Hydrol. Earth Syst. Sci., 15, 2391–2400, 2011
www.hydrol-earth-syst-sci.net/15/2391/2011/
doi:10.5194/hess-15-2391-2011

Hydrology and Earth System Sciences

HESS Opinions
"On forecast (in)consistency in a hydro-meteorological chain: curse or blessing?"

F. Pappenberger[1], **H. L. Cloke**[2], **A. Persson**[1,3,4], **and D. Demeritt**[2]

[1]European Centre for Medium Range Weather Forecasts, Reading, UK
[2]Department of Geography, King's College London, London, UK
[3]UK MetOffice, Exeter, UK
[4]Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

Received: 11 January 2011 – Published in Hydrol. Earth Syst. Sci. Discuss.: 26 January 2011
Revised: 27 June 2011 – Accepted: 5 July 2011 – Published: 26 July 2011

Abstract. Flood forecasting increasingly relies on numerical weather prediction forecasts to achieve longer lead times. One of the key difficulties that is emerging in constructing a decision framework for these flood forecasts is what to dowhen consecutive forecasts are so different that they lead to different conclusions regarding the issuing of warnings or triggering other action. In this opinion paper we explore some of the issues surrounding such forecast inconsistency (also known as "Jumpiness", "Turning points", "Continuity" or number of "Swings"). In thsi opinion paper we define forecast inconsistency; discuss the reasons why forecasts might be inconsistent; how we should analyse inconsistency; and what we should do about it; how we should communicate it and whether it is a totally undesirable property. The property of consistency is increasingly emerging as a hot topic in many forecasting environments.

1 Introduction

Flood forecasting increasingly relies on numerical weather prediction (NWP) forecasts to achieve longer lead times (see Cloke et al., 2009 and Cloke and Pappenberger, 2009). One of the key difficulties that is emerging in constructing a decision framework for these flood forecasts is what to do when consecutive forecasts are so different that they lead to different conclusions regarding the issuing of warnings or triggering other action. In this opinion paper we explore some of the issues surrounding such forecast inconsistency (also known as "Jumpiness", "Turning points", "Continuity" or number of "Swings"; Zsoter et al., 2009; Mills and Pepper, 1999; Lashley et al., 2008). We begin by defining forecast inconsistency; discuss the reasons why forecasts might be inconsistent; how we should analyse inconsistency; and what we should do about it; how we should communicate it and whether it is a totally undesirable property. The property of consistency is increasingly emerging as a hot topic in many forecasting environments (e.g. for a discussion on NWP inconsistency see Persson, 2011). However, in this opinion paper we restrict the discussion to a hydro-meteorological forecasting chain in which river discharge forecasts are produced using inputs from NWP. In this area of research (in)consistency is receiving recent interest and application (see e.g. Bartholmes et al., 2008).

1.1 What is (in)consistency?

Forecast consistency refers to the degree to which two forecasts agree about the magnitude, onset, duration, location or spatial extent of a given event. In hydrological forecasting we are typically interested in comparing the degree of consistency between consecutive point or grid based forecasts of river discharge from the same model issued at different times. However with the emergence of ensemble and grand-ensemble techniques (Pappenberger et al., 2008), this involves assessing the consistency between forecasts made for the same location and time for different model set-ups and iterations, though as we discuss below, ensemble methods lead to further complications. If forecast consistency is the

Correspondence to: F. Pappenberger
(florian.pappenberger@ecmwf.int)

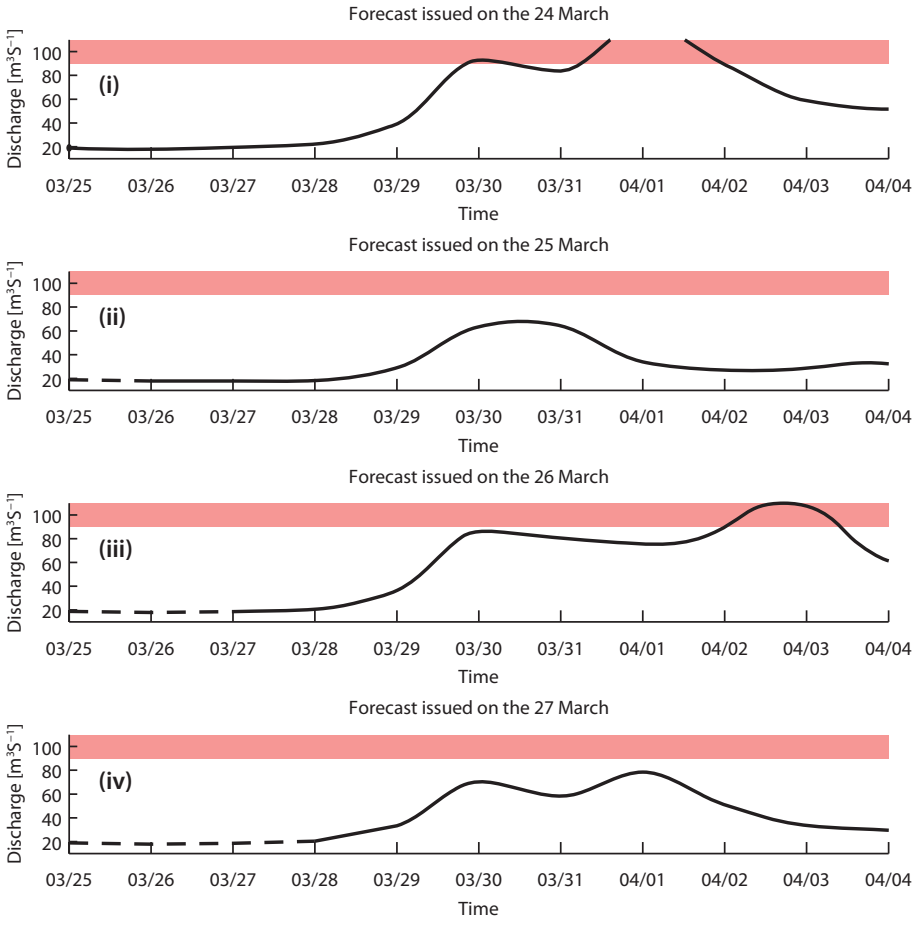

Fig. 1. Four different forecasts are shown issued on the (i) 24, (ii) 25,(iii) 26 and (iv) 27 March for a station along the river Severn (hypothetical case). The dashed line indicates the observations. The solid area represents a warning threshold. A flood alert would be issued in case (i) and (ii).

degree of agreement between two different forecasts made for the same future point in time, forecast inconsistency occurs when sequences of (temporally) consecutive forecasts of a variable develop differently and so exhibit a change in behaviour in some way from one another about their predictions of what is going to happen.

1.2 Deterministic flood forecasts

In Fig. 1 this is illustrated for a (deterministic) forecast showing inconsistency in the magnitude of forecasted peak discharge for a hypothetical case. These forecasts are from the same model with the same structure and equations. The figure shows four different forecasts for station X issued on the 24, 25, 26 and 27 March. The dashed line indicates the observations and the solid area represents a warning threshold. The first forecast (i) indicates a possibility of a flood on the 30 March and has a very clear signal, in terms of threshold exceedance, on 1 April. In the next forecast (ii), issued on 25 March, the threshold exceedance signal has disappeared. Forecasted river levels exceed warning levels again in the forecast issued on 26 March (iii), but one/two days later, while the forecast (iv) issued on 27 March does not predict flooding. As is typical for many flood forecasting systems, in our case a flood warning is issued depending on whether (or not) a river discharge is exceeded (see Fig. 2).

Table 1 shows a typical forecast overview diagram for our hypothetical case. The rows indicate the date and time that the forecast was issued and the columns indicate the date for which the forecast was issued. As the table clearly shows there is inconsistency between the forecasts, and river discharge threshold exceedance is variously forecast to occur on both 30 March and 1 April, on either date or neither. Hence inconsistency is demonstrated in the timing of the flood event as well as whether the event happens at all.

While this kind of (in)consistency in forecasts of threshold exceedance might be dismissed merely as a statistical artefact of translating continuous forecasts into binary yes/no categories for given dates and times, it is one that operational forecasters are acutely conscious of, both because threshold

Table 1. Inconsistent threshold exceedance according to Fig. 1. The rows indicate the date and time that the forecast was issued and the columns indicate the date for which the forecast was issued.

Forest day	23	24	25	26	27	28	29	30	31	1	2	3	4
22 March													
23 March								X		X			
24 March										X			
25 March													
26 March											X	X	
27 March													
28 March										X	X		
29 March													
30 March										X			
31 March										X			

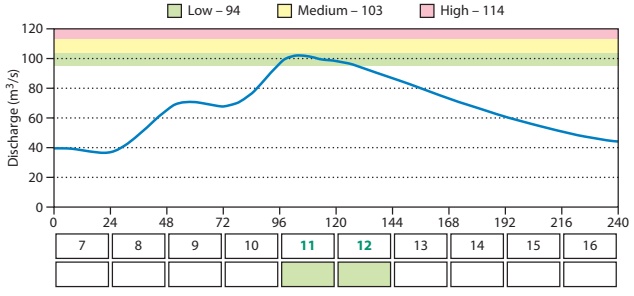

Fig. 2. Threshold exceedance on the example of a forecast issued by the European Flood Alert System. In the top figure, the blue line represents a single deterministic forecast. The plot shows three different warning levels (green, yellow and red). The threshold exceedance of the green level is shown in the table plot underneath. When an ensemble of forecasts is used (rather than just one deterministic forecast) the number of ensemble members exceeding the threshold level would be also shown in the table plot.

exceedance is the primary basis by which skill scores and the operational performance of the forecasting centre are assessed and because threshold exceedance also serves as triggers for warnings and other management responses to flood forecasts. Thus (in)consistency in forecasts of threshold exceedance is an issue of real concern for forecasters. However, it is not the only kind of (in)consistency that might be of interest. Though less commonly considered, (in)consistency could also be related to other hydrograph properties such as the length of time the water level stays above (or below) the threshold or the magnitude by which some threshold is exceeded.

1.3 Ensemble flood forecasts

Further complexity is added by the combination of various forecasts into ensemble forecasting systems. Many modern flood forecasting systems rely not only on deterministic forecasts, but also on ensemble forecasts (and a combination thereof). In this situation, in addition to the above mentioned definitions, it is necessary to define inconsistency thresholds based on the number of ensemble members[1] (either in the form of frequency or probability) over a warning discharge threshold for a given location.

Consider the example of an alert chart from the European Flood Alert System (Table 2). As in Table 1, the rows indicate the date and time that the forecast was issued and the columns indicate the date for which the forecast was issued. However, this time the table shows the number of ensembles (out of 51) exceeding a high alert level (Thielen et al., 2009a,b; Pappenberger et al., 2011a). For a series of ensemble forecasts of this sort there are different ways in which it is possible to define (in)consistency between consecutive ensembles (represented by rows in the table):

1. In terms of the number of ensemble members exceeding various discharge thresholds: in this case, the differences for consecutive forecasts range from 0 to 35 between different forecasts. A difference of 35 can be observed between the midnight and noon forecasts issued on 10–14 for day 15.

[1]Ensemble members could here be several lagged deterministic forecasts or an ensemble forecast or a combination thereof – the way an ensemble is created will have an impact on its level of consistency.

Table 2. Number of Ensemble members (out of 51) exceeding a high alert level. The rows indicate the date and time that the forecast was issued and the columns indicate the date for which the forecast was issued.

Forecast day	11	12	13	14	15	16	17	18	19	20	21	22	23
11 October, 00:00						5	11	5	1	1			
11 October, 12:00					6	33	25	5	5	2			
12 October, 00:00						45	50						
12 October, 12:00					14	47	21	21	1	1	1		
13 October, 00:00					2	41	31	5	10	12			
13 October, 12:00					30	45	7	7	7 7				
14 October, 00:00						51	28	21		12			
14 October, 12:00				5	35	11	20	11	1				

2. The onset of the flood varies between the 14/15 and 16.

3. The flood lasts from anything between 4 days to 2 days.

4. It exhibits a single or double peak.

2 Why are forecasts inconsistent?

Forecast inconsistency comes from various imperfections in the forecasting chain. In medium range NWP the most significant cause of inconsistency is errors in the specification of initial conditions for a non-linear dynamic model so that even with a "*perfect*" model (that is one with a perfect representation of the physics of atmospheric processes, if that can exist), and thus inconsistency between forecasts is unavoidable. NWP models were more consistent 20–30 years ago because the poverty of their representations of atmospheric processes and their low spatio-temporal resolutions made them less sensitive to variance in the specification of initial conditions. Thus reducing the quality of the NWP model with respect to variability, for example by reducing the resolution or putting in diffusion, would improve consistency, but reduce overall predictive skill (Simmons et al., 1995). In general increases in model complexity will generally lead to a decreased bias in calibration mode and to increased uncertainty. This may not be the case in forecasting (where higher complexity will also lead often to an increased bias). There is an interesting philosophical question here of what model "quality" might mean and whether the atmospheric modelling community has respected the principle of parsimony. That is a very important discussion, but it goes far beyond the immediate focus of this opinion paper.

At the end of the hydro-meteorological forecasting chain, inconsistency is complicated by the nonlinear interaction between all imperfections (including initial conditions, forcing, model parameterization, observations etc.; note we assume that every forecast system is always imperfect due to hydrological uncertainty, see Beven, 2006). As a result, the relative importance of different sources of uncertainty for forecast consistency will depend on exactly which aspect of forecast inconsistency (i.e. the timing or magnitude of the flood peak, its spatial extent or temporal duration) one is concerned with. In the case of convective flash flooding, for example, forecasts are typically less consistent than largely synoptic scale driven floods partially because of the high uncertainties involved in modelling convective rainfall location and timing at high resolution (Gupta et al., 2002). Indeed for flash flood forecasts inconsistency about the predicted location of flooding is common and the tendency is to remain on flood alert while the possibility of a flash flood exists even if the uncertainty about its exact location is high. Inconsistency here, is clearly defined as shifts in location and from a hydrological standpoint these shifts can have very dramatic effects if there are several flash flood prone catchments in the region or if shifts in the localization of convective rainfall simply means that the rain is falling on the 'non flash-flood producing side' of the valley.

The problem of forecast inconsistency is some way eased through ensemble forecasting as the ensemble will intrinsically "blend out" individual jumpy forecasts as well as providing a better characterization of initial condition/model uncertainty. On the other hand, however, it also makes the conceptual problem of defining in just what sense one set of model runs (individual ensemble members) might be "consistent" with the next more, not less, difficult. Inconsistency exists mainly due to the imperfection of the actual ensemble design e.g. limited number of members and underdispersivity and thus remains a significant challenge to the forecaster.

3 Quantifying inconsistency

Quantifying inconsistency can be useful but only when it is accompanied by an understanding of why the inconsistency occurred. Here we make a (unrealistic) binary divide between expert users, such as those involved in producing hydro-meteorological forecasts, and non-experts users of hydro-meteorological forecasts among the general public in order to illustrate extreme positions. We note that in reality there is less differentiation between these groups (Nobert and Demeritt, 2011; de Roo et al., 2011).

It is important for *expert users* to find robust ways to identify inconsistency and express it numerically in order to aid their decision making, clarify system limitations or assess the performance of different forecast systems and aid their operational decision making. Examples of evaluation measures include regression, root mean squared error and bias based approaches (Nordhaus, 1987; Clements, 1997; Clements and Taylor, 2001; Mills and Pepper, 1999; Bakhshi et al., 2005) and pseudo-maximum likelihood estimators (Clements and Taylor, 2001). In weather forecasting a latitude weighted root mean squared error (Zsoter et al., 2009), the Ruth-Glahn forecast convergence score (Ruth et al., 2009) and the Convergence Index (Ehret, 2010) have also been used. Pappenberger et al. (2011b) have applied the latter to probabilistic hydro-meteorological forecasts. The number of different ways in which it is possible to quantify inconsistency introduces its own level of uncertainty to the evaluation, but it remains essential to quantify it in some (or many) numerical ways and understand these relationships (similar to other skill scores see Cloke and Pappenberger, 2008).

In contrast *non-expert users* may not necessarily benefit from such quantitative measures of forecast inconsistency, as they would be able to see for themselves that the forecast has changed. Inconsistency in these circumstances has to be accompanied by an explanation of why it occurs as well as an analysis that is understandable in lay terms, which might well involve verbal qualitative (rather than a numerical) description, though it would, of course, be based on a numerically computed evaluation for and by the expert user.

4 Consistency and uncertainty

It could be argued that inconsistency is simply another term for uncertainty. Uncertainty itself is a difficult term to define and quantify (see Montanari, 2007) and therefore any exploration of the numerical relationship requires some more concrete case studies of particular models and their applications, which is beyond the scope of this opinion paper. 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.

5 Consistency and forecast performance

It could be hypothesised that consistency is an indicator of a "better" forecast. However we would like to highlight the important fact that, the theoretical basis for this is not necessarily clear cut. Persson and Grazzini (2007) demonstrated that correlation between forecast jumpiness and forecast error (typically 30 % according to investigations by see e.g. Hoffman and Kalnay, 1983; Dalcher et al., 1988; Palmer and Tibaldi, 1988; Roebber, 1990 and others) is a statistical artefact. Inconsistency is clearly related to forecast error, but consistency should not be used as a proxy for forecast accuracy (Hamill, 2003), nor does it qualify as a predictor of a priori skill.

Ehret (2011) has put forward one interesting method to explore the relationship between skill and jumpiness by using a threshold approach. Here we offer a solution based on 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 Fig. 3). 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 $\boldsymbol{ga}$ and $\boldsymbol{fg}$ is the anomaly correlation and can be used as a measure of this inconsistency. In Fig. 1 it is assumed that the two forecasts systems (f) and (g) lack predictive skill and are mutually uncorrelated therefore all three vectors ($\boldsymbol{a}$, $\boldsymbol{g}$ and $\boldsymbol{f}$) are perpendicular (90°). Whereas the analysis vector ($\boldsymbol{a}$) and the forecast vectors ($\boldsymbol{f}$ and $\boldsymbol{g}$) are perpendicular, their differences are not! Their mutual angles are 60° which implies correlations of 50 %. This can be seen in Fig. 3 which is a rotated Fig. 4.

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

$$\boldsymbol{gf}^2 = \boldsymbol{ga}^2 + \boldsymbol{fa}^2 - 2 \cdot \boldsymbol{fa} \cdot \boldsymbol{ga} \cdot \cos(\alpha) \quad (1)$$

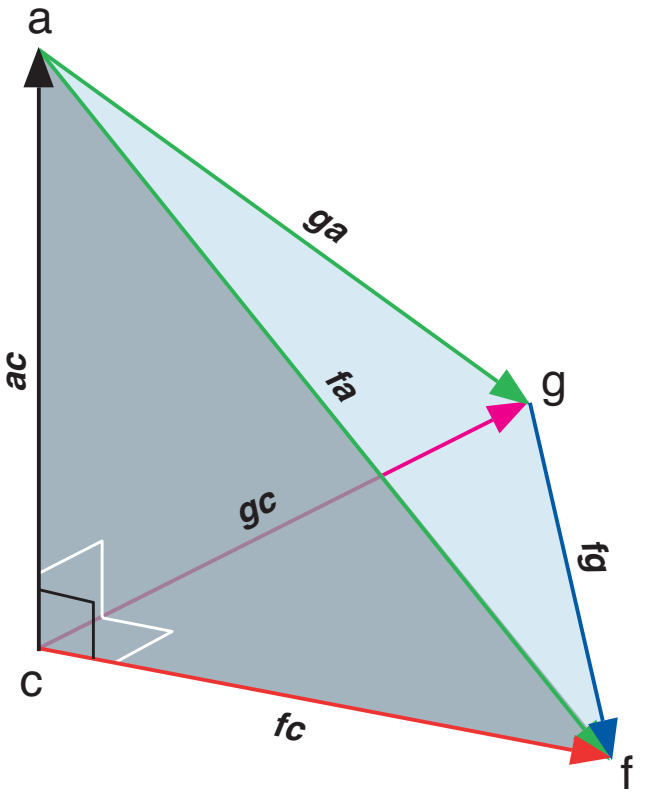

Fig. 3. Illustration of two forecasts – **(g)** and **(f)** – and observations **(a)**. Forecast errors (green line) represents the difference between forecast and analysis. Jumpiness are expressed as blue line indicating the difference between two forecast.

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

$$c^2 = c^2 + c^2 - 2 \cdot c^2 \cdot \cos(\alpha) \quad (2)$$

$$\cos(\alpha) - \frac{1}{2} \quad (3)$$

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

$$\cos(\alpha) = \frac{\boldsymbol{gf}^2 - \boldsymbol{ga}^2 - \boldsymbol{fa}^2}{2 \cdot \boldsymbol{fa} \cdot \boldsymbol{ga}} \quad (4)$$

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.

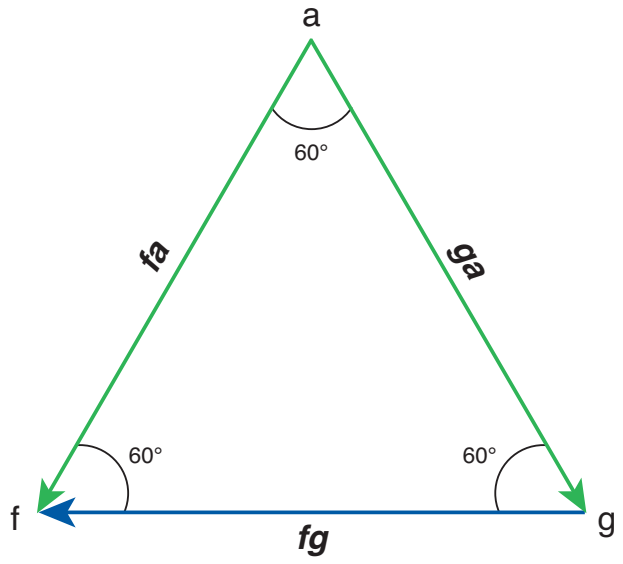

Fig. 4. Relationship between forecasts and analysis in the case of lack of predictive skill and mutual uncorrelation.

6 The problem of inconsistency

Forecasting preference is usually for consistency. Any forecaster would ideally like to issue a flood warning as early as possible, minimize the error and then update the forecast in continuous way. However, hydro-meteorological forecasts of flooding are typically subject to high uncertainty not only due to the quality of NWP-or radar-based forecasts, but also due to the rarity of flood events, which makes it difficult to validate model predictions of them. Flood forecast recipients face similar problems. Unlike daily weather forecasts, which members of the public are accustomed to using and evaluating, flood alerts and other warnings of extreme weather are so rare that there is not the same intuitive feel for how much stock to put in them or how best to respond to uncertain warnings of impending disaster.

One response to the challenge of decision-making in the face of inevitable uncertainty about forecast accuracy is to establish a cost-loss function, so as to weigh up the relative costs that would be incurred by taking precautionary action in response to the forecast against the losses that would be incurred if the forecast is ignored and yet proves correct (Murphy, 1977; Richardson, 2000; Roulin, 2007; Laio and Tamea, 2007). However, actually establishing this functional relationship is complex, and the values associated with some costs and losses cannot be easily reduced to monetary ones as is required for a cost-benefit type calculation (Davies and Demeritt, 2000). What value should be put on a life? The question is incalculable (Demeritt and Rothman, 1999), and when the values at stake are sufficiently high – whether in terms of lives and property, dread risk (i.e. nuclear accidents or terrorism; cf. Slovic, 1987), or the reputational costs of getting it wrong – then cost/loss functions often go out

the window, and pre-emptive action is taken regardless of whether one gives much credence to the likelihood of the forecasted event. Moreover hits, misses, false alarms and correct negatives often have significantly different weight from each other in flood forecasting (Demeritt et al., 2010; Ramos et al., 2010) than the weight that is implied by a standard contingency table (Bartholmes et al., 2009). There is also the key issue that what counts as a meteorologically correct forecast (i.e. rainfall $> 30\,\mathrm{mm\,h^{-1}}$, which is the design capacity for urban drainage) may not result in flooding, so that forecast recipients are measuring something slightly different than forecasters themselves when they evaluate what stock to put in the warning.

Flood forecasters are well aware of the problem of "crying wolf" and the risk that a sequence of false alarms will result in people no longer taking action and hence increase the costs of a hit (value of losses!). In addition a miss can be catastrophic for the individuals directly affected by the flooding and also for the organisation which failed to alert (Dedieu, 2010). Consideration of reputational damage plays an important role in flood forecasting and consequentially has to be added to the cost, which can be different for different people given the same event. The cumulative effect of these two peculiarities suggests why flood forecasters are often unwilling to update a previously issued warning simply based on the latest new forecast (Demeritt et al., 2010; Ramos et al., 2010; Norbert et al., 2010)[2]. Therefore, reducing the false alarm rate and strong autocorrelation[3] between warnings both play a strong role in the design of any flood warning system. But this is just one kind of error: the false positives (type 1) error.

There is also the type 2 error of missed events. While EPS helps to increase sensitivity to possible surprise, and so decreases the frequency of type 2 errors, it tends (with low thresholds needed to avoid type 2 errors) to increase the number of type 1 errors. In the case of the EFAS, lagged forecasts are used to reduce this sort of error, and this temporal consistency, or persistency, of forecasts is then built into the decision making process (Bartholmes et al., 2009): At least three consecutive flood forecasts must predict that a critical discharge threshold will be exceeded for the same river stretch, for an EFAS flood alert to be issued. This use of temporal consistency reduces the number of false alarms and at a minimal cost to the overall hit rate. It does, however, lead to under-forecasting, which may not always be desirable and strongly depends on the envisaged lead time. While Bartholmes et al. (2009) demonstrate that this use of temporal consistency is the best solution for the particular institutional context of the European Flood Alert System, it is important to recognize that different uses for forecast consistency may be necessary for other forecasting contexts.

[2]On the one hand, this may be analogous to the pre-NWP model culture that existed in meteorology and so it may be that flood forecasters eventually will also adopt the approach of always using the latest forecast the more they get used to meteo-hydrological forecasting chains. But, on the other hand, hydrological forecasting has its own longstanding traditions and these, combined with the rarity of flood event, may well mean that attitudes do not evolve as they have done in meteorology.

[3]This auto-correlation stems partially form the fact that discharge is a highly auto-correlated variable, however one could also as whether this autocorrelation stems more from an "anchoring bias" around the initial warning – a more detailed discussion is beyond the scope of this paper.

7 The uses of inconsistency

Despite the preference of hydrological forecasters for consistency one should not ignore the advantages of inconsistency. Inconsistency discourages the forecaster from relying on the latest forecast, and instead seeking out alternative information in an ensemble system in addition to the forecasted hydrograph values, as well as considering previous forecasts or information from other models. Persson and Grazzini (2007) argue that a consistent forecast may lull forecasters into a false sense of confidence in the reliability of their model, which exacerbates difficulties in decision making when sudden surprising forecasts arise. In the same way a gradually changing forecast may contribute to greater confidence than an abruptly changing one (Lashley et al., 2008) and thus the magnitude of inconsistency is of particular importance. Inconsistency can thus be an asset if it alerts forecasters to possible forecast problems and highlights alternative developments (see full details in Persson and Grazzini, 2007).

To illustrate these benefits of inconsistency, we refer back to Table 1. It can be clearly seen that a flood event could occur between the 31st of March and 3rd of April. Here we would argue that a warning should be issued at the 26.03 stating that there is the possibility of a flood between 31 and 4 April. This warning should stay in place until 28 March, when it is changed to the fact that the flood may happen on 1 April. In this way the communicated warning would have a considerable consistency but still allow for the ambiguity seen in an otherwise deterministic forecast. In reality many countries have several warning levels, ranging for example, from "flood watch" over to "flood warning" and "severe flood warning". The ramping up of a warning level from no warning to flood watch is probably a tolerable level of inconsistency; however fluctuating between flood watch and flood warning or severe flood warning could be seen as intolerable fluctuation. In order to avoid people letting their guard down too early, the rule of thumb seems to be that warning levels go up, but not down, until the crisis has passed (Demeritt et al., 2007). The hypothetical flood event described in Table 1 may in fact not have happened and a false warning would have been issued. However it is inevitable that we will sometimes get it wrong, and so we need to ensure that our process of forecast interpretation and warning is both robust, so that the prevalence of error can be reduced as much

as possible, and also clear, so that forecast recipients can assess how much confidence to place in them and that when mistakes are made, lessons are learned.

8 How to deal with (in)consistency – codes of practice

(In)consistency in forecasts is unavoidable given the imperfections of forecast systems. The challenge is how to deal with the difficulties – and opportunities – presented by it. In this paper we have shown that there are different aspects or dimensions of forecast (in)consistency: temporal persistence, value magnitude, spatial pattern etc.). It may well be that these various dimensions are more important for some purposes than others. However this is yet to be pinned down in forecasting practise.

It is clear that forecast inconsistency is one part of the total uncertainty and analysis of it needs to be communicated alongside the forecast itself as part of a wider framework for decision-making. The challenge of communicating inconsistency is thus embedded in the challenge of communicating uncertainty where a close relationship with forecast end users is key (see Norbert et al., 2010; Faulkner et al., 2007; NRC, 2006). It may well be that more trained experts are better able to deal with inconsistency whereas it may cause a loss of confidence among less well qualified audiences (Lashley et al., 2008). However, the situation may well be far more complex than this (for example the uncertainty trough as postulated by MacKenzie, 1990, and Shackley and Wynne, 1995). To date, inconsistency has not been adequately discussed with forecast end users or indeed within (and between) the connected but distinctive meteorological and hydrological forecasting communities. For these products we strongly advocate future discussion and research in this area. For example, post-processing methods may reduce inconsistencies (Bogner and Pappenberger, 2011).

As a first suggestion, a code of practice with respect to forecast inconsistency of any forecast system may be:

1. *Define inconsistency in the context of the particular forecast task:* Is a forecast which first predicts $10\,m^3\,s^{-1}$ above a "medium" warning level (of let us say $100\,m^3\,s^{-1}$) and then $5\,m^3\,s^{-1}$ above inconsistent? How much does a probabilistic forecast have to change to be inconsistent?

2. *Involve your end users in developing inconsistency forecast products:* It is important that the way one decides to illustrate and demonstrate inconsistency is developed in collaboration with the end users that information is designed to inform. Similar to warning system, these types of products cannot be designed at a scientist's/forecaster's desk alone.

3. *Establish the magnitude of inconsistency and its dependency on catchment location, hydrological and meteorological attributes.* Inconsistency will heavily depend on catchment properties such as catchment response time. Flash flood forecasts on the medium range will be highly inconsistent. In contrast forecasts which rely on a longer channel routing process with ample opportunity to be updated will exhibit less inconsistency (although perhaps at least in some dimensions, such as the size of the flood peak, its timing or the resulting spatial inundation pattern it may well be much more uncertain). Inconsistency can depend on seasons (see Pappenberger et al., 2011) and the degree of this dependency must be understood.

4. *Make it a clear part of your decision making and communication framework*

    a. *Establish the nature and magnitude of inconsistency with which you and your end users are comfortable* in issuing decisions/warnings

    b. *Anticipate forecast inconsistency in your decision making* (rather than just reacting to it in a post event analysis setting). This means, if you expect high inconsistency because of the season or domain in which you are working, make sure that you anticipate in your decision making and communication process that it could happen.

    c. *Clearly communicate in your warnings and decisions the level of inconsistency* at a level appropriate to the end user. As illustrated above, communication has to be targeted and not necessarily "numerical" (see also section on quantifying inconsistency). A good (but as yet unanswered) question is whether it would be better to be able to add this to the total uncertainty of your system in your communication process or whether it needs to be treated and communicated separately. This will be strongly end-user dependent. For untrained end-users all sources of uncertainty may best be folded into a single presentation, for trained end-users, which have to rely on additional decision making processes, the separation of uncertainty sources is vital

    d. Above all: do not confuse end-users unless they are clearly involved in the process and understand what you are talking about (and you understand what they want from you!)

9 Conclusions

Flood forecasting based on numerical weather predictions remains a relatively new field and using probabilistic forecasts is an even younger discipline and hence the guidelines above are only a very first step to initiate the discussion in this field. We expect them to be evaluated and revised. We encourage

all flood forecasters researching and practising in this area to routinely evaluate the inconsistency in their forecasts.

Is it a cure or blessing? We believe that it is a blessing in that it does not lull us into a false sense of "reliability" and it is better to know and actively approach all possible levels of uncertainty. However, a perfect system would have no issues with unreliability which complicates our decision making and communication framework. If we could honestly choose, we would prefer not to have any inconsistency in our forecast rather than learning to live with it. In that sense it is a curse.

Acknowledgements. This work was partially funded by the FP7 EU Projects KULTURisk (www.kulturisk.eu), GLOWASIS (www.glowasis.eu) and DEWFORA (www.dewfora.net).

Edited by: H. H. G. Savenije

References

Bakhshi, H., Kapetanios, G., and Yates, T.: Rational expectations and fixed event forecasts: An application to UK inflation, Empir. Econ., 30(3), 539–553, 2005.

Bartholmes, J. C., Thielen, J., Ramos, M. H., and Gentilini, S.: The european flood alert system EFAS - Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts, Hydrol. Earth Syst. Sci., 13, 141–153, doi:10.5194/hess-13-141-2009, 2009.

Beven, K. J.: A Manifesto for the Equifinality Thesis, J. Hydrol., 320(1–2), 18–36, 2006.

Bogner, K. and Pappenberger, F.: Multiscale Error Analysis, Correction and Predictive Uncertainty Estimation in a Flood Forecasting System, Water Resour. Res., doi:10.1029/2010WR009137, 2011.

Clements, M. P.: Evaluating the rationality of fixed-event forecasts, J. Forecast., 16, 225–239, 1997.

Clements, M. P. and Taylor, N.: Robustness of fixed-event forecast rationality, J. Forecast., 20(4), 285–295, 2001.

Cloke, H. L. and Pappenberger, F.: Evaluating forecasts of extreme events for hydrological applications: an approach for screening unfamiliar performance measures, Meterol. Appl., 15(1), 181–197, 2008.

Cloke, H. L. and Pappenberger, F.: Ensemble Flood Forecasting: a review, J. Hydrol., 375(3–4), 613–626, 2009.

Cloke, H. L., Thielen, J., Pappenberger, F., Nobert, S., Salamon, P., Buizza, R., Bálint, G., Edlund, C., Koistinen, A., de Saint-Aubin, C., Viel, C., Sprokkereef, E.: Progress in the implementation of Hydrological Ensemble Prediction Systems (HEPS) in Europe for operational flood forecasting, ECMWF Newsletter No. 121, p.20-2, 2009.

Dalcher, A., Kalnay, E., and Hoffman, R. N.: Medium range lagged average forecasts, Mon. Weather Rev., 116, 402–416, 1988.

Davies, A. and Demeritt, D.: Cost-benefit Analysis and the Politics of Valuing the Environment, Radical Stat., 73, 24–33, 2000.

De Roo, A., Thielen, J., Salamon, P., Bogner, K., Nobert, S., Cloke, H. L., Demeritt, D., Younis, J., Kalas, M., Bodis, K., Muraro D., and Pappenberger, F.: Quality Control, Validation and User Feedback of the European Flood Alert System (EFAS), Int. J. Digit. Earth, in press, 2011.

Dedieu, F.: Alerts and catastrophes: The case of the 1999 storm in France, a treacherous risk, Sociologie du Travail, 52(1), 1–21, doi:10.1016/j.soctra.2010.06.001, 2010.

Demeritt, D. and Rothman, D.: Figuring the Costs of Climate Change: An Assessment and Critique, Environ. Plan. A, 31, 389–408, 1999.

Demeritt, D., Cloke, H., Pappenberger, F., Thielen, J., Bartholmes, J., and Ramos, M.-H.: Ensemble Predictions and Perceptions of Risk, Uncertainty, and Error in Flood Forecasting, Environ. Hazards, 7, 115–127, doi:10.1016/j.envhaz.2007.05.001, 2007.

Demeritt, D., Nobert, S., Cloke, H., and Pappenberger, F.: Challenges in communicating and using ensembles in operational flood forecasting, Meteorol. Appl., 17, 209–222, doi:10.1002/met.194, 2010.

Ehret, U.: Convergence Index: A new performance measure for temporal stability of operational rainfall forecast, Meteorol. Z., 19, 441–451, 2010.

Ehret, U.: discussing HESS Opinions "On forecast (in)consistency in a hydro-meteorological chain: curse or blessing?", by Pappenberger, F., Cloke, H. L., Persson, A., and Demeritt, D., Hydrol. Earth Syst. Sci. Discuss., 8, 1225–1245, doi:10.5194/hessd-8-1225-2011, 2011.

Faulkner, H., Parker, D., Green, C., and Beven, K.: Developing a translational discourse to communicate uncertainty in flood risk between science and the practitioner, Ambio, 36(7), 692–703, 2007.

Gupta, H., Sorooshian, S., Gao, X., Imam, B., Hsu, K.-L., Bastidas, L., Li, J., and Mahani, S.: The challenge of predicting flash floods from thunderstorm rainfall, Philos. T. Roy. Soc. Lond. A, 360, 1363–1371, 2002.

Hamill, T. M.: Evaluating forecasters' rules of thumb: a study of D(Prog)/Dt, Weather Forecast., 18, 933–937, 2003.

Hoffman, R. N. and Kalnay, E.: Lagged average forecasting, an alternative to monte-carlo forecasting, Tellus A, 35, 100–118, 1983.

Laio, F. and Tamea, S.: Verification tools for probabilistic forecasts of continuous hydrological variables, Hydrol. Earth Syst. Sci., 11, 1267–1277, doi:10.5194/hess-11-1267-2007, 2007.

Lashley, S. L., Fisher, L., Simpson, B. J. , Taylor, J., Weisser, S., Logsdon, J. A., and Lammers, A. M.: Observing verification trends and applying a methodology to probabilistic precipitation forecasts at a National Weather Service Forecast Office. Preprints, 19th Conf. on Probability and Statistics, New Orleans, LA, Am. Meteorol. Soc., 9.4. available at: http://ams.confex.com/ams/pdfpapers/134204.pdf (last access: 18 July 2011), 2008.

MacKenzie, D.: Inventing Accuracy: An Historical Sociology of Nuclear Missile Guidance, MIT Press, Cambridge, MA, 1990.

Mills, T. C. and Pepper, G. T.: Assessing the forecasters: an analysis of the forecast records of the treasury, the London Business School and the National Institute, Int. J. Forecast., 15, 247–257, 1999.

Montanari, A.: What do we mean by "uncertainty"? The need for a consistent wording about uncertainty assessment in hydrology, Hydrol. Process., 21, 841–845, doi:10.1002/hyp.6623, 2007.

Murphy, A. H.: The value of climatological, categorical and probabilistic forecastsing the cost-loss ratio situation, Mon. Weather

Rev., 105, 803–816, 1977.
Nichols, N.: Cognitive illusions, heuristics and climate prediction, B. Am. Meteorol. Soc., 7, 1385–1397, 1999.
Nobert, S. And Demeritt, D.: Models of "good" risk communication for flooding and other water related hazards: a critical review, KULTURisk WP5 report, available at: www.kulturisk.eu (last access: 18 July 2011), 2011.
Nobert, S., Demeritt, D., and Cloke, H. L.: Using Ensemble Predictions for operational flood forecasting: Lessons from Sweden, J. Flood Risk Manage., 3, 72–79, 2010.
Nordhaus, W. D.: forecast efficiency: concepts and applications, Rev. Econ. Stat. 69, 667–674, 1987.
NRC – National Research Council: Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts, National Academy Press, Washington, DC, 2006.
Palmer, T. N. and Tibaldi, S.: On the prediction of forecast skill, Mon. Weather Rev., 116, 2453–2480, 1988.
Pappenberger, F., Bartholmes, J., Thielen, J., Cloke, H. L., Buizza, R., and de Roo, A.: New dimensions in early flood warning across the globe using grand-ensemble weather predictions, Geophys. Res. Lett., 35, L10404, doi:10.1029/2008GL033837, 2008.
Pappenberger, F., Thielen, J., and del Medico, M.: The impact of weather forecast improvements on large scale hydrology: analysing a decade of forecasts of the European Flood Alert System, Hydrol. Process., 25(7), doi:10.1002/hyp.7772, 2011a.
Pappenberger, F., Bogner, K., Wetterhall, F., He, Y., Cloke, H. L., and Thielen, J.: Forecast convergence score: a forecaster's approach to analysing hydro-meteorological forecast systems, Adv. Geosci., 29, 27–32, doi:10.5194/adgeo-29-27-2011, 2011b.
Persson, A. and Grazzini, F.: User Guide to ECMWF forecast products, available at: http://www.ecmwf.int/products/forecasts/guide/index.html, last access: 18.07.2011, 2007.
Persson, A.: Update of User Guide to ECMWF forecast products, available at: http://www.ecmwf.int/products/forecasts/guide (last access: 18 July 2011), 2011.
Ramos, M. H., Mathevet, T., Thielen, J., and Pappenberger, F.: Communicating uncertainty in hydro-meteorological forecasts: mission impossible?, Meteorol. Appl., 17(2), 223–235, 2010.
Richardson, D. S.: Skill and economic value of the ECMWF ensemble prediction system, Q. J. Roy. Meteorol. Soc., 126, 649–668, 2000.
Roebber, P. J.: Variability in successive operational model forecasts of maritime cyclogenesis, Weather Forecast., 5, 586–595, 1990.
Roulin, E.: Skill and relative economic value of medium-range hydrological ensemble predictions, Hydrol. Earth Syst. Sci., 11, 725–737, doi:10.5194/hess-11-725-2007, 2007.
Ruth, D. P., Glahn, B., Dagostaro, V., and Gilbert, K.: The Performance of MOS in the Digital Age, Weather Forecast., 24(2), 504–519, 2009.
Shackley, S. and Wynne, B.: Integrating Knowledges for climate change: pyramids, nets and uncertainties, Global Environ, Change, 5, 113–126, 1995.
Simmons, A. J., Mureau, R., and Petroliagis, T.: Error growth and predictability estimates for the ECMWF forecasting system, Q. J. Roy. Meteorol. Soc., 121, 1739–1771, 1995.
Slovic, P.: Perception of risk, Science, 236, 280–285, 1987.
Thielen, J., Bartholmes, J., Ramos, M.-H., and de Roo, A.: The European Flood Alert System - Part 1: Concept and development, Hydrol. Earth Syst. Sci., 13, 125–140, doi:10.5194/hess-13-125-2009, 2009a.
Thielen, J., Bogner, K., Pappenberger F., Kalas, M., del Medico, M., and de Roo, A.: Monthly-, medium- and short range flood warning: testing the limits of predictability, Meteorol. Appl., 16(1), 77–90, 2009b.
Zsoter, E., Buizza, R., and Richardson, D.: "Jumpiness" of ECMWF and Met Office EPS Control and Ensemble-Mean Forecast, Mon. Weather Rev., 3823–3826, 2009.