EMS Annual Meeting Abstracts
Vol. 8, EMS2011-855, 2011
11th EMS / 10th ECAM

Forecasting the weather
- ensemble techniques
in probabilistic weather prediction

The information gain of probabilistic forecasts with coarser or finer discretizations

R. Peirolo

Centro Meteorologico Regionale, Servizio Meteorologico dell'Aeronautica Militare (Italian National Meteorological Service), Milano, Italy

A verification score based on the information added by a probabilistic forecast to that contained in the climatological distribution has been presented recently by Peirolo (2011). This score, called information gain (IG), is mathematically related to the well known ignorance score (Roulston and Smith, 2002): as a measure of the ignorance that was still "contained" in a probabilistic forecast and was removed by the realization of the event k with probability f_k they considered the quantity:

$$\mathrm{IGN}_f = -\log_2 f_k.$$

Considering the climatological distribution as a reference, and considering a discretization of the continuous predictand into 2^m climatologically equiprobable intervals, the information gain (IG) of forecast f is equal to:

$$\mathrm{IG}_f = \mathrm{IGN}_c - \mathrm{IGN}_f = m - \mathrm{IGN}_f \ .$$

The information gain over a dataset of forecast-observation pairs is of course the mean value.

Besides being strictly proper (Brocker and Smith, 2007), the IG can give a simple and convincing measure of the accuracy of a single forecast or of a set of forecasts (Peirolo, 2011). The idea is to construct an "equivalent probabilistic forecast" (EPF), with the same IG of the forecasting system under consideration, but wih the lowest m as possible.

Another advantage of the IG is that it extends easily to the continuous case. It can be shown (Peirolo, 2011) that $\mathrm{IG}_{f,(m)}$ is a growing function of m, tending asymptotically to $\mathrm{IG}_{f,(\infty)} = \log_2[f(x)/c(x)]$. Instead, $\mathrm{IGN}_{f,(m)}$ is divergent for $m \quad \rightarrow \quad \infty$

The information gain score has been tested on the ensemble prediction system (EPS) of the ECMWF (Peirolo, 2011). The chosen fields were 500 hPa geopotential and 850 hPa temperature. IG was calculated for every year from 2000 to 2008 (summer), on day 1, 3, 5, 7, 10 forecast. It can be seen that the information gain decays with increasing lead time and increases over the years, in agreement with the improvement of the model. The information gain of 850 hPa temperature field is lower than that of 500 hPa geopotential, as it is a less predictable variable.

Besides being an interesting verification score, the IG can also give to the forecaster an idea of how it is worth to detail a probabilistic forecast. We have seen that, considering discretizations of the continuous predictand into 2^m climatologically equiprobable intervals, $\mathrm{IG}_{f,(m)}$ is a growing function of m.But when the growth is extremely limited, is it worth to double the number of intervals considered? An heuristic rule could be that we stop when

$$\mathrm{IG}_{f,(m+1)} \quad < \mathrm{IG}_{f,(m)} + 0.1.$$

The maximum achievable increase of the IG when we pass from 2^m to 2^{m+1} intervals is of 1 bit. If the real gain is only one tenth of that, probably the more detailed forecast is just more difficult to be communicated and understood, without a significant increase in the information content.