

## ***Interactive comment on “The regional MiKlip decadal forecast ensemble for Europe” by S. Mieruch et al.***

**S. Mieruch et al.**

sebastian.mieruch@kit.edu

Received and published: 21 March 2014

### **Review answer 3**

#### **The regional MiKlip decadal forecast ensemble for Europe**

**S. Mieruch, H. Feldmann, G. Schädler, C.-J. Lenz, S. Kothe, and C. Kottmeier**

referee comments in **red**, author reply in **black**

C2849

Main issues:

In section 2 the authors describe the experimental setup. The description is generally accurate and provides enough information to allow other people to redo the experiment. The only exception is the procedure for the generation of the regional ensemble which allegedly represent one of the key point of this paper.

We will include some more explanation in Sect. 2 on the ensemble generation to make it clearer.

When on line 15 of page 5714 the authors mentioned a 1-day lagged initialisation I assumed they were referring to the global ensemble rather than the regional one. This is not entirely clear from the text.

Right. We will include a few words to make it clear.

In the following lines the authors state “For the first regional ensemble a larger ensemble size was preferred to a higher number of starting dates” but it is not clear to me whether the size of the regional ensemble is the same as the size of the driving data or whether multiple regional simulations have been run for each driving condition. Being constrained by the lateral boundary conditions (dictated by the GCM) and by soil conditions on the other I am wondering which degree of freedom would a regional simulation have.

On page 5715 it is said that “Therefore, all 10 available realizations of the MPI-ESM-LR for 5 starting dates (decadal1960-decadal2000) were downscaled covering the whole 50 yr period.” To make it clearer, we will explain that we are only regarding to one RCM, the CCLM here. Further, we will mention explicitly that we used only the 5 starting dates (1961,1971,1981,1991,2001) rather than the 10 starting dates (every 5 years) of

C2850

the MPI-ESM-LR. This is the compromise of having a larger ensemble (10 members), but fewer starting dates.

Soil is only initialised and then interacts freely with the atmosphere. Also, due to the size of the domain, a substantial number of degrees of freedom remain. The better representation of geographical features within the RCM domain, like topography, land cover, or land-sea contrasts, modifies the meteorological pattern depending region season and meteorological parameter (c.f. Feser et al. (2011)). In addition, a better resolution of smaller scale processes - like convective precipitation creates degrees of freedom from the downscaling (c.f. Feldmann et al. (2008)).

Also in section 2 the authors describe the procedure they followed to develop a baseline regional simulation. A spin-up period of 2 years is explicitly mentioned. Given that this was felt to be important there I was wondering why there is no mention of a spin-up in the downscaling of the predicitions. While the similarity of the models used for the large scale and for the high-resolution could probably reduce the model drift is not obvious to me that this can be discounted completely.

All initial conditions except the soil are obtained from the global forcing, as the regional scale atmospheric stucture emerge very fast. GCM and RCM use different soil vegetation atmosphere transfer models, therefore the soil conditions cannot be directly obtained. The long term memory of the soil errors in the initial fields would introduce artificial fluxes and a potential model drift (c.f. Khodayar et al. (2014)). One strategy to achieve a model soil which is in equilibrium with the atmospheric fluxes is to use a sufficient long spin-up period. This method has been only used for the re-analysis driven simulations. The initial soil conditions for the hindcast ensemble are then taken from the re-analysis driven simulation, which had a sufficiently long spin-up. Therefore, no additional spin-up is needed for the hindcast simulations.

C2851

The authors use two concepts to analyse the performance of the decadal system with respect to observations: fidelity and reliability. Following DelSole and Shukla (2010) the authors define "fidelity" as a measure of the agreement between model and observational climatological distributions. The second concept the authors use is the one of reliability. This is defined as following Weigel et al. (2009) as a measure of "how consistent the forecast probabilities are with the relative frequencies of the observed outcomes". An alternative definition of reliability could be based on the inability from an observed to distinguish the observations from the model output on the basis of their statistical characteristics (e.g. Joliffe and Stephenso's book on forecasts verification). In that sense there appears to be a direct relationship between reliability and fidelity. This seems to be contradicted by the plots (e.g. fig. 5 and fig. 6) which show high fidelity in region of poor reliability and vice versa.

There is no contradiction between fidelity and reliability (Figs. 5 and 6). Both metrics measure different aspects of the systems under consideration, thus different outcomes are possible. The main difference is that the reliability includes the RMSE, thus it measures how well the model outcome (hindcast) and the observations agree at **every point in time**. The fidelity, on the other hand, is a much weaker metric, because it only compares the distributions **independently of their temporal point by point agreement**. Thus it is no conflict to observe high fidelity and poor reliability. The reason could be that the climatological distributions are quite similar (good fidelity), but the temporal point by point agreement is poor (bad reliability).

Is fidelity only referring to the relationship between each ensemble member and the observation? Even in this case I am not sure I do understand how and ensemble of high-fidelity members can lead to unreliable predictions. I think the paper could benefit from a bit more information on these skill metrics.

Yes, the fidelity refers to the agreement between each single ensemble member and the observations. The answer to the second question is given in our explanation from

C2852

the previous answer. However, we decided to remove the analysis of the fidelity from the paper, because it gives not much information. Generally the fidelity is good for all regions, all seasons, all variables and all filtering. Thus, the only conclusion which can be drawn is that the climatological distributions of the model and the observations are not significantly different.

Similarly it would be useful to explain why following the approach of Weigel et al. (2009) in the calculation of the reliability index makes sense for variable which are not normally distributed such as rainfall. While the central limit theorem is mentioned in the conclusion, it may be a good idea to add a reference to it in section 3. Also, while the assumption of normality may be reasonable for most of Europe in can potentially be challenged for summer precipitation on the southernmost part of the domain do to the limited amount of precipitation there in this season. A brief discussion on this point could also be beneficial.

We will add the information on the central limit theorem in Sect. 3 to the reliability subsection. We will look into the rainfall distributions in southern Europe and if necessary add a brief discussion.

Minor comments:

Minor comments: The statement on line 21 page 5713 “For practical applications, the information provided by global models is much too coarse” appears to be very generic and whilst probably correct for the vast majority of users express it in such general terms can be misleading. It is known that for some specific applications is more important to have skill in the large scale field than to have high resolution in the outputs (e.g. international food market).

Right, we will change the sentence in the revised version. We will substitute “practical”

C2853

by “regional scale”.

The statement on user needs on line 29 while reasonable in his formulation is not rooted in any evidence and should be better substantiated through appropriate references.

We will cite Dool (2007) and Berg et al. (2013).

Lines 7-8 on page 5719 make little sense to me.

Here we tried to express the invariance to location and scale of the correlation coefficient in prose language. What we mean is, that two time series can be well correlated, even if the values itself have e.g. a large bias.

## References

- Berg, P., Wagner, S., Kunstmann, H., and Schädler, G.: High resolution regional climate model simulations for Germany: part I—validation, *Climate Dynamics*, 40, 401–414, 2013.
- Dool, H. V. D.: Empirical methods in short-term climate prediction, Oxford Univ. Pr., 2007.
- Feldmann, H., Früh, B., Schädler, G., Panitz, H.-J., Keuler, K., Jacob, D., and Lorenz, P.: Evaluation of the precipitation for South-western Germany from high resolution simulations with regional climate models, *Meteorologische Zeitschrift*, 17, 455–465, <http://dx.doi.org/10.1127/0941-2948/2008/0295>, 2008.
- Feser, F., Rockel, B., von Storch, H., Winterfeldt, J., and Zahn, M.: Regional climate models add value to global model data: a review and selected examples, *Bull. Amer. Meteor. Soc.*, 92, 1181–1192, 2011.
- Khodayar, S., Sehlinger, A., and Feldmann, H.: Sensitivity of soil moisture initialization for decadal predictions under different regional climatic conditions in Europe, *Int. J. Climatol.*, submitted, 2014.

C2854

C2855