Response to Reviewers' comments

Reviewer # 2

1. ***Comment:*** *The ESP approach : in order to not run an ensemble, it seems that the authors used the ensemble mean for the precipitation forcing of VIC. It is not realistic and VICclim certainly has less skill than an ensemble mean flow forecasts.*

   Response: Actually, climatological forcings are specified as an ensemble. Only for ECHAM4.5 forecasts, we used the ensemble mean, since we required spatio-temporal downscaling. The sentences have been revised to reflect that daily precipitation forcings (at 1/8° resolution from Maurer et al., 2002) during the climatological period from 1957 to 1980 were used to drive the VIC model with each updated IHC for the forecasting period (1981 to 2010). For instance, to predict 6-month lead streamflow starting from January 1981, IHCs at the end of December 31, 1980 were used and the VIC model was implemented with daily forcings (1/8° resolution) from January to June period during 1957 to 1980 (all the 24 years). Finally, the ensemble mean flow forecasts were estimated by averaging estimated streamflow during 24 6-month periods at lead time 1-6 months.

2. ***Comment:*** *A bias correction is applied to the flow forecasts with no details on the approach. A bias correction will affect the analysis and needs to be presented along with an expectation on how this could affect the results.*

   Response: We have included the following details in the revised manuscript. Percentage bias correction on the mean monthly simulated flow during the calibration period (1957 to 1980) was estimated and was applied on the mean flow simulated during the evaluation period (1981 to 2010) for each month. For consistency, calibration period is updated to 1957 to 1980 instead of 1951 to 1980. The estimated values of percentage bias correction during the calibration period are provided in Table 1.

3. ***Comment:*** *With the flow observation as reference and the current calibration (overestimation of Spring flow), it means that any negatively biased seasonal precipitation will show as improvement. It would be good to add a discussion with respect to the VIC simulation forced with the observed gridded meteorological dataset as reference instead, then compare with the observed flow in the discussion section. This would perhaps also allow supporting some of the conclusions regarding VIC simulations in low flow conditions.*

   Response: We have added a comparison with respect to VIC simulated streamflow (using the observed gridded data) as a reference.

   We choose two measures of skill: Rank correlation (which tells us how well the variability in simulated streamflow is captured, but not magnitude) and the Mean Square Skill Score (which tells us about forecast accuracy comparing the mean square error of the forecast with respect to the mean square error of climatology). During spring months, when there is overestimation of simulated flows, negative bias in precipitation may result in lower

streamflow but it does not mean that the variability will also be simulated well (as determined by the rank correlation). Finally, we perform systematic bias correction on simulated streamflows (based on VIC model performance during the calibration period) so that we can compare the utility of precipitation forecasts in forecasting the observed flows.

4. ***Comment:*** *coordinate the period of calibration of the different parameters and the verification period. There are all sometimes independent and sometimes overlapping. This can drive to overfitting for some experiments and affect the inter-comparison of the different forecast approaches.*

   Response: The calibration and evaluation period of both the statistical and VIC models are independent. For consistency, calibration (training) period is updated to 1957 to 1980 instead of 1951 to 1980. The evaluation period remains the same from 1981 to 2010.

5. ***Comment:*** *"skill" is used throughout the paper for different metrics. Explaining what type of skill each metrics address would benefit the paper and clarify the conclusions. Which approaches is best for predictability, mean errors, etc.*

   Response: We considered two measures of skill: Rank correlation (which tells us how well the variability in simulated streamflow is captured, but not magnitude) and the Mean Square Skill Score – MSSS (which tells us about the mean squared errors in terms of forecast accuracy, i.e. whether the forecasted streamflow is better than or inferior to a reference forecast model). Both these skills, rank correlation (variability) and MSSS (mean errors), have been clarified in the text throughout the paper.

6. ***Comment:*** *Needs clarification and reorganization in a couple of places. In particular the analysis approach is not well defined or presented. The result section could be focused on how to best answer the scientific questions. It would give more organization in the results section as well.*

   Response: The analysis section has been revised as a separate sub-section prior to results. The results section has been modified to provide clarifications of analysis.

7. ***Comment:*** *Not sure why Flint is presented. There is no conclusion associated with this location*

   Response: We have indicated that the results and conclusions are for the Flint River basin at the beginning of the conclusion section. Flint River is presented to demonstrate that soil moisture and streamflow are simulated reasonably well in the upstream sub-basin besides the entire river basin. The goal of presenting Flint River is to highlight that beyond 3 months lead time, correlations between soil moisture and observed streamflow are not statistically significant at sub-basin scale, however, such correlations are still significant at the large basin scale.

8. ***Comment:*** *Adding a diagram presenting the full experimental design would help clarifying the paper*

   Response: Figure 2 has been updated to show the experimental design.

   Specific Comments:

9. ***Comment:*** *The analysis is specific to a location – the title is then somewhat misleading. I would suggest adding the region in the title.*

   Response: The title has been revised as "Role of Climate Forecasts and Initial Conditions in Developing Streamflow and Soil Moisture Forecasts in a Rainfall-Runoff Regime".

10. ***Comment:*** *Bias correction of the VIC flow: it is unclear how it is performed "based on calibration*

    Response: We have added more details on the bias correction. Please see the response # 2.

11. ***Comment:*** *Performance". It is all the more confusing that UW usually performs a quantile mapping- based bias correction . I would suggest the authors to clarify the bias correction approach they used.*

    Response: The bias correction is carried out only in predicting the mean monthly flow values. We did not consider any quantile mapping. For further details, please see the response # 2.

12. ***Comment:*** *P5227L21: replace "soil moisture skills" by "IHCs, in particular soil moisture" or something equivalent.*

    Response: "Soil moisture skills" is replaced by "skill in forecasting soil moisture" since we are specifically referring to soil moisture here.

13. ***Comment:*** *P5228L7: the term "updated precipitation forecasts" is confusing. This is a substitution of the precipitation forecasts from the ESP approach by precipitation forecast from GCMs? In the ESP approach there is an ensemble of precipitation. Is the GCMs seasonal precipitation forecast deterministic or is it an ensemble as well?*

    Response: The term "updated precipitation forecasts" is replaced by "updated monthly precipitation forecasts" which refers to the ECHAM4.5 precipitation forecasts that are updated every month for 6-month lead time. This term has also been used by other agencies such as National Weather Service Climate Prediction Center (NWS CPC) (http://www.cpc.ncep.noaa.gov/products/predictions/multi_season/13_seasonal_outlooks/color/churchill.php). Here, the precipitation forecasts ensemble mean time series from seven selected ECAHM4.5 GCM (from 1957 to 2010) is first spatially downscaled and then temporally disaggregated to the scale of VIC model implementation (1/8° at daily time step). Since previous studies such as Kumar et al. (2001) have shown that limited information is

available over individual members of the ensemble or the ensemble spread, we have considered the ensemble mean for downscaling the precipitation forecasts.

In contrast, the ESP approach uses climatological forcings as an ensemble. The daily precipitation forcings during the climatological period of 1957 to 1980 at 1/8° are used in the ESP approach to implement the VIC model with updated IHCs during 1981 to 2010.

14. ***Comment:*** *P5229L4: seasonal forecasts issued once a month cannot really support a "real* time forecasting system" but rather a planning system.

    Response: We agree with this comment. We have updated this in the text.

15. ***Comment:*** *P5229: please clarify the approach – what is the baseline seasonal forecast, ESP? And then you substitute the ensemble precipitation forecasts by a GCM deterministic forecast? Do you keep it as an ensemble?*

    Response: We have added more details on this. Please see the response to comments #1 and 13.

16. ***Comment:*** *Figure 1: 12.7% bias is relatively significant. What type of calibration was performed on VIC? How many precipitation gauges used to derive the Maurer et al. dataset lie in the basin? What is the degree of regulation and consumptive use in the basin that could partially explain the difference. Is there any literature evaluating the latent heat simulated by VIC over this basin with another model for example? The point of the question is that the rest of assessment of skill is relative to observations. When there is a systematic bias like this, any low bias in precipitation forecast for the Spring will drive artificially to a decrease in the mean errors and flow improvement. It would be good to add a succinct analysis of precipitation forecasts so that we can better evaluate the sources of improvement.*

    Response: Mosley and McKerchar (1993) suggested that there could be ±10% errors in streamflow measurements in the observed gauges. Therefore, the percentage bias around 10% is reasonable. Further, we accounted for the variability captured in simulated flows in comparison to the observed flows through estimating Nash-Sutcliffe Efficiency (NSE). The NSE for the evaluation period is 0.81 which is reasonably high. Finally, we corrected the systematic biases in the simulated streamflow over the calibration period and applied it to the evaluation period to remove the effects of model biases.

    Maurer et al. (2002) selected all the NCDC stations (in the multi-state region) that have consistent data of at least 20 years during 1950 to 2000. The meteorological data was gridded to 1/8° spatial scale using the Symap algorithm. The number of nearest neighbors for precipitation was selected to 15 in the Symap algorithm to avoid any sharp discontinuities in the gridded data due to low station density.

    Manual calibration was performed on the VIC model. The soil parameters that were calibrated include: (a) Ds - fraction of maximum velocity of baseflow (Dsmax) from the lowest soil layer where non-linear baseflow begins, (b) Dsmax, (c) binf – parameter that

define shape of the VIC model curve, (d) Ws – the fraction of maximum soil moisture of the lowest soil layer where non-linear baseflow starts, (e) Soil depth of second and third layers.
The effects of regulation and consumptive use have been removed in the observed streamflows as this basin is a part of HCDN network.
Unfortunately, Ameriflux data (and others) do not have any site located in the study area that measures latent heat fluxes for direct comparisons of model simulated fluxes. Comparison of VIC model simulated latent heat fluxes with other hydrologic models is beyond the scope of this study due to uncertainties added by different models. However, Sinha et al. (2010) have implemented the calibrated VIC model (NSE = 0.78, which is similar to the NSE in this study) in the Midwest (different climatic conditions) and the VIC model was able to predict monthly soil moisture and soil temperature reasonably well in comparison to observations at several sites in Illinois.

17. ***Comment:*** *Section 2.3: why would you select ECHAM4.5 GCM grid cell that have the best rank correlation (for which lead time? 6-month accumulated or monthly precipitation?) and not take the overlying grid cells?*
Response: The selected seven ECHAM4.5 grids are covering the entire study area and are adjacent to each other. The locations of the ECHAM4.5 grids are added in Figure 1a. A table with rank correlations comparison between seven selected ECAHM4.5 grids (at different lead times) and spatially average observed precipitation is added as Table 1.

18. ***Comment:*** *P5332: specify that the calibration period is also the period of the analysis. In this context the VIC model structure and gridded dataset uncertainties are known and quantified for the remaining of the analysis. If the skill of the seasonal forecasts are evaluated with respect to observations, the assessment of skill and "improvement" should take into consideration the initial model errors. In this respect it is surprising that the reference is observed streamflow instead of the VIC simulation forced with the gridded observed meteorological dataset.*
Response: The model calibration period is independent of the evaluation period. We agree that there will be model errors in estimating initial conditions and therefore, we have added a comparison with respect to VIC simulated streamflow (using the observed gridded data) as a reference.

19. ***Comment:*** *P5232: specify that the spatial downscaling is performed using the observed gridded meteorological dataset as reference.*
Response: We have added information on the reference on the observed gridded meteorological dataset (Maurer et al., 2002).

20. ***Comment:*** *P5232: many different periods are used so far : 1981-2010 is the period of the overall experiment, 1981-2010 is also the period of the VIC calibration, 1957-1980 for*

deriving precipitation monthly anomalies, 1957-2010 for the principal components, 1951-1980 for the temporal disaggregation. It would be good to use some consistency.

Response: The calibration period and statistics are updated to 1957 to 1980 for consistency. The evaluation period is from 1981 to 2010.

21. ***Comment:*** *Figure 2: please add in the diagram the different experiments: baseline, ESP, down-scaled GCMS,*

Response: We have added Figure 2b, so that it provides information on the considered three experiments.

22. ***Comment:*** *P5235 L17: replace "Ensemble Streamflow Prediction" by "Extended Streamflow Prediction".*

Response: We used ESP as "Ensemble Streamflow Prediction" on the basis of several other studies such as Shukla and Lettenmaier (2011) and Wood and Lettenmaier (2008). We imply ESP as primarily VIC forced with climatological precipitation ensemble.

23. ***Comment:*** *Specify here how you handled the ensemble of precipitation forecasts used for the ESP approach.*

Response: We have discussed this already with reference to comment #13. We considered climatological forcings in the form of ensemble to obtain the streamflow forecasts under ESP.

24. ***Comment:*** *P5235; please clarify which bias correction approach was applied? Was the objective to remove the uncertainties of the baseline simulation with respect to the observations?*

Response: Please see the response to comment # 2. Yes, the objective of bias correction was to remove consistent bias in estimating the mean monthly streamflow in baseline simulation during 1957 to 1980.

25. ***Comment:*** *P5236 Line 2: are you evaluating the "skill of the VIC model" or "trying another way to derive streamflow forecast in order to evaluate the role of model uncertainties? This is not one of your scientific question or does not seem directly motivated to address the scientific question. Please clarify the added value of this additional experiment.*

Response: We are trying to derive streamflow forecasts using a well calibrated hydrologic model. Then the simulated streamflow are bias corrected on monthly basis to compare them with the observed streamflow. Thus, it is better to look at the performance of forecasted streamflow in retrospective runs and compare it with the observed streamflow. We have also added comparisons with respect to VIC simulated flows (when forced with the observed meteorological forcings) as a reference.

26. ***Comment:*** *P5236 section 3.2.1: introduce the metrics you are using for the analysis – what aspect of the forecast they represent; mean errors, variability, predictability, etc*

    Response: Rank correlation addresses the variability and predictability while the metric MSSS represent mean errors and added advantage of using precipitation forecast with respect to a reference forecasts (streamflow climatology). This has been clarified in the text.

27. ***Comment:*** *P5236L10: Specify the baseline for the PCR: observed streamflow?*

    Response: The PCR model was developed from 1957 to 1980 as training period or base line (same as calibration period for the VIC model). Then the estimated PCR regression parameters during the training period were used to estimate streamflow during forecasting period 1981 to 2010.

28. ***Comment:*** *P5237: in the transition, perhaps introduce the analysis and which question it is supposed to address. Present the analysis before the results.*

    Response: Analysis is now introduced prior to results.

29. ***Comment:*** *P5238L13: The statement is not supported. If it was due to VIC poor performance it would be seen on both VICfcst and VIC clim experiments.*

    Response: This statement has been deleted.

30. ***Comment:*** *I suggest having the results section more organized, either by season, by metrics. Not necessary by lead time.*

    Response: The results section has been modified to provide clarifications of analysis.

31. ***Comment:*** *P5239L25: the VIC simulation of low flows for the baseline (Figure 1) seems better than the high flow season. There seem to be no real support for the "VIC model's inability to simulate low flows". It is possible, but just not supported here.*

    Response: We agree that percentage bias is high in VIC simulated streamflow during the high flow season than low flow season, however, the Nash-Sutcliffe efficiency in the low flow season is much lower. Although, the mean of simulated low flow season streamflow is closer to the observed flows, the variability in simulated streamflow differs a lot than the actual flows (Please see Table 1, low rank correlations during Sep to Nov). In contrast, high flow season variability is much better simulated. Hence we conclude that VIC model's performance is relatively poor in simulating low flows.

32. ***Comment:*** *P5240: the skill of ECHAM4.5 is assessed during ENSO conditions. Were different traces considered for VICclim in order to have a similar ENSO conditions? Did the spatial and temporal disaggregation training period got aligned with ENSO modes?*

    Response: We are not sure what the reviewer is alluding to here. The experimental design and analyses are the same. The only difference is that we estimate the skill of streamflow

forecasts from ECHAM4.5 forecasts and climatology during ENSO conditions. Basically, we assess the skill of the streamflow forecasts obtained using ECHAM4.5 forecasts and climatological forcings under ENSO conditions. For this purpose, we identified whether Nino3.4, the ENSO index, is under warm pool or cold pool over the forecast period. If so, during those years, the streamflow forecasts obtained from ECHAM4.5 and climatology are grouped and their skill is assessed.

With regard to climatological forcings, we did not consider different traces for VICclim to have similar conditions. We wanted to keep the VICclim climatological period consistent with the PCR training period (using 24 years of data from 1957 to 1980) for fair comparison. The spatial and temporal disaggregation are not directly aligned to the ENSO modes, however, in looking for nearest neighbors during temporal disaggregation in the historical years, similar ENSO conditions would have been used.

33. ***Comment:*** *Did the PCR get calibrated over specific ENSO years for the same evaluation? The ECHAM4.5 might have de factor an ENSO signal. It makes then the comparison with VICclim and PCR relatively unfair.*
Response: PCR was not calibrated (trained) over specific ENSO conditions as less than 20 years of data being available as training period. In order to have a fair comparison between VICclim and PCR, we used all the 24 years data (1957 to 1980) in the same way by not accounting differently for traces due to ENSO conditions. The main objective of the analysis is to understand the utility of forecasts and climatology in improving streamflow predictions as a function of lead time.

34. ***Comment:*** *"skill " is used for many metrics instead of "predictability" "mean errors" etc. Please be more specific so that we can summarize at the end what skill means.*
Response: We have mentioned the appropriate measure wherever needed. Please also see response # 5.

35. ***Comment:*** *Section 4.4 line 1: the first sentence justifies the reminder of the paragraph by stating that VICclim and VICfcst have "good skill" at one month lead time. Please justify if this is mean errors, or reliability which will in turn allow assessing if the spatial variability is accurate in terms of predictability, mean errors, etc.*
Response: It has been now clarified that when we refer skill to variability and mean errors throughout the paper.

36. ***Comment:*** *Why is Flint presented? I t does not seem to bring any value.*
Response: Since we wanted to evaluate the streamflow forecasts skill in a rainfall-runoff regime over the Southeast US, we considered the Flint River. Please also see the response # 7.

37. ***Comment:*** *P5244L15-18: again the statement is not supported if both VICclim and VIC fcst does not show the same pattern.*

Response: We have revised the sentence stating that statistical model performed better than both VICclim and VICfcst during the fall months.

38. ***Comment:*** *P5246L20: the ensemble mean precipitation forecast should not be used to drive the hydrology model.*

Response: We used ensemble mean precipitation forecasts for downscaling the precipitation forecasts on the basis of previous studies such as Kumar et al., (2001), which indicate that limited information is available over individual members of the ensemble or the ensemble spread.

39. ***Comment:*** *Need to assess even briefly the performance of the spatial and temporal disaggregation. It seemed that some of the figures were assessing the point but are not related to metrics used in the analysis.*

Response: The performance of temporal disaggregation is also added to figure 3b.

40. ***Comment:*** *Table 2 – why show Flint?*

Response: We evaluated the performance of simulated soil moisture at an upstream sub-basin "Flint" to demonstrate that our results are applicable to the entire river basin. Please see the response # 7.

41. ***Comment:*** *Conclusion 2: skill of soil moisture forecast cannot be supported – rather look at soil moisture patterns.*

Response: In this study, we compared the skill in forecasting soil moisture by comparing it with observed streamflow at upstream sub-basin and the entire river basin due to lack of observed soil moisture measurements. We have also added a comparison of RMSE (in Table 2) in soil moisture forecasts during different seasons over the 1981 to 2010 period by using VIC simulated soil moisture (obtained using observed forcings) as a reference. Finally, we also considered the deviations of forecasted monthly soil moisture conditions during La Nina conditions from the normal (climatological) conditions, since La Nina conditions associate with droughts over the region.

42. ***Comment:*** *P5238L12: if VIC had an issue simulating the low flow season in September then it should be on both VICfcst and VIC clim experiment. Could it be due to the GCM model forecast? I believe you actually mean the climate forecast model. Please specify in the text.p5239L25 confirms that you ment VIC. Although low flows are usually hard to simulate, it should be on both VIC experiments. The statement is not supported.*

Response: This statement has been deleted.

43. ***Comment:*** *Figure 1 seemed to show good skill. In the text, difficult to see to which basin it applies – what are the conclusions based on the difference between the 2 basins?*

Response: We have focused all the analysis at the Apalachicola River at Chattahoochee, FL which comprises of both Flint and Apalachicola River sub-basins. In addition, we evaluated the performance of simulated soil moisture at an upstream sub-basin "Flint" to demonstrate that our results are applicable to the entire river basin.

44. ***Comment:*** *Section 4.4: there is no evaluation with respect to observations, and the section does not bring skill assessment of the seasonal forecast. I would suggest the authors to elaborate on it and refer publication over expected effects of La Nina/El Nino. There are no real conclusions drawn from the section*

Response: Soil moisture observations are limited in both spatial and temporal scales. So, it is difficult to validate with observations. Therefore, we used VIC simulated soil moisture as our reference to compare the soil moisture anomalies due to precipitation forecasts.

45. ***Comment:*** *Adding a diagram presenting the full experimental design would help clarifying the paper*

Response: We have included Figure 2b to describe the experimental design. Please see the response # 8.

Thanks for the detailed comments!

References used:

Kumar, A., A. G. Barnston, and M. P. Hoerling.: Seasonal predictions, probabilistic verifications, and ensemble size, Journal of Climate, 14, 7, 1671-1676, doi: http://dx.doi.org/10.1175/1520-0442(2001)014<1671:SPPVAE>2.0.CO;2, 2001.

Mosley, M.P., and A. I. McKerchar.: Streamflow. In: Maidment, D.R. (Editor-in-Chief) Handbook of Hydrology. McGraw-Hill, chapter 8, 1993.

Shukla, S. and Lettenmaier, D. P.: Seasonal hydrologic prediction in the United States: understanding the role of initial hydrologic conditions and seasonal climate forecast skill, Hydrol. Earth Syst. Sci., 15, 3529–3538, doi:10.5194/hess-15-3529-2011, 2011.

Sinha, T., Cherkauer, K. A., and Mishra, V.: Impacts of historic climate variability on seasonal 25 soil frost in the Midwestern United States, J. Hydrometeorol., 11, 229–252, 2010.

Wood, A. W. and Lettenmaier, D. P.: An ensemble approach for attribution of hydrologic prediction uncertainty, Geophys. Res. Lett., 35, L14401, doi:10.1029/2008GL034648, 2008.