Response to Anonymous Referee #1

I thank the reviewer for these comments. They improve the clarity of the manuscript and bring up a number of interesting points. The manuscript has been changed in the following ways in response to the reviewer's concerns (in bold)

"The main concern I have with the presented paper is the lack of sufficient details regarding the way forecasts are produced. On page 14436 the author mentions "…extended use of rainfall forecasts, and improved flood forecast model". Later on "…use of satellite-based precipitation estimates to supplement the sparse ground-based rain gauge network". How meteorological data are chosen? Which sources are used? Is rainfall/precipitation the only meteorological input? I'd like to see a more specific description on the input data and the subsequent processing to obtain discharge/river stage estimates. This would help a lot the understanding of the forecasting system and how decisions are taken."

Some of this information was contained in other sections, but the requested details have been collected into a new section labelled "forecast methods". The section also contains more detail. The entirety of that section is reproduced here:

The RFMMC relies on observed river height data as well as precipitation estimates as inputs for models and to develop situational awareness. Ground-based stations are primarily selected based on their realtime availability. In recent years, the RFMMC has expanded its use of satellite-based precipitation estimates to supplement the sparse ground-based rain gauge network. The RFMMC uses two satellite-based products from the National Oceanic and Atmospheric Administration - Satellite Rainfall Estimation and the Tropical Rainfall Measuring Mission (MRC, 2010). The RFMMC has developed statistical methods for removing bias from the satellite-based products.

The RFMMC inherited several forecasting tools, including the Streamflow Synthesis and Reservoir Regulation (SSARR, Rockwood, 1968) installed in 1967 to simulate flows in the main river from Chiang Saen to Pakse (Johnston and Kummu, 2012). Following the recommendations of a comprehensive review (Malone, 2006) the forecasting system was updated in 2008 to use additional data sources, improve and extend use of rainfall forecasts and adopt improved hydrologic models.

The RFMMC currently uses human expertise and a combination of statistical, hydrologic and hydraulic models to generate flood forecasts. Empirical methods such as statistical regression are used downstream of Pakse, for example, estimating the recent rate of change of river height at the upstream river station and regressing this against the downstream station height change to make a future forecast. The statistical model output serves as a "sanity check" for the other model outputs, but is also useful when a lack of rainfall observations prohibit the running of other models.

In 2008, the RFMMC shifted to the Delft-FEWS platform using the URBS event-based hydrologic model with Muskingum hydraulic routing (Tospornsampan et al., 2009). URBS can be forced with spatially semi-distributed station and/or satellite based rainfall. Manually-tuned loss parameters control the rates of rainfall excess. The routing model is then forced with the rainfall excess and the observed recent streamflow. MM5 (Fifth Generation Mesoscale Model operated by the US Air Force, Cox et al., 1998) gives three, 24-hourly forecasts of rainfall for consecutive days and zero rainfall is assumed subsequently (Malone, 2006).

The RFMMC also uses the ISIS hydrodynamic model, a generic one-dimensional model for the simulation of unsteady flow in channel networks, by providing an implicit numerical solver for the Saint Venant equations. At selected intervals, it computes water levels and discharges on a non-staggered grid. The ISIS model is used for forecasts from Stung Treng to the ocean, receiving tributary inflows from the URBS model. ISIS is more computationally intensive than URBS and therefore the latter is run routinely whereas ISIS is run for retrospective analyses and as demand arises.

Over time, the operational forecasters have improved and gained experience with the system. The system was tested by major floods in 2008 and 2011, after which the forecasters re-tuned the URBS model parameters. Hydrologists use their situational awareness to quality control data, adjust model parameters/outputs and synthesize the results before generating the official forecasts.

" Also, the author states, that (p14439) "Total travel time between Chiang Saen and Phnom Penh is about 10 days". That means that the skill of rainfall forecasts might be not as important as that of a good rainfall estimation approach and a good routing model, considering that 5 day is the longest forecast lead time chosen. Also, correlation techniques between stations might be useful. I suggest the author to comment on this."

The other reviewer had similar concerns, that discussion should be given to the relative importance of hydrology versus hydraulics. The tool section mentioned above includes correlation methods. Beginning line 403, the following text has been added as well

Despite the large range of error standard deviations from one location to another, the CP indicates that the skill of forecasts is relatively even across the basin. There is a larger difference in 1- and 5-day ahead CP for the upstream locations than there is for the downstream locations between Kratie and Neak Luong, which may be the attributed to the greater uncertainties in initial conditions, recent and future precipitation and other meteorological influences at the smaller scale watersheds found upstream. Indeed, the lowest performing forecasts (5-days ahead at Chiang Saen) rely almost exclusively on the signal contained in observed upstream flows due to the lack of access to rainfall observations in China. Downstream, where hydraulic routing effects have a greater influence than local precipitation, there is nearly no loss of skill with leadtime. The exception is the two furthest downstream forecast points, where low flow forecasts have relatively high error when the river height is affected by the ocean (e.g. observe the poor performance of Tan Chau forecasts in June-July, relative to those in September-October in Figure 2).

"At page 14447, the error standard deviations are difficult to evaluate as they are now, because they depend a lot on the shape of the riverbed and consequently on the typical ranges of values. I suggest showing them together with the standard deviations of observations (or a ratio between the two values), perhaps in a Table."

The reviewer had a good insight about the standard deviation being related to the riverbed shape. This text has been added to section 7 and the observed standard deviation has been added to figure 3 and table 2.

Most locations upstream of Phnom Penh have a wet-season observed standard deviation near 2.5 meters although Kratie has a value as high as 3.6 and Chiang Saen (the most upstream point) is as low as 1.4 meters. The river height at Kratie is naturally more variable than neighboring locations because of Kratie's W-shaped channel cross section and nearly vertical 15-meter tall banks. Below Phnom Penh, the observed standard deviation is typically close to 1.5 meters. Some of the observed variability is due to the seasonal cycle. The standard deviation of August observations (near the peak of the wet season) is also shown at the top of Figure 3.

"Specific comments p14435, l1: "underdeveloped" does not read very well. I'd suggest removing it or finding a politically correct alternative. ,l 16: "respectively" is not needed."

"underdeveloped" was replaced by "less developed". The word respectively was used to distinguish structural versus non-structural measures and so the sentence has been restructured like so

The RFMMC and the flood forecasts it produces are part of a broader water management plan that includes both structural measures designed to keep floods away from people and non-structural measures designed to keep people away from floods.

p 14437, l2: "and" should be "is". ,l 20: ": : :"? Please amend. ,l 9-26: I would put a reference to Fig 1 to facilitate the understanding of the text.

Accepted as suggested

p 14437-38: Please make uniform the way to cite MRC (2005) (later on cited as Mekong River Commission, 2005)

Accepted as suggested

p 14438, l9: "(e.g. 11.8 m)". I suggest specifying where (e.g., at Pakse). , l 23: From Fig 2 it looks July to October. Please clarify.

Accepted as suggested. The reviewer was correct. There was a mistake in the figure and it has been corrected.

p 14439, l4: provide a reference for this. , l9: "is fair" should be made more specific ,l 20: "and they are" should be "as they are" (the spreadsheets) or "and is" (the layout).

I calculated the travel time myself and my analysis largely agreed with numbers provided in an email from the MRC. However, I haven't seen such analysis published in a journal so I added a reference of a personal communication with the Mekong River Commission. The final numbers in the article were from MRC, not my analysis.

The word "fair" came from an article describing the network and the original did not provide more detail. However the word has been changed from "fair" to "sufficient" (a phrase used in Malone's report). Some extra quantitative information was included and so the text has been changed to

Rain gauge density (but not spatial distribution) in Thailand and Viet Nam is sufficient, but the networks are inadequate in Cambodia and Laos (Pengel et al., 2008). There is little automation and telemetry of measurements, in part because human observers remain relatively inexpensive and provide reliable quality data. In 2006, the RFMMC had realtime access to 20 rainfall stations across 250,000 km^2 between Chiang Saen and Pakse. This is less than one tenth the density recommended by the World Meteorological Organization (Malone, 2006).

The text was changed to

The layout of the spreadsheets has changed over time and is designed to be human-readable (as opposed to having a strict and consistent format for machine-readability).

p 14441, l1: 1) Bulletins, 2) Operational Database, 3) IKMP

Because the numbering does not say which is the highest priority the text was changed to

The data were merged in order of priority (lowest to highest): Bulletins, Operational Database, IKMP.

p 14442, l 1-4: Please reshape the sentence, now difficult to read (particularly the part in brackets). , l 19: "high" should be "highest" or similar.

This text has been changed to

Plate et al. (2008) demonstrated general evaluation concepts using water level forecasts from the SSARR model during July – October 2005 (wet season) as examples. The study included standard performance measures such as the Nash-Sutcliffe (NS, Nash and Sutcliffe, 1970). The NS is the mean squared error of the forecasts, relative to the error if the long-term average water level were used in place of forecasts (1 is perfect, 0 is no-skill).

p 14445, l14: This seems the same as the quality score (Plate, 2008) described at page 14442. Can the author clarify this point? ,l 20: Not the best way to describe it mathematically.

The text right after the above was expanded to

Plate et al. presented a "Quality Index", which is similar to NS but uses persistence instead of long-term average water level as a baseline and has a reverse orientation (i.e. 0 is perfect, 1 is no-skill). The formula for this index is the same as the Coefficient of Prediction (CP, described in the next section) except the orientation is reversed. This is a more difficult baseline to outperform and Quality scores at Pakse were 0.47 for 1 day ahead degrading to 0.74 for 5 days ahead (CP of 0.53 and 0.26, respectively) .

The formula was changed to the correct mathematical syntax for an if-then-else statement.

$$\mathrm{PS}(\mathrm{loc},\mathrm{lead}) = \frac{1}{N}\sum_{i=1}^{N} \begin{matrix} |f_i(\mathrm{loc},\mathrm{lead}) - \mathrm{o}_{i+lead}(\mathrm{loc})| < \mathrm{B}(\mathrm{loc},\mathrm{lead}) \rightarrow 1 \\ |f_i(\mathrm{loc},\mathrm{lead}) - \mathrm{o}_{i+lead}(\mathrm{loc})| \geq \mathrm{B}(\mathrm{loc},\mathrm{lead}) \rightarrow 0 \end{matrix}$$

p 14448, l17-19: Are the new benchmarks derived over all available years of forecasts?

Yes. Clarified as suggested

p 14451, l1-5: This part doesn't read very well. I suggest clarifying it and make it more specific.

The text has been changed to

The forecasts should be visualized in the context of the recent observations and historical climatology to ensure that unreasonable forecasts are not issued. For example, the recent observation can be extended into an envelope of possibilities in the future based on simple autocorrelation of historical river levels at a given location (e.g. the river depth has rarely changed more than 1 meter per day); the operational forecast can go outside this envelope if anomalous conditions are predicted (e.g. significant rainfall has occurred and/or a flood wave has been observed upstream).

Table 3: Unusual layout. I suggest showing the POD and ETS as additional columns on the right of the FAR.

The original layout was chosen so the scores would be next to the data used to calculate them, but I have accepted the reviewer suggestion to put them in an extra column.

In Figure 2, circles corresponding to 1 to 5 day forecasts are unreadable. I'd choose 1 lead time or use a 2-column layout with 1 and 5 day lead time.

The figure has been changed to include the 1 and 5 day head forecasts only. Also the error is now plotted, highlighting the differences between the two.