Nonlin. Processes Geophys. Discuss., 2, C554–C556, 2015
www.nonlin-processes-geophys-discuss.net/2/C554/2015/
© Author(s) 2015. This work is distributed under
the Creative Commons Attribute 3.0 License.

Nonlinear Processes in Geophysics
Discussions
Open Access

NPGD
2, C554–C556, 2015

Interactive Comment

Interactive comment on "Predicting climate extremes – a complex network approach" *by* M. Weimer et al.

Anonymous Referee #2

Received and published: 25 November 2015

In this paper, the authors construct evolving climate networks from the output of a regional climate network (CCLM) and suggest to use the network correlation threshold as an indicator of heat periods in 8 different regions in Europe.

The presentation is clear and concise, and the paper is quite well written apart from some typos. The figures are informative, although the captions should be extended in order to provide more information.

I do, however, have strong concerns with the general approach. If these issues can't be resolved, I would be reluctant to recommend this article for publication.

1. General Comments:

The title and abstract are in my opinion misleading. The approach presented here does not aim to improve the prediction of climate extremes. Instead, the authors propose a different way to quantify the skill of an initialized regional climate model to reproduce observed numbers of heat periods. The standard way to define such a skill would be to compare the numbers of observed and modeled heat periods. As an alternative way, the authors propose to compare the (suitably normalized) number of observed heat periods with the correlation threshold used to construct climate networks from the model data. Their result is that in some regions and for some decades, the model skill quantified using the correlation threshold is better than the standard one, and sometimes it is not. It should be emphasized here that the result does not imply any improvements of the model's predictions, but only shows that a different and less direct skill definition (based in the correlation threshold instead of directly comparing the numbers of heat periods) leads to different results, with sometimes higher skill. Furthermore, the result is not really quantified (see Fig.7 which compares the two skill definitions in terms of better/worse/tie), and no tests of statistical significance of this result is performed (in the sense of: "what is the probability of getting a rank matrix like the one in Fig.7 from some suitable null hypotheses"). See also point 4. below.

2. Specific Comments:

1. For all networks, a constant edge density of 0.3 is enforced, implying that the network correlation threshold is simply the 70th percentile of all correlation values for a given region and time span (P70). Hence, i think that the network terminology is quite overdrawn in this work, since the only "measure" employed here is this P70. The wording "link strength" is furthermore unfortunate in this context, since strength also refers to the sum link weights of a node in a weighted network.

2. The authors first show that in artificial (Gaussian) time series, P70 increases linearly with the number of heat periods included in the time series (Fig.3). Furthermore, based on the observations (E-OBS), they show that there is some dependance between the number of heat periods and P70 (Fig.4). The authors infer from this that P70 is a good predictor of the number of heat periods. I don't agree to this statement: In the case

NPGD
2, C554–C556, 2015

of artificial time series, you trivially increase their correlations by putting in additional heat periods, because the time series become “more similar” by construction. P70 can be interpreted as an estimator of the overall correlations between given time series at hand, and hence their coherence. If a region experiences a spatially extended heat period (more than 20% of the region with temperatures above a threshold, as you defined it), it is to be expected that the corresponding time series behave more coherently, as expressed in higher P70. However, any other spatially coherent behavior, such as a cold period, would lead to higher values of P70 as well. P70 measures nothing but the “spatiotemporal” coherence, and heat periods can be examples of this, but many other phenomena could be just as well.

3. I agree with you that comparing the observed number of heat periods with the modeled number of heat periods suffers from the “threshold problem” that the model amplitude might for a given observed heat period just closely not exceed the threshold. However, I can think of some simple ways to overcome this: In particular if an ensemble of projections is available, on could define that a heat period is detected in the model runs if some percentage of the ensemble members exceeds the temperature threshold. I am not convinced that taking P70 instead, which is ambiguous in this context for the reasons explained above, is a good alternative. Instead of taking the modeled number of heat periods, you propose a much less precise “proxy” - namely P70 - for heat periods. Loosely speaking, this proxy could be seen as a “randomization” around the modeled number of heat periods, and the results shown in Fig.7 could hence simply reflect this randomization: Sometimes the model skill is better according to your skill definition than the standard skill, and some times vice versa.

Interactive comment on Nonlin. Processes Geophys. Discuss., 2, 1481, 2015.

NPGD

2, C554–C556, 2015

CC BY