Response to Referee Comments by Dr. Nathalie Voisin

The authors would like to thank Dr Voisin for taking the time to review our manuscript and for her constructive comments.

Response to Specific Comments:

The referee made three Specific Comments, which we will summarise and to which we will respond.

1. *In the analysis, the authors assume that the mitigation time is linked to the forecast lead time, and then base the analysis on the lead time instead.*
The referee notes that in the analysis, we have chosen to estimate potential loss analysis as a function of lead time instead of mitigation time. Indeed, in reality there is a difference between the two (as mentioned on p6644, lines 14-16 and in Fig. 1) due to the time required to prepare and disseminate a forecast and to take and communicate a decision. Essentially, we assumed this required time to be of zero length. This is done for reasons of simplicity, and to thus avoid confusion. We think this is justified as it improves readability and affects all cases in similar fashion. We will stress this in the text of the manuscript.

The value La was not made a function of time. I would think that if the mitigation time is less than 1 hour, La is 0.
In the manuscript, avoidable losses (La) *are* a function of lead time. This has been explained in p6653, lines 4-5 and shown in Fig. 4. However, the caption to this figure does not explain this very well; we shall correct this.

I would think that if the mitigation time is less than 1 hour, La is 0.
Correct. Indeed, the analysis starts with a lead time (assumed to be equal to mitigation time) of 1 hour, not less.

I refer to the authors to explain how the current forecasts are made, and how current warnings are issued in order to argument an analysis made on lead time 1 to 6 hours.
We agree with the referee that lead times used in the analysis are quite short. These lead times, however, are based on the situation in the White Cart basin. The time of concentration of White Cart at Overlee (upstream area approximately one hundred square kilometres) is approximately three hours. It is a statutory requirement to the Scottish flood forecasting and warning authorities (SEPA) to issue flood warnings no less than three hours in advance (Werner and Cranston, 2009). The operational forecasting system includes one source of forecast precip only – radar now/forecasts – which has a maximum lead time of six hours. These are the main reasons for choosing similar time scales in our analysis. While this does not allow the at risk community to take extensive mitigating action, some actions can – and indeed are – taken. The type of actions (as a function of lead times on similar time scales, i.e. several hours) are described by Carsell et al. (2004). We shall add these reasons (for using 1 through 6 hour lead times) to the manuscript.

2. Adding a discussion on the relative value based on the time of concentration in the

text and conclusion. ... The reliability of the probabilistic forecasts varies significantly for lead time shorter than the time of concentration and those longer than the time of concentration.
For the re-forecast analysis, we used precip *observations* rather than precip *forecasts* to force the hydrologic model. This has mainly been done for practical reasons (availability of forcing data). We think this is justified as while the resulting levels of predictive hydrological uncertainty are different, this has no effect on the actual analysis: demonstrating the difference between the four cases. Indeed, had we used a mix of observed and forecast precipitation (which we didn't), a jump in sharpness of the probability forecasts would have been likely. We will add a note to the discussion section of the manuscript to stress that observations rather than forecasts were used, and how this affects the results of the case study.

But a deterministic forecast is still required in order to gain on computation time.
This would indeed be the case if predictive uncertainty would be estimating by forward propagation of forcing uncertainty using ensembles. In the present case study, however, predictive uncertainty is estimated by a post-processor that uses a single predictor: the deterministic forecast. This post-processor does not require a lot of computation time (less than five seconds), making it attractive for use in near real-time forecasting systems.

3. *How to make sure that the deterministic threshold, which is used for both deterministic and probabilistic forecasts, is optimized with the probabilistic thresholds? As end users will determine new rules with probabilistic forecasting, they might use a lower/higher deterministic threshold with an optimized associated probabilistic threshold with it.*
The 1.5m threshold used in the analysis is taken from the operational forecasting system in use in White Cart. It was based on a flood risk analysis carried out some years ago and coincides with the water level above which damage starts to occur. We shall add a note to this effect in the manuscript.

Response to Technical Revisions:

"Surprises" versus "missed events".
The use of two terms to indicate the same is confusing. We will make sure to consistently use "missed event".

p6642, line 1: define EAD
We shall add the meaning of the abbreviation EAD after its first use.

p6642, line 13: situationS
we shall change the manuscript accordingly

p6643, line 6: in detailS.
"Detail" here is part of the expression "to explain in detail" which uses the singular "detail".

P6644, line 14: increase in mitigation time decrease potential losses but also increase the cost of an alarm: people not working, staff on hold, business closed, etc. Optimized, but not maximized.
The "maximised" refers to the first part of the sentence, which discusses the benefits of mitigating action only, and not also the costs. To avoid confusion, we shall incorporate a change to clarify that

maximisation of mitigation time does not necessarily mean an optimisation of benefits.

P6645, line 2: La = 0.5 Lp Â˝ U arbitrary . How sensitive are the results to this assumption?
This value of La was based on research carried out by DEFRA (citation in p6644, line 30) and is therefore not as arbitrary as the last sentence of the paragraph may be suggesting. We will change the wording slightly to avoid confusion.
The case study results are quite sensitive to the value of La. This, however, only affects the absolute values of mitigated loss for the perfect / single value / probability forecast cases with respect to the no forecast case. It does not affect the performance of the perfect / single value / probability forecast cases versus one another. A note to this effect is included in the Discussion section (p6662, second paragraph), but this note shall be extended to also state *how* choice of La affects performance, rather than only that it does.

why use climatology and not persistence? Climatology is used for seasonal forecast. Persistence for short term flow forecasts.?
By "climatology" / "climatological frequency", we mean the long-term ("climatic") average frequency of event occurrence. The use of this frequency as a forecast is often used as a reference, unskilled forecast. In that sense it is different from persistence, which in short-term forecasting is used to indicate forecast runs of several days with similar characteristics. We shall add a note to clarify this.

Pp6648: the paper would benefit of including in table 1 the different cost associated with each frequency.
We agree, and will change the manuscript accordingly.

P6651: Please clarify which flood level was chosen for the analysis, i.e. to which quantile the flood level correspond; what is the reference quantile corresponding to 1.5 m; is this the quantile threshold or which one do you use exactly?
The 1.5m flood level was derived in a flood risk analysis for the White Cart basin, and coincides with both the level above which flood damage begins to occur and one of the warning thresholds at the flood warning location. We have not related this to a flood frequency analysis, but the available data record (that was used for the analysis) shows that, on average, this level is exceeded more than once annually. We shall add a note to the text clarifying the significance of the 1.5m flooding threshold.

Pp6651 Â˝ U line 19-20; The quantiles were derived over the period used in the analysis? Is this a dependent verification then? Please clarify how it affects your results Â˝ U i.e. both deterministic and probabilistic the same way?
The quantiles were derived using a re-forecast analysis spanning a 5-year period starting April 1, 1991 ending March 31, 1996 (p6651, line 18). Re-forecast analyses (both deterministic and probabilistic) for value assessment were carried out for the approx. 11 year period starting April 1, 1996 ending February 20, 2007 (p6654, line 5 and p6655, line 25). This thus consitutes an independent verification, as the quality of quantiles was assessed on a different period from the period that was used for their derivation. We shall add a comment to clarify.

Pp6656, line 3 Â˝ U unfinished sentence (rule of zero Â ˇ E)

The last few sentences of this paragraph should read:
"From these pairs of forecasts and observations, for every decision rule, the number of resulting hits, misses, false alarms and quiets was determined. Table 4 shows these numbers for forecasts with a 3-h leadtime. For the decision rule 'warn if forecasted event probability is equal to or higher than 0 per cent' – i.e. always issue a warning – the number of hits is equal to the number of observed events (h = o = 15) with the number of false alarms being equal to the number of forecasts made, minus the number of hits (f = N – h = 15857). At the other extreme, the decision rule 'warn if forecasted event probability equals 100 per cent' results in zero hits, zero false alarms and all events missed."

There was also a mistake made in Table 4. The first row should read: threshold = 0, h = 15, m = 0, f = 15857, q = 0. This shall be corrected.

Pp6657: clarify that the envelop is giving the best improvement from probabilistic fore- cast, but it also means that a different probabilistic threshold needs to be optimized for different cost-loss ratio.
It is assumed that users apply the "optimal decision rule" (p6656, lines 10-12) which means that each user issues a warning if and only if the predictive probability of event occurrence equals or exceeds his/her cost-loss ratio, thus optimising his/her expected expense or utility (Murphy, 1985). This indeed means that users with differing cost-loss ratios should apply different decision rules to fully benefit from probability forecast. If they use a decision rule that does not coincide with their cost-to-loss ration, their benefit will be less. The envelope shows the benefit that is realised using the optimal warning rule. We shall add this clarification, and an explanation of how this rule is derived, to the manuscript.

P6658, line 25-28: as mentioned above, for different lead time and different cost ratio, a different probabilistic forecast thresholds needs to be optimized in order to claim the benefits upon a deterministic forecasts.
This is correct. We refer to above clarification.

Conclusions: I would suggest keeping it to the main points only.
We shall list the main points of the article in reduced form.

P6664: line 17-19: I would say that the improvement is noted throughout different lead times and cost lost ratio, instead of saying that it is independent. Also, this is given that the probabilistic threshold is optimized. This is important to clarify so that future users understand the steps to ensure a benefit from using probabilistic forecasting.
We shall amend the text accordingly and add the suggested clarification.

Table 1: add to cost associated with each frequency
We will change the manuscript accordingly.

Table 2: add relative improvement. Specify that lead time is not mitigation time.
We shall add a column listing the EAD values for each of the 'perfect forecasts' as a percentage of the EAD in the 'no warning' case. We will change the words 'lead time' in 'mitigation time'.

Figure 2: please add arrows specifying in which direction all variables increase.
We will change the manuscript accordingly.

Figure 6: add the line V=0 in the figure. Specify the range in cost lost ratio for which warning benefit is negative.
We will change the manuscript to show the V=0 line. We shall add a note to the caption explaining when warning benefit is less than zero.

Figure 8: specify each probabilistic threshold for the different lines, or at least p=0, p=1. Add line V=0.
We shall amend the graph so that it becomes clear which line is associated with which probability threshold. We shall also emphasize the V = 0 line.

Figure 9: specify p=0 and p=1 lines.
We shall amend the graph so that it becomes clear which line is associated with which probability threshold.

Figure 10: please add longer lead timesÂ ˇ E
Please refer to the response to the first comment. We have chosen to, as per the Referee's suggestion, how current warnings are issued in order to argument an analysis made on lead time 1 to 6 hours.

Delft, August 29, 2011
Jan Verkade and Micha Werner