Authors' responses to the comments of Dr. Hagemann

We would like to thank Dr. Hagemann for helpful and constructive comments. Below, we respond to the comments in a point-by-point manner. All revisions are visible in the supplement document ("Revised manuscript_visible changes.doc").

Major remarks

The authors present a robust analysis of a large ensemble of GCM-HM simulations to investigate the impact of internal variability on simulated river runoff. The study is interesting and worth publishing but a few things need to be addressed before.

1. It needs to be pointed out clearly that the considered time scales are important for the validity of results of the study. To separate the study from other research working on longer, climatological time scales, time scales longer than one year should be notably excluded, i.e. the impact of internal variability diminishes compared to other uncertainty sources if, e.g., multi-year monthly or annual means are considered (e.g. Déqué, M., D. Rowell, D. Lüthi, F. Giorgi, J.H. Christensen, B. Rockel, D. Jacob, E. Kjellstrom, M. de Castro and B. van den Hurk (2007) An intercomparison of regional climate models for Europe: assessing uncertainties in model projections. Climatic Change 81, Supplement 1, 53-70)

We in agree with this comment and are thankful to the reviewer for providing important references (also from the following comment) that are relevant to the subject of our study. These papers are now cited and the remark of the reviewer is now pointed out in the Introduction. We note, however, that although our study mainly presents uncertainties of the seasonal cycle component statistics, uncertainties of long-term climatic trends in annual component are also considered (in particular, in Fig. 12, where 34-yr trend distribution is shown) and may be relevant also for near-term climate predictions.

2. Studies such as Deque et al. (2007) or (Hagemann, S., H. Göttel, D. Jacob, P. Lorenz and E. Roeckner, 2009: Improved regional scale processes reflected in projected hydrological changes over large European catchments. Climate Dynamics 32 (6), doi: 10.1007/s00382-008-0403-9: 767-781) considering uncertainty introduced by internal variability at longer time scales should also be referred to in the introduction section.

These papers are now cited.

3. In the conclusions section it would be interesting to address the following question based on the results: What are the implications for seasonal to decadal predictions using GCMs?

Our general implication for GCM prediction is basically in line with conclusions by Deser et al. (2012; 2014) who indicate an importance of large ensembles of climate model realizations. This is now added in the Conclusions section.

> Our results, in line with the conclusions of Deser et al. (2012; 2014) who analyzed temperature and precipitation changes, suggest an importance of performing large ensembles of climate change projections with climate models also for making robust estimates of uncertainty and externally forced signal in hydrological response on decadal to multi-decadal time scale.

4. Technically I recommend a careful checking regarding the use/non-use of 'a' and 'the' in the manuscript. These seem to be missing at many places.

We have revised the manuscript regarding the use of the articles.

Minor Comments
In the following suggestions for editorial corrections are marked in *Italic*.

p. 2306 – line 25
… mean *value, which* indicates …

Corrected

p. 2306 – line 26
It is written: "…a considerable portion of the observed trend can be externally driven."
As you only deal with simulations I would not recommend using the word "observed" in this context.

We might have not properly formulated this sentence. The simulated ensemble mean Lena River discharge is statistically different from zero (when estimated from model ensemble spread) and fits well to the observed value. This allows us to suggest that observed trend has a contribution from external to the atmosphere forcing (SST, sea ice). This is not the case for Northern Dvina River. Now, this sentence is corrected both in the Abstract and Conclusions.

p. 2311 – line 25
In *Section 5, runoff* characteristics …

Corrected

p. 2313 – line 14
… Geophysics*; Motovilov et al. 1999a) has* been…

Corrected

p. 2313 – line 23-24
(SWAP*; Gusev and Nasonova 1998) has* been …

Corrected

p. 2314 – line 10-11
It is written: "Some key-parameters of the models are calibrated against streamflow measurements and …"
Some more information on the calibration and the respective parameters is desirable.

Some information about the parameter calibration procedure was added at page 8 of the revised manuscript. As to the calibrated parameters, we believe that including the list of the respective parameters is not too usefull without description of both models (that is unreal within the framework of the manuscript). The issues concerning the choice and justification of calibrated parameters can be found in (Motovilov and Gelfan 2013; Gusev et. al, 2011; ; Krylenko et al., 2014, Gusev et al., 2015). All these publications are cited in the revised manuscript

p. 2314 – line 25-27
It is written: "In particular, ECHAM5 similar to majority of climate models (Flato et al., 2013; IPCC AR5) simulates colder climate in winter in high latitudes of the Northern Hemisphere …"

I doubt this statement. Hagemann et al. (2006, 2013) show a distinct warm bias of ECHAM5 (AMIP simulation, but also coupled to an ocean model) in the winter over the high northern latitudes land area (or the area covered by the six largest Arctic rivers).
References: Hagemann. S., K. Arpe and E. Roeckner, 2006: Evaluation of the hydrological cycle in the ECHAM5 model. J. Climate, 19, 3810-3827 Hagemann, S., A. Loew, A. Andersson, 2013: Combined evaluation of MPI-ESM land surface water and energy fluxes. J. Adv. Model. Earth Syst., 5: 259-286, doi:10.1029/2012MS000173.

We agree that cold bias found in our simulations (see the Figures 1 and b for SAT and SLP biases) may not be a characteristic feature for the ECHAM5 model. Different biases may result, in particular, from different setup (e.g., we employ the old cloud scheme (Roeckner et al., 1996, ECHAM4 description), not the statistical-dynamical approach based on Tompkins (2002)). Also, coupled and uncoupled results may considerably differ. We, therefore, reformulated the indicated statement in a more generalized manner with citing the outlined papers.

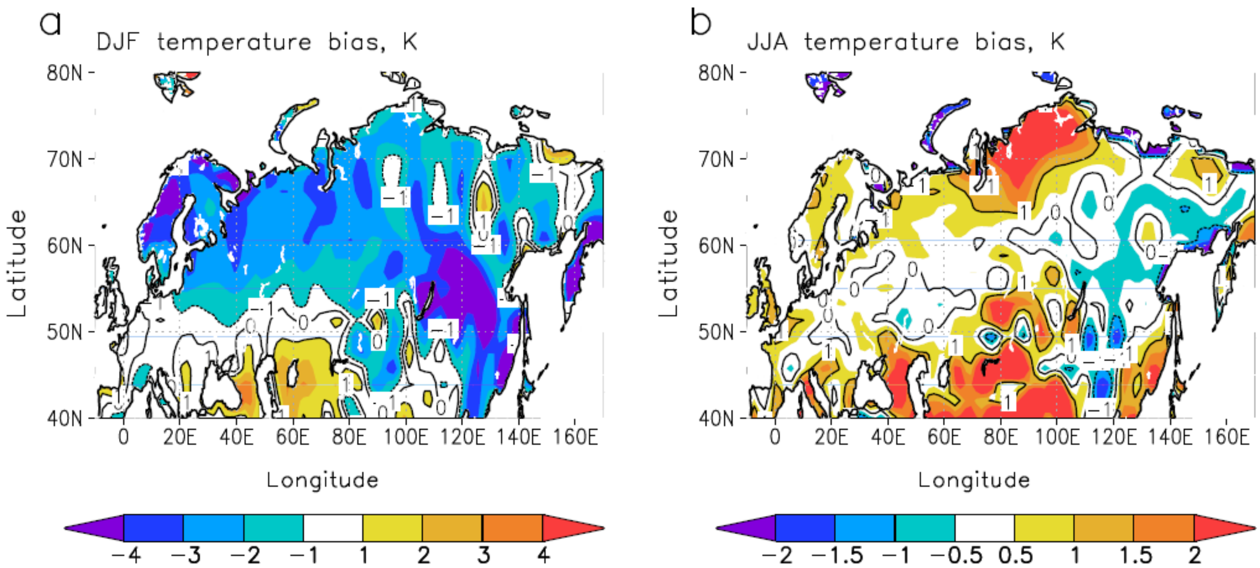

Figure 1. DJF (a) and JJA (b) surface air temperature difference between ECHAM5 (ensemble mean) and NCEP reanalysis, averaged for 1979-2012 period, K.

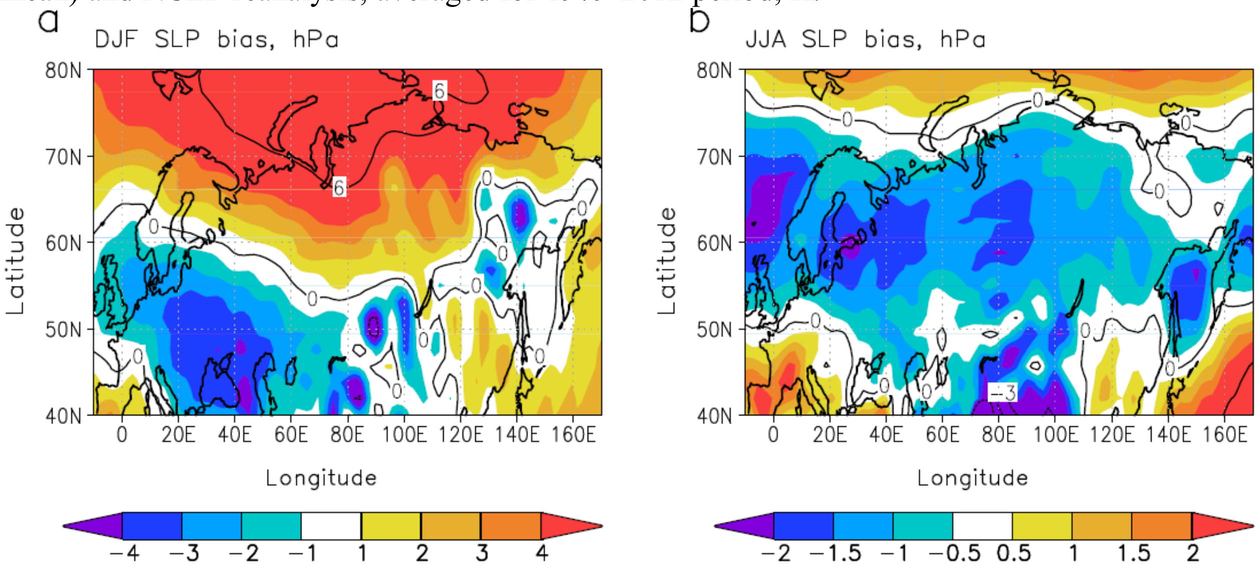

Figure 2. As in Fig. 1 but for sea level pressure, hPa.

p. 2317 – line 10-11

It is written: "One can see from this Figure that the applied post-processing allowed us to obtain rather similar fields of the above listed variables.."

The similarity between the model data and observations used for corrections is rather trivial as this can be expected from a bias correction approach. It would be of interest to show the uncorrected (original) fields in addition to see how large the correction actually is.

We do not fully agree that the similarity is always trivial result of the bias correction procedure. This result was, on the contrary, rather surprising for us, because the assigned correction factor equals difference between the model data and observations averaged spatially (over the very large basin) and temporally (over large time interval). In this case, similarity between the areal averages is trivial indeed, however similarity of spatial patterns is not. We have clarified this issue in the corresponding paragraph of the revised manuscript

> In addition, the similarity is rather surprising taking into account that the assigned correction factor is based on the model-observation differences averaged over the very large basin. Thus ECHAM5 demonstrates good performance in simulating spatial distribution of deviations from the basin averaged values of precipitation, air temperature and humidity.

p. 2325 – line 12

It is written: "… which is particularly noticeable for the winter season, when the SD-estimates are sometimes lower by hundreds percent in comparison with their observed variability."

Maybe it should be noted that discharges in winter are usually small for high latitude rivers so that even absolute small differences may yield large relative differences.

This is now indicated in the text.

p. 2327 – line 7-8

It is written:

"Importantly, the role of the internal atmospheric variability is most visible for the time scales from years to first decades …"

This is only true if one does not consider multi-annual monthly or annual means. See major remark [1].

We now added "in the presented simulations" to this sentence, which makes it clear that we do not imply multi-annual climatic averages.

p. 2328 – line 2 … runoff *trend, were* estimated.

Corrected

p. 2338 – Fig. 2 The top left panel is a duplicate of the top right panel. I assume, it should show temperature, not precipitation.

Sorry for misprinting. The figure is corrected.

p. 2343 – Fig. 7 Panels for the same river should be merged to allow an easier comparison between the two models. If this is not feasible, please use at least the same y-axis scaling for panels belonging to the same river.

p. 2346 – Fig. 10 Panels for the same river should be merged to allow an easier comparison between the two models. If this is not feasible, please use at least the same y-axis scaling for panels belonging to the same river.

According to your suggestion, we use the same y-axis for panels belonging to the same river

p. 2340 – Fig. 4 Instead of showing one curve per panel, the panels for the same river should be merged to allow an easier comparison between the two models.
p. 2344 – Fig. 8 Instead of showing one curve per panel, the panels for the same river should be merged to allow an easier comparison between the two models.
p. 2342 – Fig. 6 Panels for the same river should be merged to allow an easier comparison between the two models
p. 2345 – Fig. 9 Panels for the same river should be merged to allow an easier comparison between the two models.

We prefer keeping the listed Figures as are because of two reasons
1. The comparison between two models is not the purpose of our study. The models require different input data, are differently structured and parametrized, so their responses to the internal atmospheric variability can be incomparable. This is the case for Figures 4 and 8.
2. Figures 6, 9 are not readable if the corresponding panels would be merged.