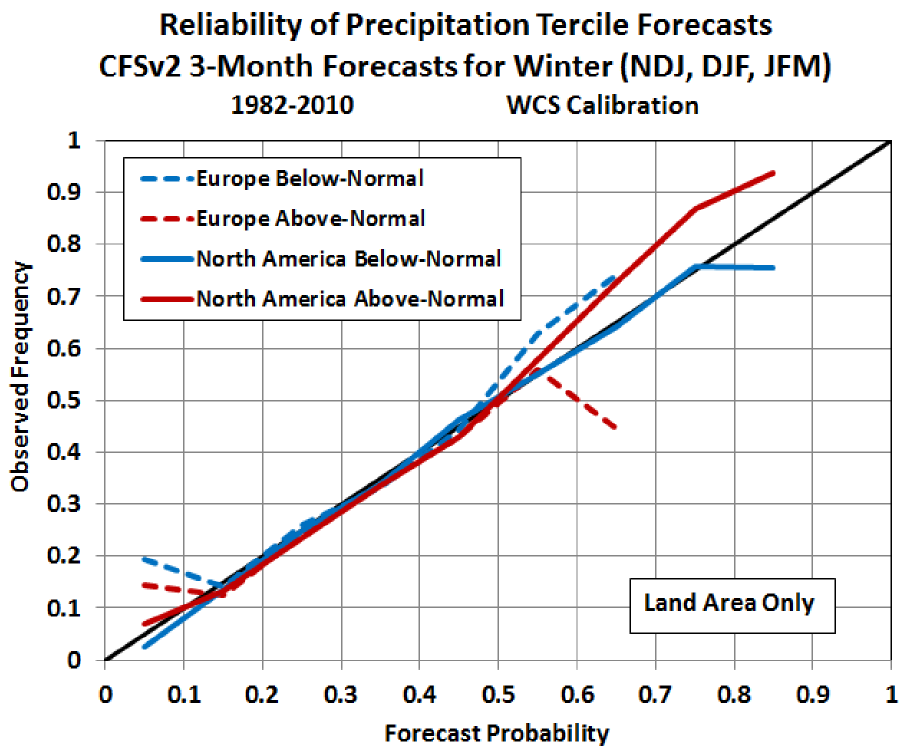

Figure 6. As in Figure 5, but for forecasts calibrated by the WCS.

c. Fraction Correct

Users often ask, "How good is the forecast?" With a deterministic forecast, this question is relatively easy to answer, because simple metrics such as mean absolute error are sufficient to describe the forecast accuracy. However, it is more challenging to describe the accuracy of probabilistic forecasts, because a probability value does not unequivocally point to any specific outcome.

The scientific literature contains many alternative measures of performance for probabilistic forecasts, but the WCS has chosen to use a single metric that is relatively intuitive – the "fraction correct". In the context of the WCS probability forecast maps, **the fraction correct statistic answers the simple question, "How often does the forecast map color shading indicate the correct tercile?"** Recall that the forecast map shading indicates which of the three terciles has the highest probability; therefore we are interested in determining how often the observed outcome is within the highest-probability tercile. It would be possible to extract more information about how the other tercile probabilities perform, but we restrict ourselves to the most likely tercile, which is the one shown on the map.

In practice the fraction correct is calculated by generating a long history of forecasts and comparing them to the observed outcomes. For all forecasts in which the above-normal tercile probability was highest, we count the number of times that above-normal was observed. After doing the same for below-normal and near-normal, the overall fraction correct is obtained from the ratio of the number of correct forecasts to the total number of forecasts. Figure 7 illustrates the fraction correct for CFSv2 seasonal forecasts of 2m temperature over North America and Europe. Note that random skill-less forecasts would have a fraction correct of 0.33 or 33%, so fraction correct values above 0.33 indicate an improvement over random chance.

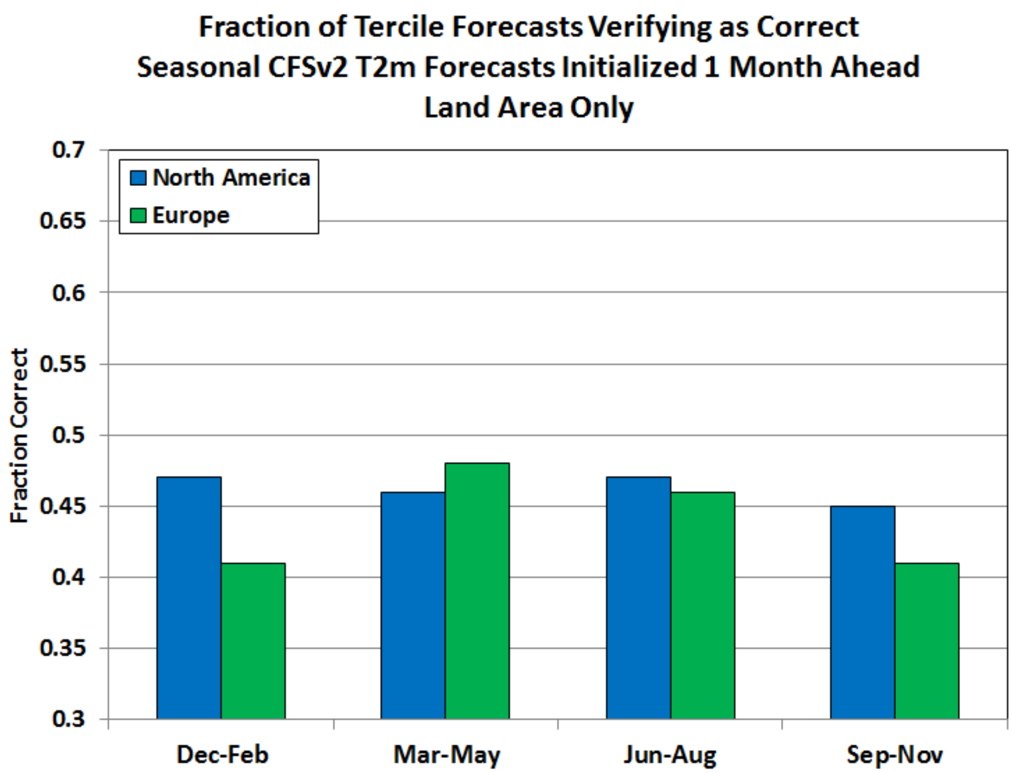

Figure 7. Fraction correct for CFSv2 seasonal forecasts of 2m temperature over North America and Europe (land area only), for a 1-month lead time. Random forecasts would have a fraction correct of 0.33.

Given that WCS forecasts are designed to be reliable, it is clear that the fraction correct will be higher when the forecast probability is higher; in other words, higher confidence translates into a greater likelihood of success. The fraction correct shown above in Figure 7 indicates the forecast performance for all probability levels; for example, this performance is representative of forecasts with tercile probabilities that might be as diverse as (34%, 33%, 33%) or (1%, 8%, 91%). To explore how confidence affects performance, we also sub-divide the historical forecasts by probability level and obtain the fraction correct for different threshold values of the highest tercile probability. Figure 8 shows the fraction correct for all forecasts with a tercile probability of at least 50%; the performance is much better than in Figure 7.

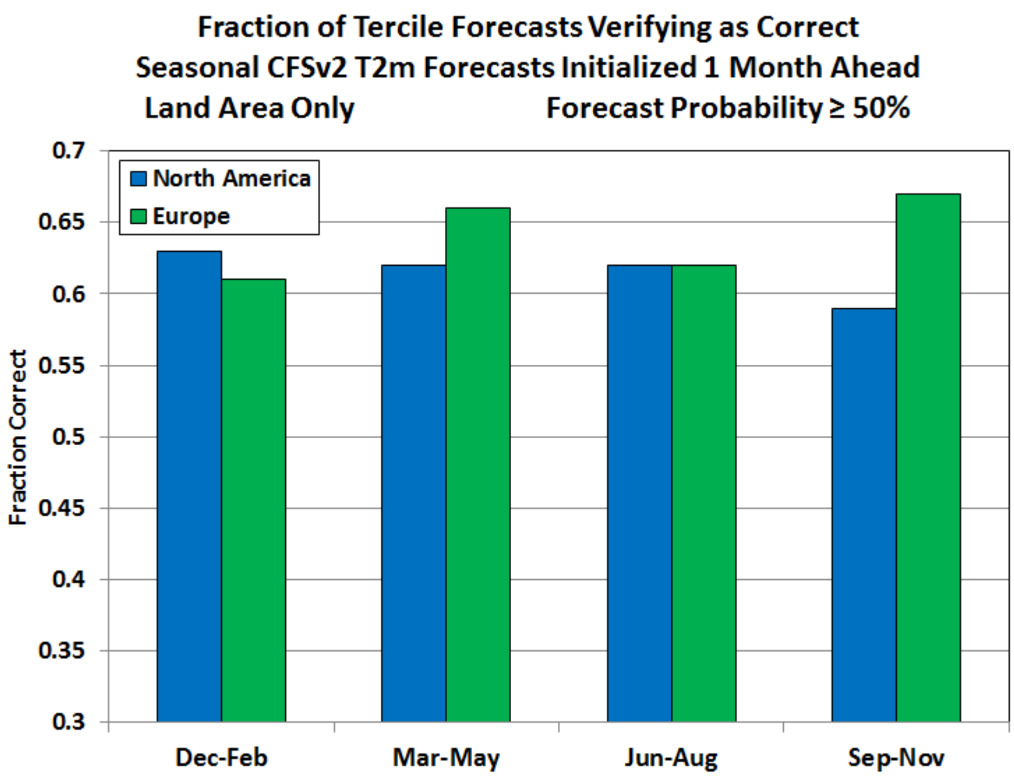

Figure 8. As in Figure 7, but for forecasts with a probability of at least 50%; forecasts with lower confidence are excluded.

Maps of the fraction correct skill metric are available on several of the WCS model forecast pages; a checkbox to the right of the map title allows users to toggle the “skill map”, which shows the fraction correct (e.g. Figure 9). The fraction correct values are determined from the historical forecasts for each location, forecast variable, lead time, and month of the year separately. The calculation is performed using all tercile forecasts regardless of confidence level, so the skill maps reflect the average performance for all forecasts. The fraction correct skill maps provide users with the ability to discern locations or seasons in which the historical forecasts demonstrate particularly good or poor performance. The example shown in Figure 9 indicates that CFSv2 temperature forecasts made in June for the subsequent December through February are moderately skillful in most of the tropics and also show modest skill over the northern North Atlantic Ocean, but skill is marginal over most of North America and minimal in most of Europe.

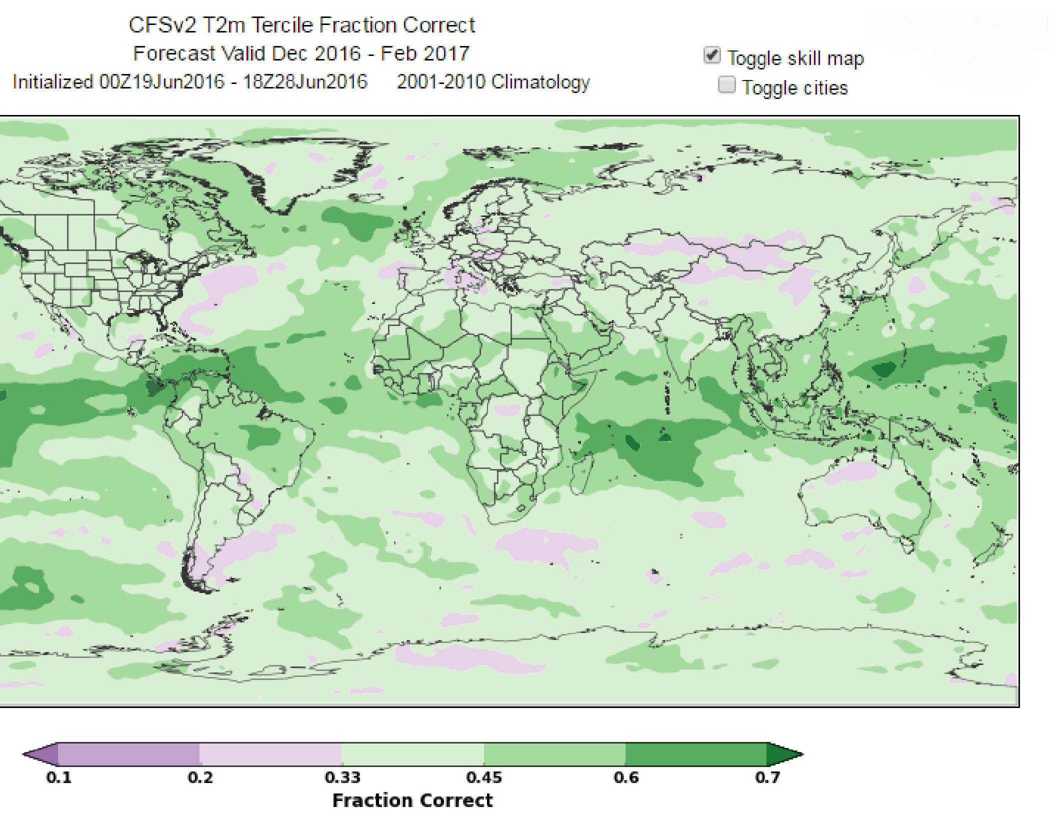

Figure 9. Fraction correct corresponding to the probability forecast shown in Figure 2.

3. Decision Systems

The power of probability forecasts lies in their ability to inform quantitative decision systems that produce known and optimal financial outcomes over the long-term. This is possible because reliable probability forecasts provide a true indication of the likelihood of an outcome, and this can be paired with the known financial implications of action or inaction based on the forecasts. The result is a powerful tool for risk management through informed decision-making. A simple example using short-term probability forecasts will illustrate the process, but the same procedure applies to S2S forecasts for any weather-sensitive activity.

Consider a small business whose primary activity is dependent on having dry weather in order to operate profitably. A company that pours concrete will serve as a good example. Each day, the business owner uses a weather forecast to decide whether or not to proceed with normal operations for the day. If the forecast suggests wet weather, then the owner may decide to postpone concrete-pouring until the next day, but if the forecast suggests fine weather, then normal operations will proceed. Either decision carries some risk, because the weather may turn out differently from the owner's expectation. In this example, a significant financial loss may be incurred if unexpected rain occurs during or after pouring concrete. On the other hand, the decision to sit idle will create an unnecessary loss if the weather turns out to be dry.

How is the business owner to make this decision? In many instances, the business owner will know from experience when it is safe to proceed with a pour, and when to postpone. However, a reliable probabilistic forecast provides the necessary information to make the precisely correct decision. A simple calculation using the profit and loss information from the business, together with the probabilistic forecast, provides the exact answer for an optimal outcome over the long-term.

Three pieces of financial data are needed from the business, as follows:

- L is the loss that occurs when operations proceed but rain occurs
- P is the profit that occurs when operations proceed and rain does not occur
- C is the cost of postponing operations (e.g. overhead expenses, salaries)

A probabilistic weather forecast is also needed:

- R is the forecast probability of rain occurring

The expected profit if the business owner proceeds with operations is:

$$profit_{proceed} = P(1 - R) - LR$$

The expected profit if the business owner postpones operations is

$$profit_{postpone} = -C$$

Therefore the business owner should proceed if, and only if,

$$P(1 - R) - LR > -C$$

$$-R(P + L) > -C - P$$

$$-R > -\frac{C + P}{P + L}$$

$$R < \frac{(P + C)}{(P + L)}$$

Conversely, the business owner should postpone operations if

$$R > \frac{(P + C)}{(P + L)}$$

Consider two examples of specific profit-loss scenarios. In Example A, a "high-risk" scenario, the loss from unexpected rain is severe: L=$3000, P=$1000, C=$500. In this case, the owner should proceed if R<0.375 (37.5%), and postpone if R>0.375. Note that if this business operates in a rainy area, where rain is likely on many days, then business will often be postponed.

Example B is a low-risk scenario in which the loss from unexpected rain is not severe, and the cost of postponing operations is relatively large: L=$1500, P=$1000, C=$1000. In this case, the owner should proceed if R<0.8, and postpone if R>0.8. This business scenario is quite tolerant of high rainfall probabilities; rainfall is not a major risk.

Note that if C>L, operations should always proceed regardless of the forecast, because it costs more to postpone than would be lost in an unexpected rain event.

A similar calculation can be performed for any business decision in which the financial outcome depends on the weather. For example, an electric utility may find that monthly average temperature is the key weather variable that affects profitability; let us also suppose that hedging activity can be undertaken if unfavorable weather is expected. If no hedging is performed and the monthly average temperature is below 10°C, then a profit P is obtained from increased electricity demand. However, if the temperature is above 10°C and no hedging is undertaken, then a loss L is incurred. A hedge may be placed in advance at cost C and with payout L if the temperature exceeds 10°C. In this example, a probabilistic temperature forecast is used to decide whether or not to place the hedge; the forecast must provide the probability of the average temperature exceeding 10°C for the month in question. It is simple to show that the hedge should be placed if the probability of unfavorable conditions R>C/L.

The following steps summarize the process that is needed to use probabilistic weather forecasts in a quantitative decision system similar to these examples.

- Identify a weather threshold that significantly affects the financial outcome of the business activity. This must be a discrete "yes/no" threshold that may or may not occur in a specified time period. A reliable probabilistic weather forecast must be available for this particular threshold.
- Compute L, P, and C from the financial history of the business. In the generic case, these metrics are identified as follows:
    - L is the average loss that occurs when business proceeds as normal, but unfavorable weather occurs.
    - P is the average profit that occurs when business proceeds as normal, and favorable weather occurs.
    - C is the average cost of disrupting operations or taking action to mitigate loss in anticipation of unfavorable weather.
- Use the conceptual framework illustrated here to compute the probability threshold R to use in making the "go/no-go" decision. When the forecast probability of unfavorable weather is less than R, proceed with business as normal. When the forecast probability is greater than R, take alternative action to avoid the potential loss.

4. Conclusion

World Climate Service S2S probability forecasts are designed to provide the necessary information to empower users to manage weather risk by making informed and confident decisions. While most business decisions are vastly more complex than the simple examples outlined in Section 3, the principles of decision-making in response to probability forecasts are still applicable, and indeed these principles may be employed in a very broad range of human endeavors. The WCS believes there is wide scope to expand the use of probability forecasts based on these principles and aims to develop tools to facilitate this goal. We also welcome interaction with customers regarding the issues discussed in this paper.