Response to interactive comment by Anonymous Referee #1

In this study performance of streamflow forecasts for Kharif Season (April-September) in the Upper Indus Basin of Pakistan is assessed. Streamflow forecasts are generated using the Bayesian joint probability (BJP) approach. Several predictors such as antecedent flow, climate indicators, and ESP based streamflow forecasts are used to test the performance of the streamflow forecasts. The study finds that in general BJP streamflow forecasts based on predictors antecedent flow and climate indicators perform the best. Variation in the skill is found for the focus basins, and for the early and late part of the season. In general, the manuscript is well organized and methods are technically sound. I do have a few comments/suggestions, some of which are moderate to major, which need to be addressed before publication.

Response: Thank you for your encouraging and helpful review.

Major comments:

(1) It would be helpful, mostly for the readers who are not well aware of the seasonal cycle of climate in the region, to add a figure for both basins that show the seasonal cycle of precipitation, temperature, runoff/streamflow. Similar to Fig. 2 of this manuscript http://journals.ametsoc.org/doi/full/10.1175/JHM-D-14-0213.1. Such a figure would provide a needed background to the readers about the region and also help interpret the results of streamflow forecasts evaluation.

Response: We will add a new figure to show the seasonal cycle of streamflow, precipitation and calculated potential evapotranspiration for our two study basins.

(2) The authors mention lack of climate forecasts skill in this region. I would encourage them to show a map(s) of the long-term skill of at least rainfall (winter) and temperature (winter and summer) in the region. I think a case for using statistical forecasts such as ones presented in this study can be made better if statistical forecast skill is demonstrated relative to the skill of dynamical forecasts, not just climatological forecasts. As of now, there are several global dynamical forecasts systems that provide operational seasonal forecasts. One of them being the North American Multimodel Ensemble (NMME, http://www.cpc.ncep.noaa.gov/products/NMME/).

Response: We agree with the comment that it will be useful to assess statistical forecast skill together with dynamical forecast skill. The reason that this has not been done in this study has do with the strong practical application focus of the tool development. Because of the need to model glacier and snow processes, sophisticated hydrological models that could ingest climate forecasts are of limited availability. Furthermore, the region does not have ready access to real-time seasonal climate forecasts, although this problem can be overcome. The purpose of this study is therefore to develop tools that can be easily implemented based on infrastructure that exists today. As infrastructure (like hydrological models) improves, an extended study on a fully dynamical forecasting system will be highly appropriate. As dynamical forecasts need to be post-processed for use in hydrological forecasting, the linkage of dynamical climate model forecasts with hydrological models in the region would require substantial additional research.

In the Introduction, we will provide more information on the background of this study and clarify the strong practical application focus of this study, including why a fully dynamical forecasting system has not been investigated. We will also add to the discussion, regarding potential further research with dynamical models.

(3) The authors use March streamflow is the only predictors reflecting antecedent conditions, it is not clear why other variables such as snow water equivalent, soil moisture, total water storage were not used. Nowadays observations (through remote sensing) or simulations (e.g. through GLDAS https://ldas.gsfc.nasa.gov/index.php) of those variables are readily available. Especially in a region where snowmelt runoff is dominant, I would think snow and soil moisture would provide some streamflow forecast skill.

Response: We did investigate a MODIS-based snow cover product (post-processed to remove cloud cover effects) and did find a relatively high correlation with streamflow. For example, for the Jhelum catchment the MODIS snow cover at the end of March has a 0.68 correlation with Kharif season streamflow. However, the MODIS data only commenced 2000 and so this limited availability of data resulted in its exclusion from the final set of model evaluations. As March flow correlation with Kharif season streamflow is also 0.68, a snow cover predictor is not expected to be a demonstrably better predictor than March flow in this case.

(4) I would also encourage the authors to provide some more details regarding the PIT plots in the method section. To my knowledge PIT is not a typical metric used for forecast evaluation so it would help the readers to get a bit more details on them and also briefly describe what each type of the figures (a through e) highlights regarding the forecast skill.

Response: We will add a more detailed explanation, as well as further cite published literature regarding the use of PIT plots to evaluate probabilistic forecasts.

Minor comments:

(5) P2, L24: Not only P and T but other atmospheric forcings as well.

Response: We will change the text to remove specific mention of P and T: "*Dynamical approaches use hydrological models initialised with observed inputs up to the beginning of the forecast season to account for antecedent conditions, that can be driven either by historical or modelled climate inputs.*"

(6) P2, L29: This statement regarding the skill of dynamical forecast skill should be made more specific, e.g. mention the regions and seasons etc.

Response: We will add a summary, as a specific example, of a published evaluation of dynamical climate forecast skill for the Asian summer monsoon from ECMWF and NCEP forecast systems.

(7) P3, L23: Summer streamflow would depend upon winter T too, as winter T would influence snow accumulation. Please revise.

Response: We will revise to include reference to winter T: "*The predominant source of flow in the UIB is snowmelt, with glacier melt a secondary source, with 80% of flow occurring during the June-September summer period. Interannual flow variability is thus controlled by two processes, snow accumulation as determined by winter precipitation and temperature and meltwater generation as determined by summer temperatures. Hence snowmelt-generated flow is a function of winter precipitation and temperature and also summer temperature, whereas glacier melt is primarily a function of summer temperature, although glacier melt rates are also influenced by snow cover.*"

(8) P4, L5-10: These sentences are confusing and hard to understand.

Response: These sentences will be replaced with a clearer statement: "*Useful climate indices should relate to the weather prior to the forecast season, providing an indication of snow accumulation, and* *also to the weather within the forecast season, influencing temperature and hence snow and glacier melt rates*."

(9) P5, L21: Please see comment #2.

Response: See response to comment #2 above.

(10) Results in Table 1 and 2: It is not clear if those results are after cross-validation or before? Or are the results presented in Figure 3 onward are cross-validated? I would suggest comparing the cross-validated skill vs the skill calculated using the entire period.

Response: These are cross-validated results, and the Table's headings will be updated to reflect this. In the paper we will re-emphasis our case that cross validated results are more representative of the real skill and reduce the chances of overfitting. Hence we do not want to show results without cross-validation given statistical methods are prone to artificial skill and overfitting.

(11) Section 3.3: Suggest dividing this section into three sub-sections to discuss each of the verification scores separately.

Response: As we rely on cited references for the provision of detailed descriptions of the skills scores, we feel there is not enough stand-alone material for each to justify its own sub-section.

(12) P8, L2: It is surprising to see that MEI (May-June) from the previous year is a skillful predictor. Could you comment on why that may be? During May-June, ENSO events are in initial development stage and sometimes may change signs in the later part of the year so it is surprising that in this case, you are finding MEI May-June to be a skillful predictor for the streamflow of the following year.

Response: The MEI (May-June) predictor skill relates to autumn/winter snow accumulation, a lag of 4 months to snow accumulation from October onwards. It is thus not unreasonable that circulation systems bringing moisture into the region during autumn/winter are influenced by the forcing initiated by ENSO processes during the summer.

There are many supporting references presenting details of ENSO/precipitation teleconnections for the region, with several cited in the manuscript. As an example, Mariotti (2007)[1] notes: "*The associated circulation pattern during El Nino (La Nina) involves a southwesterly (northeasterly) moisture flux that brings more (less) moisture into this region* [southwest central Asia]. *This flux flows along the northwestern flank of the large-scale high pressure anomaly over the Indian and western Pacific Oceans, broadly, the western pole of the Southern Oscillation see-saw pattern. Unlike many ENSO teleconnections in which lower pressure and weaker subsidence leads to more precipitation, this mechanism does not require a change in the local pressure, but rather involves a change in the tropical moisture supply to subtropical-midlatitude storms.*"

(13) P8, L14-15. I thought that in some cases March flow was the highest skill predictor and adding any predictor didn't increase the skill so why are you using both March flow and climate predictors here?

Response: For Indus at Tarbela (Table 2), inclusion of a climate predictor in addition to the March flow predictor increased skill in all cases. For Jhelum at Mangla, there were mixed results as adding the climate predictor slightly decreased the skill for the full 1975-2015 period but increased the skill

[1] Mariotti, A.: How ENSO impacts precipitation in southwest central Asia, Geophysical Research Letters, 34, 10.1029/2007GL030078, 2007.

for the 2001-2015 period. As the 2001-2015 period is the comparison period for SRM, we think it is acceptable to use these predictor combinations.

(14) Figure 5 and 6. These figures are used to compare the skill of BJP vs SRM based streamflow forecasts. I think it would be better to combine them both into one figure. Maybe just show SRM forecasts with a different color.

Response: We are concerned that combining multiple forecast evaluations onto single figures would look cluttered in some circumstances, such as Panel (d), hence out preference is to maintain these figures as they are.

(15) Conclusion: The last two bullet points are not really findings. I suggest to separately discuss them after listing the findings. Also please mention here the current state-of-the-practice for generating streamflow forecasts in the region and the value the methods explored in this study will add.

Response: Agreed, these last two bullet points will be presented as a separate paragraph after the list of findings. We will add a description of current methods and add a discussion of the value added by our new approach.