

Interactive comment on “Initial assessment of a multi-model approach to spring flood forecasting in Sweden” by J. Olsson et al.

P. Crochet (Referee)

philippe@vedur.is

Received and published: 21 August 2015

The authors propose three new approaches to spring flow forecasting in Sweden and compare their performances to the method currently in use at SMHI, on three selected basins. Despite their complexity, none of these new methods consistently outperform the one in operational use. When all methods are combined into a multi-model approach, the improvement becomes substantial. Seasonal hydrological forecasting is a challenging topic with important implications for reservoir operations, hydropower production planning and water resources management. Spring flow forecasting is especially important in regions where snow accumulation and melt play an important role in streamflow seasonality. Overall, I found this paper very interesting and the proposed

C3186

multi-model approach promising. However, although the authors have acknowledged the fact that basic versions of the proposed methods were used, I think that several aspects of this study could be improved, regarding in particular the methods implementation and the analysis of results.

Major comments:

1) A few additional figures would be welcome to help analysing the results. Showing the distribution of SFV and some statistics (min/median/max) would help understanding how large is the inter-annual variability on the different basins. Scatter plots of forecast vs. observed SFV, or time series plots would help the reader to better understand and compare the quality of the different SFV forecasting methods. This should also help analysing some discrepancies in the results, when e.g. $FY \geq 50\%$ while RI is negative (See Table 3). What makes RI negative in these cases? a few outlier years? Can we conclude that the new methods in question are worse than the reference one in these cases? Which criteria is the most important in the present context: RI or FY?

2) It appears that HBV has been calibrated on the period used to evaluate the forecasts rather than on an independent period (Table 1). The skill of the hydrological model is most likely overestimated, compared to what it would be in an operational environment. This approach could be acceptable if HBV-based methods only were compared in the study. However, in the present case, I would think that the SD method is disadvantaged. This should be discussed at least, if not reconsidered.

3) The use of ERA40 and ERA-Interim to run the CP-based analogue method in forecasting mode is questionable. The present application of the method does not reflect the expected skills in operational conditions and the improvements relative to the baseline method are possibly overestimated. In an operational environment, operational NWP analyses rather than ERA-40 or ERA-Interim will be used to define the CPs to be compared to historical CPs derived from ERA-40 and ERA-Interim. Therefore, I would expect that discrepancies or inconsistencies between ERA-40 and operational NWP

C3187

analysis will impact the method's skills. In order to better reflect expected shortcomings in operational conditions, I would recommend to either re-calculate the anomaly $g(i,t)$ and corresponding CPs for the years to be forecasted, using available operational NWP data, or change the validation period so that ERA-Interim only is used in forecasting mode and compared to ERA-40 historical CPs.

4) Since three of the new methods (two analogue methods and ECMWF seasonal forecast) aim at improving the meteorological forecasts, they should have been evaluated with that respect before being evaluated with respect to SFV.

5) In particular, considering the coarse horizontal resolution of ECMWF seasonal forecasts, biases can be expected to affect precipitation and temperature forecasts at the basin scale, especially in complex terrain. This should have been investigated before using them into HBV rather than after. A bias correction or an adaptation strategy is probably necessary. On the other hand, the frequent model updates can be an obstacle to the development of robust correction strategies. This should also be discussed.

6) The strength and limitations of analogue methods in general and the proposed ones in particular need to be discussed. What about: i) the impact of the archive length on the correct identification of analogue years, ii) the impact of the number of selected analogue years on the quality of the reduced ensemble and its median value, iii) the validity of the assumptions behind the proposed analogue methods, i.e. the degree to which antecedent meteorological conditions, 1 to 6 months prior to the forecast issuing date, are relevant to the prediction of future meteorological conditions up to springtime, iv) the validity of the optimisation method for CPs, which is based on precipitation only and not temperature, yet a key variable in the formation and melting of snowpack. Also, why is the optimisation conducted on one catchment only ?

7) Concerning the SD method, it seems that forecasts (rather than an analysis) are used to define the predictors for the calibration of the predictors-predictant relationships. The forecasting errors may deteriorate the nature of the underlying relationships.

C3188

What is the rationale behind that ?

8) I was wondering whether the skill of the multi-model approach could be further improved after eliminating the worst method(s) and/or by defining a weight proportional to the skill of the method rather than just the rank. I am also a bit surprised to see that the multi-model approach can perform better than the baseline method, when all new methods seem to perform worst.

9) Ensemble forecasts are developed but only the ensemble median is evaluated. The advantage of making use of ensembles is not discussed. In particular, ensemble forecasts should be reliable.

Specific comments

1) p. 6082: lines 7-8: is the regulation of Angermanalven and Ljusnan rivers taken into consideration by HBV ? Does this have an influence on the skills of the different methods, including those not using HBV ? This should be kept in mind when analysing the different results.

2) p. 6082 line 14: do you mean 900×10^6 and 8000×10^6 ?

3) p. 6083 line 1: I would suggest to write somewhere that the output of HBV is daily Q, although this might be obvious to those familiar with HBV.

4) p. 6085, lines 14-15: Are you using all historical time series to run the CE baseline method, including the current year under investigation, or only those prior to current hydrological year ?

5) p.6087 line 6: can you explain how is persistence defined and used ? Are the TCI calculated for individual months or for the entire period (1 to 6 months) ?

6) p. 6087 lines 11-15: have you tried to use one TCI only rather than a combination of 2 or 3 TCIs ? what is the rationale behind this choice ? is it the result of an optimisation ?

C3189

7) p. 6087 (TCI) and p. 6089 (CP?s): the number of selected analogue years differs from year to year for each method and lead times. How is this impacting the quality of the ensemble and resulting median ? Please give the number of selected analogue years for each forecast year in a table or at least some statistics (min, max, median). If too many years are selected, compared to the total number of years, then the method will converge toward the climate ensemble baseline method. If too few years are selected, the uncertainty of the reduced ensemble will be very large.

8) p.6088: how do you go from Eq. 2 to CPs? Do you define several $g(i,t)$ classes ? Please explain better.

9) p. 6089 line 18: how many CPs have you defined ? Can you give statistics on the frequency of occurrence of each CP over all years ?

10) p. 6089 line 19: how do you define persistence and how is it used ?

11) p. 6089 line 20: did you arbitrarily define the rule of using the 2 most frequently occurring CPs or did you try other possibilities ? did you validate this choice against forecasted P and T ? Please explain better.

12) p.6092 lines 4-9: I would suggest to move this paragraph to Section 3.4. Also, a figure illustrating the practical application of the statistical downscaling method would be appreciated, such as a scatter plot of the relationship(s) in calibration mode (median prediction against observed SFV).

13) p.6092: lines 4-5: If I understand correctly, the seasonal predictors are calculated for different periods. For a forecast issued on 1 January, the predictors are defined for the period Jan-Feb-Mar, for a forecast issued on 1 March, the predictors are defined for the period Mar-Apr-May and for a forecast issued on 1 May, the predictors are defined for the period May-Jun-July ?

14) p. 6093 line 10: According to Eq. 6, MAE is defined in $m3$ and not in %. I would suggest to either redefine MAE in Eq. 6 to make it in %, or to give results in $m3$

C3190

whenever MAE is used.

15) p. 6094: line 23-26: A direct assessment of weather forecast errors will help to clarify the situation.

16) p.6096 lines 1-4: What do you mean by climate phenomena ? can you explain better ?

17) p. 6097, line 1-2: This demonstrates the importance of calibrating the method for each catchment separately. Why not doing so ?

18) p. 6097, line 3: was it in Section 3.1.2 or 3.2.2 ?

19) p.6097: lines 5-10: You claim that a higher uncertainty is observed because discrepancies between ERA-40 and ERAInterim datasets are leading to inconsistencies in the CP classification. Therefore, if the method was applied operationally, operational ECMWF analysis would have to be used to define the CPs for the past 1 to 6 months, and there would be discrepancies too. So in practise, I am not sure that you will be able to obtain this 10-20% improvement.

20) p.6097, lines 16-17: this should have been done first.

21) p.6098, lines 1-7: I don't quite see the rationale behind the use of each of all these parameters and some explanations would be welcome. Also, I would prefer to see the list of predictors for each forecast issue date.

22) p. 6098, line 16-17: same with analogue methods.

23) p. 6099, line 10-11: This should be mentioned in the presentation of the CP approach.

24) p. 6099, lines 18-21: why ? could this be related to a few poor forecasts only ? Scatter plots may help understanding this better. Also, add information about FY in Table 6.

C3191

25) p. 6109, Table 3: Please give a global average for RI and FY for each method, over all rivers and lead times, as in Tables 4 and 5.

26) p. 6110: Table 4, TCI6 1/1: results are slightly different than in Table 3. Please check values.

27) Tables 4 and 5 could be skipped.

28) p. 6113, Fig. 1: Please add information about the spatial domains used to define CP and SD.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 6077, 2015.