Geosci. Model Dev. Discuss., 6, C2533–C2536, 2014
www.geosci-model-dev-discuss.net/6/C2533/2014/

Geoscientific Model Development Discussions
Open Access

GMDD
6, C2533–C2536, 2014

Interactive Comment

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

Anonymous Referee #3

Received and published: 10 February 2014

The paper provides an interesting analysis of the decadal predictability in Europe.

While a number of other authors have looked at similar issues in the past the authors of this paper conducted the analysis with a different set of models, explored the value of downscaling and analysed the impacts of the use of a long-pass filter on the results.

The paper is generally of a good quality and would certainly be of interest to the science community. TSaid that, tere are a number of aspects of the paper that need to be improved.

Major comments:

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 alegedly represent one of the key point of this paper.

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.

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.

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 predicitons. 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.

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 Stephenson'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.

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.

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 idead 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.

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 whist 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).

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.

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

GMDD

6, C2533–C2536, 2014

Interactive comment on Geosci. Model Dev. Discuss., 6, 5711, 2013.

GMDD

6, C2533–C2536, 2014