Performance and usefulness of ensemble based probabilistic forecasts

Frédéric Atger, Météo-France, Toulouse (frederic.atger@meteo.fr)

1. Introduction

Ensemble prediction systems (EPS) have been run operationally for several years. Yet forecasters make rather little use of probabilistic guidance for routine tasks. Most end products, in Europe and elsewhere, are still expressed in a purely deterministic form. Although meteorologists are well aware of the limitations of weather prediction, they seem to ignore this intrinsic uncertainty when elaborating end forecasts. Correspondingly, most weather forecast users are not convinced that they could get benefit from a probabilistic formulation. Many scientific, technical and human factors probably contribute to this failure in generalising the production and use of probabilities in weather forecasting. In this paper we focus on 3 conditions that are required for the success of a probabilistic approach. Section 2 deals with the design of probabilistic forecasts for the general public. Conditions for a proper estimation of the reliability of probabilistic forecasts are analysed in section 3. The crucial issue of the usefulness of probabilistic forecasts is discussed in section 4.

2. Designing probabilistic forecasts for the general public

As far as weather forecasts are concerned, it may be useful to distinguish two categories of users : (i) those who know precisely what weather information they need for taking their decisions; (ii) those who do not. At least in principle, satisfying the first category of users is relatively straightforward with tailor-made products, automatically prepared from a reliable estimate of the probability density function (pdf) of the future atmospheric state. Most professional users belong to this category. They are able to determine which meteorological conditions their activities are sensitive to. A probabilistic forecast is seen as indicating a risk that can be quantified in terms of consequences to these activities. A number of applications of probabilistic weather forecasting to various sectors of the economy have been described in the literature (e.g. *Taylor and Buizza* 2001).

Unfortunately, most users belong in fact to the second category, especially the general public. People do appreciate weather forecasts when they have a decision to take, but most of the time their concerns are rather erratic : cloud cover, temperature variations, occurrence of precipitation, risk of strong wind, etc, almost everything might have an impact on their decisions. Sometimes users know which parameter they are sensitive to (e.g. strong wind, heavy precipitations) but have no idea of an accurate critical threshold. A comprehensive list of probabilities of various weather events (different parameters and different thresholds) is unlikely to satisfy these users for obvious reasons.

The main requirement of the general public is a forecast that emphasizes the "important" aspects of the meteorological future. Although this is highly subjective, it seems that the description of the weather that is most likely to occur has some importance. At the same time, a low but significant risk of a severe event is obviously an aspect that is expected to be emphasized. To summarize, end forecasts should emphasize two aspects:

- A main forecast option: the prominent weather features (i.e. forecast with high probability) and/or the most probable range of various weather parameters (i.e. confidence intervals).
- Indications about significant alternatives: risk of detrimental or severe events, even when their probability of occurrence is low.

The first aspect can be seen as the deterministic component of the forecast, e.g. based on an ensemble mean or similar (*Atger* 1999). The second aspect is more specifically probabilistic, even if probabilities are not quantified but expressed by an appropriate wording. One reason explaining that most users accept reluctantly the introduction of probabilistic forecasts may be that the deterministic component of the forecast is missing. For example, a high probability of rain does not replace the statement: "a rainy day". On the other hand, the current practice of surreptitiously mentioning significant risks in the body of a deterministic forecast is not satisfactory, since it prevents users to evaluate actual probabilities. A clear separation between deterministic and probabilistic aspects, both emphasized in the formulation of weather forecasts, might be a key condition for users accepting probabilities.

3. Estimating the reliability of ensemble based probabilistic forecasts

Both users and forecasters often argue about the performance of probabilistic forecasts. Probabilistic forecasts would be useless just because they are not sufficiently performant. Reliability is an essential and the most intuitive attribute of this performance. Reliability can be seen as a probability bias: it indicates the correspondence between a given probability, and the observed frequency of an event in the case this event is forecast with the given probability. The reliability of ensemble based probabilistic forecasts appears highly variable in space and time (*Atger* 2002). For example, large discrepancies can be found over Europe, sometimes between locations that are not very far one from another.

A proper evaluation of the reliability of ensemble based probabilistic forecasts thus requires a local verification. An evaluation that would not take into account the spatial variability, e.g. an evaluation based on data gathered from different locations, is likely to lead to an overestimation of the average performance. This is because reliability is a bias, and because biases tend to compensate one for another. Furthermore, as far as weather forecasts are concerned, the point of view of the user is generally local. Local variations are thus an important aspect of the performance that cannot be neglected.

Ensemble and model characteristics are changing rapidly, so that accumulating large samples of data is a difficult challenge. A high stratification of the data is needed when one wants to consider each location separately. This is likely to lead to ill-sampling and therefore to the loss of significance of the results. A compromise has thus to be found between significance and credibility. When considering extreme events, i.e. events that occur rarely, this compromise might just be a dream.

The high variability of the reliability of ensemble based probabilistic forecasts is partly a consequence of model biases. Therefore, one way to solve the dilemma described above might be a local correction of model biases, e.g. through statistical procedures as model output statistics methods (MOS). After reduction of the variability, the evaluation of the reliability could be based on larger samples of data from different locations.

4. Estimating the usefulness of probabilistic forecasts

The usefulness of end forecasts is a crucial issue for analysing the conditions of success of a probabilistic approach in meteorology. The question is whether, and to which extent, a probabilistic weather forecast (and more generally a weather forecast) has some value for the user. Most recent studies dealing with this question have been based on a cost/loss analysis, i.e. on a very simple model that mimics the process of taking a decision in an economic context (e.g. *Zhu et al.* 2001). This model obviously matches the behaviour of some professional users belonging to the first category described in section 2. On the other hand, the decision process appears not as straightforward when considering the large majority of users, those belonging to the second category, and even some users of the first category. There are (at least) two assumptions when using the cost/loss model for estimating the value of probabilistic forecasts:

- Decisions are taken according to the expected expense (or benefit). This means that the subjective value of the forecast (i.e. the real value, as it is appreciated by the user) is its objective (i.e. economic) value.
- Decisions are effectively taken according to the forecast probability. This means that the subjective probability (i.e. the probability used for taking the decision) is indeed the probability that has been provided.

The first assumption is the neutrality to risk. Many studies have shown that people are in fact very sensitive to the level of risk, i.e. their appreciation of an expected loss (or benefit) depends on the probabilities involved. For example, most people prefer an 80% probability of losing £1000 (i.e. a 20% probability of not losing) than a 100% probability of losing £800, while the expected loss is identical. This inclination to risk is a common feature that has been widely described in the literature (e.g. *Alhakami and Slovic* 1994). On the other hand, there is generally an aversion to risk when considering expected benefits instead of losses. These features are highly variable according to individuals and cultures, but they clearly indicate that neutrality to risk might not be an appropriate assumption for evaluating the subjective value of probabilistic forecasts.

The second assumption is that people take rational decisions from objective pieces of information that are made available to them. Obviously this is not the case most of the time. The way people elaborate subjective probabilities has been extensively studied in the field of social sciences (e.g. *Tversky and Kahneman* 1981). For example, it has been shown that human beings underestimate uncertainties. High probabilities tend to transform into certainties, while low probabilities tend to just be ignored. At the same time, subjective probabilities do not follow the axioms of the theory of probabilities. For example, adding the probability of event *A* and the probability of event *not-A* may not lead to 1. This has certainly important consequences on the way people interpret probabilistic weather forecasts.

Research is certainly needed in order to confirm whether these general results are relevant for orientating the way we evaluate the value of probabilistic weather forecasts, especially the "social" value, i.e. the global value for the whole community of users. The characteristics described above might indicate that probabilistic statements have to be expressed carefully so that subjective judgments of end users eventually reflect the actual risks. Because their decisions are affected by the same "biases" as the users, operational forecasters have probably a crucial role to play in interpreting the probabilistic guidance and communicating to the public their own subjective assessment of weather related uncertainties.

References.

Alhakami, A. S. and P. Slovic, 1994. A psychological study of the inverse relationship between perceived risk and perceived benefits. *Risk Analysis*, **14**, 1085-1096.

Atger, F., 1999. Tubing: an alternative to clustering for the classification of ensemble forecasts. *Wea. Forecasting*, **14, 5**, 741-757.

Atger, F., 2002. Spatial and interannual variability of the reliability of ensemble based probabilistic forecasts. Consequences for calibration. *Mon. Wea. Rev.*, submitted.

Taylor, J. W. and R. Buizza, 2000. Using weather ensemble predictions in electricity demand forecasting. *Int. J. Forecasting*, submitted.

Tversky, A. and D. Kahneman, 1981. The framing of decisions and the psychology of choice. *Science*, **211**, 453-458.

Zhu, Y., Z. Toth, R. Wobus, D. Richardson and K. Mylne, 2001. The economic value of ensemble based weather forecasts. *Bull. Amer. Meteorol. Soc.*, **83**, 73-83.