Seasonal forecasting and daily probability

Michel Déqué and Jean-Philippe Piedelievre
Météo-France/CNRM, CNRS/GAME, Toulouse France (deque@meteo.fr)

Seasonal prediction is not a deterministic prediction. One can also consider that a 24h forecast contains a probabilistic part, in particular as far as local extreme events are concerned. But seasonal forecast is a climatology forecast. One attempts to predict some statistical properties of one or several seasons in advance, but not the chronology. Predicting the date (or rather the week) of the monsoon onset, even with a probability aspect, is a meteorological forecast. It should be considered as monthly or extended-range forecast, not as seasonal forecast. The validity of seasonal forecasting has first be proved in deterministic approach. Indeed, if the deterministic approach fails completely, the probabilistic approach cannot work. The probability theory does not create any information, it expresses our partial knowledge which leads to imperfect deterministic prediction. In order to validate the deterministic prediction in a robust way, scientists have used root mean square error or anomaly correlation of the seasonal mean. When going to probabilistic prediction (Doblas Reyes et al., 2000), they attempted to predict the seasonal mean with a probability density function (pdf).

From a scientific point of view, this approach is partial, because the model (or the nature) offers a succession of daily situations which form the weather (if chronology is taken into account) or the climate (if only the statistical properties are considered). From a user point of view, the seasonal average does not represent a useful information, in particular if it is given by a pdf. For heating or agriculture, degree days above or below a threshold, number of consecutive dry days . . . are more useful. One case in which seasonal average is the needed product is management of water in big dams: the seasonally accumulated precipitation is the right input variable in the decision making process.

EUROSIP is a consortium born after the DEMETER European project. Three partners (ECMWF, Met Office and Météo-France) agree to produce each month at the same date 41 members of a 7-month forecast. A hindcast period (to evaluate model climatology and prediction scores) is based on 1987-2007 with 11 member ensembles. Here we examine 2 m temperature forecasts for December January February obtained from the early November issue.

Verifying a probability forecast is less straightforward than a deterministic one. A good choice is to use square differences because they are additive with respect to time and space. The square difference between observed and predicted pdf could have been a simple and good idea, but in the case of temperature and precipitation, it does not take into account the fact that they are ranked variables. For example if the predicted and observed pdfs are quasi-deterministic (Dirac distributions), an error for temperature of 1°C or an error of 10°C yield the same distance. For this reason, the distance between the cumulative density functions (cdf) is preferred. Such a distance is named ranked probability score (rps). It is often used for a sall number of categories, traditionnally quantiles:

$$\text{rps} = \frac{1}{n}\sum_{i=1}^{n}\left[\text{prob}(T_{pre} < Q_i) - \text{prob}(T_{obs} < Q_i)\right]^2 \qquad (1)$$

where Q_i are the quantiles of the climatological distribution. Remember that here we deal with daily data of a season, so observation is also probabilistic. The continuous form (corresponding to $n \to \infty$) is:

$$\text{rps} = \int_{-\infty}^{+\infty} \text{pdf}_{cli}(t)\left[\text{cdf}_{pre}(t) - \text{cdf}_{obs}(t)\right]^2\, dt \qquad (2)$$

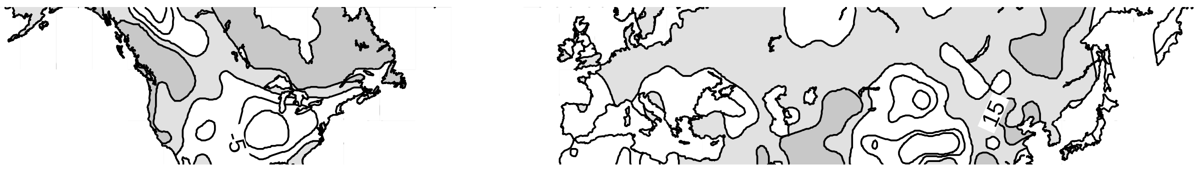

Figure 1: Ranked probability skill score for winter daily 2m temperature (land points only); contours ±5 and ±15%, shading above 5%.

In (2) the integral is calculated as a sum for bins of 0.5°C wide between -60°C and 30°C. The rps is a dimensionless quantity, with the size of a square difference between probabilities. To measure forecast skill, we need a reference. This reference is the minimum expected rps when prediction and observation are independent (i.e. no skill at all). It is obtained by replacing cdf_{pre} by cdf_{cli} in equation (2), which yields rps_{cli}. One can define a simple skill score as

$$\mathrm{ss} = (\mathrm{rps}_{cli} - \mathrm{rps})/\mathrm{rps}_{cli} \tag{3}$$

When this score is negative, the prediction is worse than the climatology prediction, which does not require costly numerical models. In this study we want to evaluate the skill for the mid-latitude (30°N-60°N) land points in EUROSIP. The first stage consists of calculating the climatological pdf on 21 winters (1987-2007) for ERA40 (extended by ERA interim beyond 2001), and the 3 models. Then a quantile-quantile correction (Déqué, 2007) is applied to each model, in order to correct the bias in pdf. For each temperature bin described above, the pdf and cdf are calculated by simple counting. Since the resulting pdfs are jaggy, a gaussian kernel smoothing is applied with 1 K standard deviation. Then the rps and ss are computed according to equations (2) and (3). The area mean rps is 0.0113 for climatology and 0.0106 for multi-model, leading to a skill score of 6.3%.

This method is not fully fair because the climatological pdf used in the correction introduces data to be predicted, that are not available in a real prediction. In fact, we do not correct the predictive behavior of each model, but only its climate properties, irrespective of the year-to-year chronology. In addition, this "cheat" favors also the climatological prediction which is the competitor. We have recalculated the scores, using the traditional "leave but one" method. The climatology rps becomes 0.0126 and the multi-model rps is 0.0118. The skill score remains unchanged. Thus the mid-latitude score is significantly positive and higher than traditional scores obtained with seasonal means, in particular over Europe. The geographical distribution (figure 1) indicates that the score is mainly positive (75% of the domain). This distribution depends not only on the models, but also on the period 1987-2007. Because of the reduced sample size, we did not attempt to improve the scores by *a posteriori* correction of the reliability of probability forecasts.

Conclusion

The winter daily temperatures provided by EUROSIP have a non-negligible skill versus climatology over most parts (except eastern US and western China) of the midlatitudes.

References

Déqué, M., 2007: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. *Global and Planetary Change*, **57**, 16–26.

Doblas Reyes, F.J., Déqué, M. and Piedelievre, J.P., 2000: Model and multimodel spread in the PROVOST seasonal forecasts : application to probabilistic forecasts. *Quart. J. Roy. Meteor. Soc.*, **126**, 2069–2088.