Interactive comment on "The skill of seasonal ensemble low flow forecasts for four different hydrological models" *by* M. C. Demirel et al.

S. Jörg-Hess (Referee)

stefanie.joerg@wsl.ch

Received and published: 27 June 2014

Summary

The research article presents a comparison seasonal low-flow forecasts with four hydrological models. Two of the models are conceptual models with a different level of complexity. The other two models are artificial neural networks (ANN) that are calibrated with historical data and with one hidden neuron. The authors combine different forecast inputs to evaluate the effect of the input uncertainty on the low-flow forecasts.

General comments

The article addresses an interesting topic an highlights the sensitivity of different model

structures to the representation of the meteorological input. The purpose of the work and the conclusions are well elaborated. The structure of the article is clear and technically sound. The Figures and Tables are well selected with room for improvement in helping the reader follow the analysis. The article would benefit from clarifying and deepening some parts.

(1) Introduction: You spend a lot of time in introducing runoff forecasts with climate indicators and forecasted meteorological variables and present studies on different rivers. For me this is not really relevant for the following article. I would rather prefer to read more about ensemble predictions and the effect of ensembles and historic data on the runoff predictions. You could already introduce here the difference of the conceptual and the artificial neural network (ANN) models.

(2) Methodology: Especially section 3.1.3 and 3.1.4 are difficult to follow and would benefit from some more details.

(3) Results: It would be interesting to see the effect of the ensembles also on the skill scores. I would suggest to validate the skill scores for 2 or 3 cases and add the skill scores to the figures in section 4.2.

Specific comments

(1) Abstract: P 5378 L21: For avoid confusion I suggest to change 'over-predict low flows' to 'over-predict runoff during low-flow events'.

(2) P 5380 L16: On the previous page the work by Wang et al. (2011) is cited in the context of statistical approaches and here the work is cited in the context of dynamic approaches. Please clarify this.

(3) Section 2.1: Some more information of the catchment would be helpful. What are the characteristics of the catchment and the dominant runoff processes? Further a Figure of the catchment with the distribution of the stations would be interesting. It was not clear to me how many stations are used to estimate P and PET in the sub-basins

and what is the size of the sub-basins.

(4) Section 2.2.1: Please mention 'h' from the Table in the Text.

(5) Section 2.2.2: The ensemble forecast is available for 184-days. For your evaluation you are using the first 90 days. Is there a reason why you stop the evaluation after 90 days?

(6) P 5383 L13: The Link of the reference ECMWF (2012) has been changed to http://old.ecmwf.int/publications/newsletters/pdf/133.pdf. In this newsletter I could not find any information about the MARS system 3. Please state in section 2.2.2 whether you are using daily or weekly meteorological forecast data.

(7) P 5386 L15: Here you could describe in one sentence what is the characteristic of the global approach.

(8) Section 3.1.2.: You could refer to Table 3, when you describe the model structure.

(9) Section 3.1.3: I do not fully understand the concept of ANN models. For me some more background information would be helpful. For example it is not clear to me what is the main difference between ANN-E and ANN-I. Or what are the n inputs? For me the inputs are P and PET and Q. But from equation (1) and Table 7 it seems that you use four inputs.

(10) Section 3.1.4: This section needs some clarification. Please explain the meaning of the numbers (population size, reproduction elite count size, etc) and how you selected these numbers. For the calculation of observed low-flow days the Q75 of the simulation is used. How do you account for systematic biases of the model by using this threshold for observations?

(11) Section 3.1.5: How is the climate mean of the ensembles defined? For which period? How many members are used? Is it calculated with a moving window?

(12) P 5389 L4: Does 'N' (equation 6) and 'n' (equation 3) both refer to the total number

of days? If yes, please be consistent.

(13) P 5389 L17: You describe non-exceedance probabilities for medium to high flows. Please change this accordingly.

(14) P 5390 L17: You begin the sentence with 'These probabilities...'. For me it is not clear which probabilities. Please specify this.

(15) Section 3.2.4: Please add some information what is the meaning of this score.

(16) Section 4.1: The Figure shows that MAE is lowest for 1 hidden neuron. Did other studies find similar results concerning the optimal number of hidden neurons?

(17) Section 4.2: There is a large uncertainty of the predicted runoff with the first three models. For most low-flow events the most ensembles overestimate the runoff. Can you explain why the spread in the conceptual models is larger than with the ANN model? Do you have an explanation why the runoff is over-predicted? I do not see your statement that the GR4J and HBV over-predict low flow after August. For me all models over-predict low-flows during the entire period of the two years. From the two years chosen my expression is that the conceptual models perform best during fall and the performance is lowest during spring. Do you have any explanation for this? In this context it would be interesting to see some scores for the forecasts (e.g. Brier skill score). Do you have an idea why the low flow in spring 2003 are not captured in the models? May be the simulation of the snow cover during winter can explain this behaviour.

(18) P 5393 L12: State that the uncertainty range is larger in Figure 4a than in Figure 3b for the conceptual models.

(19) P 5393 L 19-24: Do you have any explanation why the low-flows are not captured in Figure 4b. Please explain why the spread of the runoff forecast is narrow in this case.

(20) Figure 3: Please enlarge the points of the observation, specify the points in the

caption and label the plots with the according model. I would appreciate it if you could apart from the visual validation, add minimum one of the scores to Figures 3-5. Is there a reason to put different grey-scales for the ensemble forecast with the different models?

(21) Figure 6: It would be interesting to see a validation of these scores with ongoing lead time also for other cases (e.g. case 2).

(22) Figure 7: Please add the number of low flow events per Figure.

(23) Table 5: change caption to: '... of low-flow events based on the Q75.'

(24) Table 6: This table could be simplified by only showing the cases that are relevant for this article.

Technical corrections

P 5380 L12: change recipitation to precipitation

P5380 L27: Start a new paragraph with: 'The first approach...'.

P 5385 L3: Please rephrase the sentence as PET is not observed.

P5386 L6: Replace NN-E with ANN-E.

P 5386 L11: Please introduce G also in the text.

P 5387 L17: The formula needs to be embedded in a sentence.

P 5389 L3: delete 'where'

Table 3: There is a shift in the first column of the table.