Reviewer #1 (Prof Renata Romanowicz)

The authors thank Dr Renata Romanowicz for her constructive comments on the manuscript. We agree with most of the points of view she expressed and we explain how we will modify the text to account for her comments.

Comment 1) From the practitioner's point of view it would be useful if the authors presented a table with minimum and maximum prediction errors from the ensemble for each model for both examined time-periods (years 2002 and 2003).

Reply from authors: We will include a table in the revised version of the manuscript that will show the minimum and maximum prediction errors of the models.

Comment 2) The authors have set themselves a difficult task in comparing the models with different input variables and different basic assumptions. Unfortunately, the comparison does not come out sufficiently clearly. It would help if the authors did not include the ANN- I model in their comparison. It uses different inputs and obscures the message the authors want to put across.

Reply from authors: We agree with the reviewer comment. We will remove the ANN-I model from the revised version of the manuscript.

Comment 3) Lines 225-229: The authors distinguish between daily P and PET data and historical Q as an input. "The first model, i.e. ANN-E, requires daily P, PET and historical Q as input. Historical Q from the previous day is used to update the model states (Table 3). This is a one day memory which also exists in the conceptual models, i.e. GR4J and HBV (Figure 1). The ANN-E is assumed to be comparable with the conceptual models with similar model structures. The second model, ANN-I, uses historical Q to update initial model conditions and three low flow indicators, i.e. P, PET and G, as model input."
Does this mean: observed flow Q up to the date when the forecast is issued?

Reply from authors: We mean observed discharge value on the forecast issue day i.e. Q(t) to forecast Q(t+1) using the three conceptual models and Q(t+90) using the ANN-I model.

The sentence below will be revised for clarity.

"Historical Q from the previous day is used to update the model states (Table 3)"

Revised version:

"Observed discharge on the forecast issue day is used to update the model states (Table 3)"

Comment 4) Line 459 Case 5: zero P and ensemble PET forecasts as input for the other three models (GR4J, HBV and ANN-E). – the figure should be shown for completeness of the discussion.

Reply from authors: We will include the figure for case 5 in the revised version of the manuscript.

Comment 5) Line 477: "The decrease in false alarm rates after a lead time of 20 days shows the importance of initial condition uncertainty for short lead time forecasts. For longer lead times the error is better handled by the models." It is not clear to me how the initial conditions can affect the false alarm rate. It rather seems that the "correct negatives" are increasing in number and may be this particular indicator is not working properly for forecasts longer than 20 days? Please explain that statement in more detail.

Reply from authors: The initial conditions, i.e. model states, are important components of a reliable low flow forecast. Therefore, an estimation error in the model states can affect especially the short term forecasts as the model improves after the spin up period. We will elaborate the discussion about Figure 6 in the revised version of the manuscript.