Response to reviewer Uwe Ehret

We thank Uwe Ehret for his outstanding review, which added significantly to the discussion of the paper.

Relation between forecast consistency and uncertainty

We do fully agree with Uwe and Alberto Montanari that there are many definitions of uncertainty (although we are uncertain whether it is realistically possible and desirable to develop a general scientific framework to apply and quantify uncertainty – but that is a discussion which is beyond the scope of this paper). However, we do agree that the relationship between inconsistency and uncertainty needs to be explored, both conceptually and numerically, though in this opinion paper we restrict ourselves to first task on the grounds, as Uwe himself notes, that uncertainty schemes are so model specific that any exploration of the numerical relationship would require some concrete case studies of particular models and their applications.

Focusing on the conceptual issues, it might be said that (in)consistency is a distraction from the real issue, namely uncertainty. While we would acknowledge that (in)consistency is a manifestation of underlying uncertainties, we would insist that it is important to understand in its own right. First, the different types of heretofore poorly defined (in)consistency may help improve the understanding of different kinds and causes of forecast uncertainties. Second, it is has been documented that it is common in operational practice, to look to (in)consistency heuristically (sensu Nichols 1999) as a quick and dirty indicator of forecast uncertainty without always acknowledging that (in)consistency, like uncertainty itself, comes in many shapes and sizes. (In)consistency can be in the temporal, spatial and magnitude domains or any combination thereof. Spatial (in)consistency over an area can manifest itself as temporal/magnitude uncertainty at a point. However, it can be quantified (in what ever way) giving information about system attributes, which are different from the measure of uncertainty itself, and so specifying the kind of (in)consistency and calculating it objectively provides additional information.

So to summarize: we agree that (in)consistency is part of the overall uncertainty, however, it importantly can be quantified separately (although as rightly pointed out by Uwe Ehret, this requires more research). We will add a paragraph to the paper reflecting the discussion above.

Relationship between skill and jumpiness

Uwe Ehret has already offered one solution to explore the relationship between skill and jumpiness. This is a very valuable exercise and we would like to extend the concept using threshold exceedance (which is one example of a forecast property and not the only way which can be used to evaluate forecast inconsistency) as presented by Uwe to a continuous forecast framework (albeit using deterministic forecasts for demonstration). Our analysis is based on the publication by Persson and Grazzini (2007).

Lets assume that forecast accuracy is measured as the Root Mean Squared Error. We have two forecasts (g and f) and an analysis (a, observation) to which these forecasts are compared. This can be expressed in vector geometry (see figure 1). The difference between g and f is a measure for the jumpiness or inconsistency (blue line labelled f-g). The cosine of the angle between the vectors $\overrightarrow{ga}$ and $\overrightarrow{fg}$ is the anomaly correlation and can be used as a measure of this inconsistency.

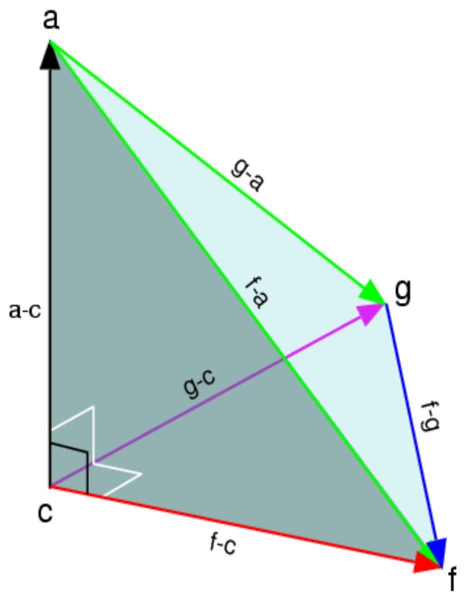

Figure 1: Illustration of two forecasts (g and f) and observations (a). Forecast errors (green line) represents the difference between forecast and analysis. Jumpiness are expressed as blue line indicating the difference between two forecast.

In figure 1 it is assumed that the two forecasts systems (f) and (g) lack predictive skill and are mutually uncorrelated therefore all three vectors ($\vec{a}$, $\vec{g}$ and $\vec{f}$) are perpendicular (90°). Whereas the analysis vector ($\vec{a}$) and the forecast vectors ($\vec{f}$ and $\vec{g}$) are perpendicular, their differences are not! Their mutual angles are 60° which implies correlations of 50%. This can be seen in figure 2 which is a rotated figure 1.

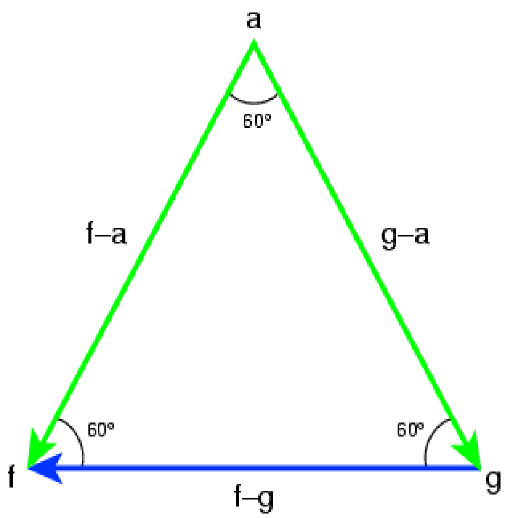

Figure 2: Relationship between forecasts and analysis in the case of lack of predictive skill and mutual uncorrelation.

This concept can now be extended to climatological forecasts in which it can be proven that this correlation is always 50%.

$$\overline{gf}^2 = \overrightarrow{ga}^2 + \overrightarrow{fa}^2 - 2 * \overrightarrow{fa} * \overrightarrow{ga} * \cos(\alpha) \qquad \text{Equation 1}$$

At long forecast ranges, the individual forecasts should converge to climatology meaning that c=$\overrightarrow{ga}$= $\overline{gf}$ =$\overrightarrow{fa}$, therefore:

$$c^2 = c^2 + c^2 - 2 * c^2 * \cos(\alpha) \qquad \text{Equation 2}$$

$$\cos(\alpha) = \frac{1}{2} \qquad \text{Equation 3}$$

qed

This also allows us to derive a relationship between jumpiness and skill (simply re-arranging equation 1)

$$\cos(\alpha) = \frac{\overline{gf}^2 - \overrightarrow{ga}^2 - \overrightarrow{fa}^2}{2 * \overrightarrow{fa} * \overrightarrow{ga}}$$

This means that if the skill of forecast increases then the correlation decreases (assuming that the spread between the forecasts is constant). Equally if the skill is kept constant more dissimilar forecasts will lead to an increased correlation. The concept presented could be extended to multiple forecasts or a weighting of forecasts according to their importance (e.g. a jumpiness in the latest forecasts maybe more unsettling as suggested by Ehret (2010)) and we will include this as an appendix to the paper

Comment on Presentation quality:

Section 3:

Uwe Ehret suggests that we should add equations and show how they are computed given the individual examples. This was deliberately avoided. We would have to compute different measures for the probabilistic and deterministic forecast and even may have to split them according to features (e.g. timing or magnitude). We do not believe that this added detail would contribute to the discussion. However as a compromise we decided to show how a consistency measured as RMSE (see above) will look like and what the effects are.

Section 5: agreed title will be changed

Section 6: agreed will be changed

Example forecasts: Will be changed to be coherent and address the issue regarding timing and amplitude (we agree with Uwe that the major inconsistency seems to be currently related to magnitude only and not timing.

All minor comments raised by the reviewer will be addressed in the revised version.

References

Demeritt D, Nobert S, Cloke H, Pappenberger F 2010 Challenges in communicating and using ensembles in operational flood forecasting ***Meteorological Applications*** 17: 209-22 DOI: 10.1002/met.194

Nichols N (1999) Cognitive illusions, heuristics and climate prediction *Bulletin of the American Meteorological Society* **7**: 1385-1397