Object : Submission of a revised version of the manuscript (hess-2012-434) by L. Alfieri et al.

We thank Dr. Hopson (Reviewer 3) for the useful comments and the constructive discussion he raised within the review of the article. We have carefully considered the reviewer's comments and worked to include in the revised version of the manuscript the proposed suggestions.
Please find below our point by point responses to the reviewers' comments. The original comments from the reviewer have been reproduced below. They are interspersed with our responses.

specific comments
The system's bias-correction method of ERA-Interim to GPCP needs to be discussed further; in particular, the issue of spatial and temporal scale-mismatch and the accounting of rain/no-rain frequency mismatch (especially important for dry regions with pronounced nonlinear catchment response).

Reply: We added more details on the ERA-Interim bias correction using GPCP (section 2.1.2). In brief, this is based on large-scale (2.5°x2.5°) rescaling of the monthly means of ERA-Interim. There was no rain/no-rain correction, since this would require the use of a different observational dataset. The reviewer made a very good point, but we believe that it is out of the scope of the current study. The current system is based on ECMWF products "as they are", and future development will of course focus in improving the different components, including the reference climatology simulation. Furthermore, Balsamo et al. 2012 also found that using the surface and sub-surface runoff from ERA-Interim/Land (as we use in this study) improved the monthly mean discharge in several large-scale basins in the world, when compared to using the original ERA-Interim fields.

It appears there are no temperature forecast (among other variables) corrections applied to the varEPS (at least as compared to ERA-Interim). The significance of this on the accuracy of snowmelt-derived discharge (in particular) should be discussed.

Reply: Snowmelt driven runoff is generated directly by varEPS, where HTESSEL if fully coupled (online) with the atmosphere. Such temperature correction would require re-running HTESSEL offline forced with bias corrected varEPS fields. This is a possibility for future developments of the system, which could also include other types of bias corrections to the forecasts.

The authors do discuss very briefly an "initial condition" bias adjustment for the early warning system, but the discussion on this needs to be expanded, along with potential sources of error in this approach (even though some errors are less important given the authors are concerning themselves primarily with forecasting "model-space" relative frequencies).

Reply: An additional part has been added in Sect. 3.2 to stress the importance of the quality and the uncertainty of the initial conditions to the forecast skill. Two relevant references have also been added, which give more insight and provide the reader with thorough discussions on the topic.

The interface between HTESSEL and Lisflood needs further discussion. In particular, given that HTESSEL itself produces surface and sub-surface runoff through its four layer subsurface model, clarify what missing features (besides overland flow and river routing) require the inclusion of the additional Lisflood 2-layer subsurface model.

Reply: In the context of global flood modeling, the transformation from precipitation to surface and sub-surface runoff is done by the HTESSEL module of the IFS, which accounts for vertical water fluxes and water/snow storage on a pixel basis. However, HTESSEL is not capable of simulating horizontal water fluxes along the river network. To this purpose, Lisflood global is set up to simulate the groundwater and routing processes, using surface runoff and sub-surface runoff (soil to groundwater percolation) from HTESSEL as input fluxes on a resolution of 0.1°. Additional details have been included in the description of HTESSEL and Lisflood.

A discussion needs to be included on calibration and parameter estimation for HTESSEL and Lisflood land-surface and sub-surface schemes, along with the Lisflood routing scheme.

Reply: We have added to the text that river parameters like channel gradient, Manning's coefficient, river length, width and depth were estimated by taking the digital elevation model, the river network and the upstream area into account. Further details of the Lisflood model can be found in van de Knijff et al. (2010). Within EFAS, the parameters to control percolation to the lower ground water zone, the residence time of the upper and lower zone and the routing parameters (a multiplier to Manning's roughness) are some of the parameters used for calibration (see Feyen et al 2007). In this setup of GloFAS these parameters are set to literature values. Parameter estimation through calibration is subject of future works.

How the 45 day forecasts are produced needs further discussion; i.e. what assumptions are used (persistence? Climatology?) to extend beyond the 15-day varEPS forecast horizon?

Reply: From day 16 to day 45, input maps of surface and subsurface runoff are set to zero, therefore the hydrological model (i.e., Lisflood) will simply convey towards the outlet water already within each river basin. This was clarified in Sect. 2.3.

Are individual discharge ensemble members used in any way in the warning system or in the forecasting scheme evaluation (besides just in generating the ensemble mean forecast)?

Reply: Yes, they are used to compute the probability of the ESP of exceeding the three warning thresholds. An example of persistence diagram has been added in Figure 11, showing the probability [%] of exceeding the high and severe alert level over five consecutive forecasts issued from 24th to 28th July 2010 (for the forecasts shown in Fig. 10). Some description was added in Sect. 4.3.

In section 3.2, I presume all CRPSS and AROC values apply to daily values of the forecasts; or are forecasts averaged (or exceeding a threshold) over a window of time, e.g. a 5-day forecast window at lead-times 41- to 45-days?

Reply: Yes, as stated in Sect. 4.2, "CRPSS maps for the 2 years of ensemble streamflow prediction (i.e., 2009-2010) were calculated for each selected forecast lead time from 1 to 45 days". Indeed, in Sect. 3.2 a general description of the skill scores is presented, while technical details on how the scores were applied in the analysis are shown in Sect. 4.2. In the same section, we modified the text to clarify that also AROC values refer to daily values with fixed forecast lead time. The text now reads "The threshold exceedance analysis is evaluated through the use of ROC curves and specifically the area under these curves, which was calculated for each of the 45 daily forecast lead times".

technical corrections
Note: “P” refers to the page number; “L” refers to the line number within the page
P12299 L6 “. . . given in the following sections”

Reply: Amended.

P12300 L10 – how does surface runoff account for subgrid variability of orography? Explain.

Reply: The overland flow to the pixel outlet does only take land cover into account. The river network routing takes the channel gradient derived from Hydrosheds and therefore some subgrid variability is taken into account. Some additions were included in the text for clarification.

P12301 does Lisflood use a separate grid resolution than HTESSEL?

Reply: HTESSEL has the same resolution of the meteorological model, which is about 80 km for the reference climatology (ERA-Interim) and 32 to 65 km for VarEPS operational forecasts (see Sect. 2.1.1 and 2.1.2). In Sect. 2.2.1 we clarified that for this work, operational ensemble forecasts of surface and sub-surface runoff are extracted from the daily output of the ECMWF forecasts and then resampled to 0.1° resolution to be used as input by Lisflood.

P12301 L6-8 – is the surface runoff routed to the outlet of each cell via overland flow, or assumed to be in-channel flows, and thus, what do the surface roughness coefficients apply to?

Reply: Surface runoff is routed via overland flow, therefore no channel parameter is assumed for the routing. The surface roughness is based on land cover classes as specified in Sect. 2.2.2

P12301 L16 – discuss the limitations of having all subsurface flows routed to the outlet in one time step.

Reply: The sentence has been reformulated as it was not completely clear. Indeed, the amount and timing of the outflow from the respective groundwater reservoirs to the outlet of each grid cell are controlled by two parameters which reflect the residence time of water in the upper and lower groundwater zone.

P12301 L16-18 – is the kinematic wave approach discussed here a repeat of what is discussed in L8 above?

Reply: Yes, it is the same approach though applied with the corresponding parameters for the channel. The text has been modified accordingly. River parameters like channel gradient, Manning’s coefficient, river length, width and depth were estimated by taking the digital elevation model, the river network and the upstream area into account.

P12302 L23-26 – is 2-yr return period based on daily-averaged discharge values? This section is not entirely clear.

Reply: Yes, both the warning thresholds and ensemble forecasts are based on daily values. This has been clarified in Sect. 2.2.

P12303 L3-5 – how do you define the “medium warning threshold”? Does this imply that you only issue warnings if the reach is already in a “medium” flood stage?

Reply: This part is just related to the selection of points to be displayed on the website. Warning thresholds are defined according to ranges of return periods, as described in the last paragraph of Sect. 2.3. The rules to issue flood warnings have not been discussed in this paper, as these are usually defined after analysing a number of flood events and the corresponding forecast probabilities. For example, in EFAS this corresponds to a probability of 30% to 60% (depending on the forecast product) of exceeding the 5-year return period threshold.

P12307 L27 to P12308 L5 – how the quantile matching is performed on the initial conditions is not clear; are you applying this to the soil moisture states, rainfall fields, etc.?

Reply: We have included more details and discussion to this part. Also, we have clarified in the text that the correction is used only for validation purpose, while model state variables are not affected by it.