Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-372-AC1, 2017

Hydrology and Earth System Sciences Discussions
Open Access

Interactive comment on "Monthly streamflow forecasting at varying spatial scales in the Rhine basin" *by* Simon Schick et al.

Simon Schick et al.

simon.schick@giub.unibe.ch

Received and published: 14 July 2017

General comments

The authors thank for the careful review and the clear comments. In general we agree with the comments and think that they bring up some very interesting ideas. We also recognise that we need to use the terms 'option', 'season', and 'operational' more precisely.

Specific comments

The MOS method is presented as an option for streamflow forecasting at the seasonal time scale (Page1-Line 16, Page 2-Line 18, Page 3-Line 9). However, the results show

Printer-friendly version

Discussion paper

only forecast skill relative to climatology for the first month ahead. This is usually not what is called seasonal forecasting (rather medium- or extended-range forecasting).

Yes, we agree, the term 'season' needs to be used more carefully, especially when it comes to the actual results of the study (e.g. page 1, line 14).

In the introduction, however, we tried to give a general overview of the approaches that one could apply to integrate GCM output for seasonal (or subseasonal) streamflow forecasting: The traditional model chain GCM - hydrological model; runoff from the land surface component of the GCM run; and PP and MOS based forecast strategies. We do not claim any of these approaches to be superior nor that using GCM output is superior to the ESP approach and its variants. Rather it is an attempt to summarise the existing toolbox, without stating that any particular model is skillful at the (sub)seasonal time scale.

The fact that skill of the tested methods in case of the Rhine basin is restricted to one month ahead forecasts does not invalidate this list – these options still exist (from a technical point of view) and might work or not, depending on the study region, skill of the GCM, validity of the model assumptions, and so on. A forecast at the seasonal time scale remains one, be it skillful or not.

So the conclusion must be drawn that no skill was found at the seasonal time scale for any of the models (including ESP and H-TESSEL). This is indeed concluded for the MOS models (Page 17-Line 15), but the suggestion that the performance may be better for particular calendar months (Page 17-Line 15,16) is unfounded and should be removed.

We agree; it was not our purpose to suggest that the models might better perform in particular months, but that the findings are valid only for forecasting all calendar months. Moreover, we should stress that the model might perform better or more worse in particular calendar months, so the results represent somehow an "average" month.

Also, the conclusion that H-TESSEL is an interesting option for seasonal (i.e. beyond 1 month lead time) streamflow forecasting (Page 1-Line 13, Page 17-Line 26) is not supported by the results shown in Figure 2 (at least not clear to me).

Yes, we agree, this conclusion is too loosely formulated (again it boils down to the correct usage of the term 'season'). It also implies several arguments that we would like to reformulate here.

The present study tries to test approaches for seasonal streamflow forecasting that have not received much attention in the scientific literature so far, i.e. a MOS approach and runoff simulations from a coupled GCM. The latter simply emerged out of curiosity: While downloading the precipitation and temperature hindcasts from MARS, we realised that runoff is also contained in the list of available parameters. We then decided to request runoff as well and started to use it as a benchmark for the MOS approach. To avoid any misunderstanding: We did not run H-TESSEL at all, we only applied a linear bias correction. Thus, the simulations correspond to the model configuration used within ECMWF's S4 setup.

To our surprise it turned out that these runoff simulations are quite hard to beat, for which reason we included them in the present study (please note that the MOS approach works on an acceptable level when fed with the best available input data, independently of the lead time). Clearly, this finding does only hold for the Rhine basin and H-TESSEL, but nevertheless we think that this point is interesting both from an operational as well as scientific perspective.

Operational, since the runoff simulations are already present after the GCM finishes his run (so no need to run a model by oneself, ECMWF and other institutes already did the job, and notably in a fully coupled way). Scientific, since the land surface model is not aimed to forecast streamflow, but to provide a lower boundary condition for the simulation of the atmosphere. Thus, if we already beat streamflow climatology with runoff from the land surface model, to which extent can we benefit by using models specifically tailored to forecast streamflow? For example, what are the benefits of including river routing algorithms, groundwater or lake modules, or a higher horizontal resolution? Studies that use GCM output for seasonal streamflow forecasting need to download the necessary parameters anyway, so why not add runoff and use it as a benchmark?

Moreover, if the MOS method is to be considered for operational streamflow forecasting, it would need to be tested against the more traditional approach, which uses ESP- or GCM-driven hydrologic models.

Indeed, it would be very interesting to contrast the two tested approaches with the 'traditional' forecast strategy. Since to the best of our knowledge no study exists that we could use for such a comparison, it would be left to us to set up a model and run it with the seasonal climate predictions. This, however, is in our opinion not feasible in the context of the present article, but could lead to a next study.

I am not sure which version of H-TESSEL you are using, but since you mention it does not include routing (Page 16-Line 26), this must be a relatively limited model. When comparing the MOS method to this H-TESSEL for zero lead time, the results are not very convincing: according to Table 4 the MAE of the S4 models is worse than for H-TESSEL at Lobith and only marginally better at Basel. Given these results, would the conclusion not be that the MOS methods are not a viable option for operational streamflow forecasting? This conclusion is missing on Page 17.*

Yes, we agree. With respect to the Rhine at Lobith and Basel, the runoff simulations from ECMWF's S4 in combination with a linear bias correction score the same skill as the tested MOS approach does. This needs to be more stressed in the summary as well as in the conclusion. However, we think it is not valid to conclude in general that MOS methods are not a viable option for operational streamflow forecasting. Neither can we speak for the whole MOS family nor for other input data nor other parts of the world. Rather, we argue that MOS in general is part of the "seasonal streamflow forecasting"-toolbox – with all the disadvantages and advantages of a black box model.

My final concern with this paper is that it is not clear how the skill of the various models at zero lead time is composed. Probably, the forecast skill for the first 5 days is higher than for the last 5 days of the 1-month lead time. By averaging over the entire first month this information is lost. It could very well be that the average positive skill for lead times 1-30 days is entirely due to the positive skill for the first few days. Moreover, this positive skill for the first few days may be a result of the persistence of weather patterns in the GCM, similar to that for a normal short range weather forecast.

Related to this is the ESP-revESP analysis. The paper states that there is no clear difference in skill between the ESP and revESP at zero lead time (Page 16-Line 2). But the zero lead time is actually an average for lead times of 1 to 30 days. A separate analysis for lead times of 5, 10, 15, etc days would probably reveal a cross-over from dominance of initial conditions (higher skill of ESP) for short lead times to dominance of meteorological forcing (higher skill of revESP) for longer lead times. But this cannot be seen in the monthly average. Therefore I encourage the authors to do an skill assessment at higher temporal resolution.

Again, we agree, this is a very interesting point. If desired, we can run the hindcast for lead times in five day steps. However, we would like to note two arguments to not do so: Basically, it would lead to a new study – our intention was to test a MOS approach with the focus on the spatial scale, not to provide a detailed ESP-revESP analysis for the Rhine basin. The ESP-revESP analysis is used to aid interpretation, and to compare the results with the already existing literature (which is on a monthly basis, too). In addition, we believe that such an experiment mainly tests the assumption of linearity. The present MOS formulation might work to forecast average streamflow of three or even two weeks, but for five days very likely will fail. Therefore, it would be tricky to separate the effects of model misspecification and hydrological persistence.

Detailed comments

Printer-friendly version

Discussion paper

Page 3-Line 23: What approaches do these earlier studies use? Are these (bias-corrected) hydrologic model forecasting studies or do they use MOS/PP?

Yes, this are studies using the GCM - hydrological model chain. We agree, this must be specified.

Page 6-Line 2: I believe the reference for H-TESSEL should be Balsamo et al., 2008 or 2009.

Yes, we agree, but suggest to retain the reference to the ECMWF IFS documentation, since the information is taken from there (it also describes the actual H-TESSEL configuration).

Page 7-Line 6: Sample size is 31? 1981-2011 is 31 years, but you leave out the year of forecast and (according to Section 4.1.3) also the two preceding and subsequent years, so n must be 26.

Yes, we agree, the sample size equals n=26 for model fitting. We decided to write n=31 since at page 7, the reader only knows that the period amounts to 31 years, but does not know any details about the cross validation. If the article will be accepted for a revision, we will follow your suggestion.

Page 16-Line 14: It is found that the S4PT model outperforms the ESP model for subcatchments with smooth terrain and weak influence of initial conditions. Can you explain why? I would expect the opposite: the S4PT model (which includes forecast temperature) should do well for catchments that are dominated by snowmelt (rough terrain and strong influence of initial conditions).

We only can speculate: GCM skill for the Rhine basin is on a low level, and thus hard to detect. When the initial conditions are strongly relevant like in the case of a snow dominated catchment, any error in estimating these initial conditions produces larger errors than the GCM skill can reduce. Thus, we suggest that GCM skill is better detectable in catchments where the relevance of the initial conditions is small. For example, we would argue that it is hard to successfully force a hydrological model with seasonal climate predictions in a catchment situated in the Alps – if we get the snow pack wrong, the small skill contained in the precipitation and temperature forecasts vanishes.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-372, 2017.