Answer to comment of referee #1

Predicting climate extremes - a complex network approach

M. Weimer, S. Mieruch, G. Schädler and C. Kottmeier

Dear referee,

Thank you for your detailed review of the paper. In the following, you can find our answers to your comments which are written in red text color.

1 General comments

Is your main goal to promote a new heat wave detection method or is it an improved prediction of heat waves? [. . .] Please make your presentation clearer in this respect.

The point is that no new method for the detection of heat periods is needed, if past observational data is used. The WMO based definition works well and also lots of EVT (Extreme Value Theory) applications, like the POT (Peak Over Threshold) package written in the programming language R, have implemented this definition in a more or less sophisticated way. But, for detecting heat periods in initialised (every 10 years) forecast/hindcast data, new methods are needed. On the one hand to overcome threshold problems and on the other hand to consider climate (long-term) predictions in contrast to short-term forecasts.
Thus, our main goal is to improve the decadal prediction of heat periods by using a new method for determining them. We agree that we did not clearly present how to achieve this aim with this method and will adapt the paper by describing the method more precisely (which also was one of the specific comments below) and by pointing out how to forecast the number of heat periods with this method in the sense of decadal prediction.

Answer to the second general comment concerning the practical value of the network approach and the forecasting methodology

We agree with the referee that this point is not clearly represented in the paper. However, the number of heat periods and the link strength shown in the figures are the absolute values and no rescaling has been performed.
As a first approach, we would suggest to assume a simple linear relation between the number of heat periods and the link strength. Regarding Fig. 4, a link strength of e.g. 0.89 would correspond to a number of heat periods of about 2.25.
However, the question is how to actually forecast real future heat periods, where no observational reference is available. To achieve this, a standard approach known as recalibration is applied. This is also done for short-term prediction, where e.g. the model bias is corrected by using past hindcast data, which has been compared with observations. In our case we would simply have to

linearly convert the link strength to the number of heat periods using the conversion factors from the past hindcasts and apply them to the real future predictions.
Assuming $\hat{m}^r_{d,\tau,W}$ to be the number of future heat periods calculated from the future link strength $\hat{W}^r_{d,\tau}$ at time τ and $W^r_{d,\tau}$ and $o^r_{d,\tau}$ the past link strength in the model and number of heat periods in EOBS by standard approach, the following formula is applied to convert link strength to number of heat periods:

$$\hat{m}^r_{d,\tau,W} = \frac{\hat{W}^r_{d,\tau} - \min^{50}_{\tau=1}(W^r_{d,\tau})}{\max^{50}_{\tau=1}(W^r_{d,\tau}) - \min^{50}_{\tau=1}(W^r_{d,\tau})} \cdot \left(\max_{\tau=1}^{50}(o^r_{d,\tau}) - \min_{\tau=1}^{50}(o^r_{d,\tau})\right) + \min_{\tau=1}^{50}(o^r_{d,\tau}) \quad (1)$$

2 Specific comments

I suggest you make the title more specific

We totally agree with that comment. As we think that the issue of decadal prediction should appear in the title as well as the difference to Ludescher et al. (2013) we suggest the following title for the paper:
"Decadal prediction of heat periods based on regional climate model data - a complex network approach"

p1482l10: Please briefly explain the "standard approach"

p1482l10: The way you word it I get the impression that nowhere the network approach works worse than the standard approach while, in fact, Fig. 7 shows that this is not the case; so please reword

We change page 1482, line 8: “We show that the skill of the network measure to predict the low frequency dynamics of heat periods is similar to the one of the standard approach, with the potential of being even better in some regions.” to “We show that the skill of the network measure to predict the low frequency dynamics of heat periods is superior to that of using a fixed temperature threshold for estimating the number of heat periods in some European regions.”

p1482l22: "have been also followed" $\rightarrow$ have followed?

Yes, that is right. We will change this in the revised version.

p1483l5ff: The abbreviations should be in brackets and the explanation in front of those brackets (here and everywhere else in the manuscript).

We will change this.

p1483l23: Do you mean on a time scale/lead time of five years

Instead of "in the order of five years" $\rightarrow$ "within the lead time of five years"

p1484l1: "On the other hand": What is on the one hand?

We change “On the other hand an innovative approach in climate research has been established in the recent years, namely the complex climate network approach. ” to “Another relative new field in climate research has been established, namely the complex climate network approach.”

p1484l19ff: Think about where in the manuscript you want to explain why you think your network approach to heat wave detection works. Explain it well in this one place and refer to this explanation from other places if necessary. Right now, you address the explanation in several places scattered across the manuscript, in part just repeating things, in part adding new aspects; that makes it hard to comprehend.

We totally agree with that comment and will modify the Methods’ section to improve its presentation at this point.

p1486l8-10: Explain the standard approach better please: standard approach to doing what? Are you talking about detecting heat periods or about evaluating a model w.r.t. its ability to detect heat periods?

It is the standard approach for determining the number of heat periods in time series depending on a fixed temperature threshold (see WMO definition). Actually, we here also account for the spatial extent of a heat period by defining it as a temperature threshold of 3 K which has to be exceeded at not less than 20 % of the grid points on five consecutive days at least. We will adapt the explanation in the revised version.

p1486l13: "the correlation threshold between time series": I understand what you mean, but you can write this better; please rephrase.

We will replace this sentence by the following: "Generally, networks of gridded data are constructed by correlations between the time series at the grid points and a threshold in the correlation coefficient which is called link strength W_τ hereafter. A constant edge density for all the networks, i.e. a constant number of connections in the networks, leads to a temporally varying W_τ which will be used as an alternative estimator for the number of heat periods in this study."

And we will discuss the properties of the method in a more detailed way at this point.

p1486l17: Make clear here an important difference between Ludescher and colleagues’ (2013) approach and yours: They make predictions based on observational data alone, i.e. they really use the NETWORK to forecast something while you actually use a CLIMATE MODEL to forecast something and then analyze those forecasts using networks

We totally agree with this comment and will change this in the revised version.

p1486l20ff: This sounds as if the main reason for a model to under- or overestimate the number of heat waves is a bad simulation of the magnitude of heat waves. Yet later (1488l9-19) you detect simulated heat waves based on the daily maximum temperature percentile corresponding to a 3 K anomaly in the observations, which implies that you correct for a possible bias in heat wave magnitudes prior to detecting simulated heat waves. Thus, your main argument as to why the network approach can be expected to work better than the standard approach (to detecting simulated heat waves) does not apply to the application case presented below…

First, "‘yes:"’ we correct for a model bias for the whole region and whole time (1961-2010). Second, no our argument remains valid. The point is that due to the bias correction we maximise the possibilities to detect heat periods using the standard approach. Nevertheless, we still have the problem of a fixed, static threshold, although this threshold is possibly optimal. Our approach is independent of any such static threshold and thus has the potential to work better.
We will make this more clear in the revised version.

p1487l5: networks → networks’

We will change this in the revised version.

1487l7-14: You will get the same relationship between the number of events and the mean correlation if you don’t include heat but cold waves in those Gaussian noise time series. So, how can you be sure that when applied to real data, your method detects heat waves and not also cold waves or some other form of spatially and temporally extended temperature anomalies?

This is a very good comment and we actually did not really think about this point previously. In a first attempt, we stress that we use the daily maximum temperature in summer months. So, it is rather unlikely that cold waves occur in the data. Additionally, figures 4 to 6 in the paper show that the method works which obviously means that more heat waves occur than cold waves. However, it is true that cold waves could lead to uncertainties in our heat period estimator.

We think, there are two possibilities to clarify this question and we will try both ones:

1. determining the number of cold waves by reduce the temperature threshold to -3 K instead of $+3$ K:
   If the detected number of cold waves is significantly lower than that of heat periods we could show that the influence of cold waves is negligible

2. leave the 10 % percentile of the bias-corrected data and then detect heat periods with the same method and calculate link strength accordingly

p1487l29: I am pretty sure you don't want to remove the variabilities from the time series.

Yes, we correct for long-term bias. We will rephrase the sentence in the revised version.

p1488l19: Why did you expect this?

We decided to omit this sentence in the revised version.

p1488l22: I assume you mean "simple graph"

Yes, this is right.

p1489l8 "acceptable results": What do you mean by that?

We want to register the dynamics, i.e. the variability, of the network which means that the correlation threshold should not be too high and not too low. Too high thresholds would result in a network without connections (extreme case). Too low thresholds in turn would lead to a network where each node is connected to each other. No statistical analysis can be inferred from these two kinds of network.

So, a correlation threshold in between these two extremes has to be applied. It is clear that a kind of subjectiveness lies in the correlation threshold but we tried several ones and correlation thresholds of 0.7 - 0.9 showed the most variable dynamics, which we think is suitable for our purpose.

p1489l20: You construct one network per JJAS season with roughly 120 days. That are pretty short time series. What about the uncertainty of your estimated correlation coefficients?

For uncorrelated data of length $N = 120$ a correlation coefficient of $r \geq 0.18$ is statistically significant on the 5 % level. Of course daily maximum temperature data are definitely autocorrelated for several days, thus the $r \geq 0.18$ threshold is too optimistic. Nevertheless we use very strong correlations in the order of 0.7 to 0.9, which are definitively significant.

p1489l21: You give an uncertainty here but talk about a constant edge density in the same breath. Please clarify.

Previously, we implemented an iterative adaption method for the link strength to determine the edge density which worked as follows: Beginning at a specific link strength the edge density was calculated. The difference between current edge density and goal edge density was then converted to a difference in the link strength because they are anti-correlated in a non-linear manner. When getting closer to the goal edge density these differences got lower which meant that the algorithm was very slow near the goal edge density. That is why we decided to break the iteration at an edge density uncertainty of ± 0.0005.

However, according to a comment of the other referee we will implement the edge density 0.3 as 70 % percentile in the correlations between the time series. This will also be a much more efficient determination of the link strength at constant edge density. The uncertainty will be removed accordingly.

p1490l9: The filter is applied to which data?? To W_τ or to the temperature data? And why?? I don't understand.

The filter is applied at the very end before plotting to both the W_τ and the number of heat periods (either m or o). We will explain that better in the revised version. We are using *decadal* prediction data which means that we do not expect the model to predict the exact number of heat periods of a year but the general trend in the number of heat periods within the decade. The same argumentation is valid for the link strength.

p1490l11: You apply a 10-year moving averige filter to 10 years of data. What do you do at the ends of the time series?

At the edges of the decades, the time range for the running mean is shortened: At the beginning of the decade we use only the 6 years' mean, in the second year 7, and so on. We will add this information to the revised version.

p1490l13: What is the "absolute mean difference"?

We will refer to the equations in the text.

p1490l14: "heat periods o": Do you mean the number of heat periods?

Yes, we do and we will add this in the new version of the manuscript.

p1490l18: Please use different symbols for (or somehow indicate the difference between) the original and the rescaled m, and similarly for o and W_τ .

We will change this in the revised version.

p1490l21: Is this the usual way the simulated number of heat waves is evaluated (using the rescaled quantities)?? Does CCLM perform better with respect to the rescaled o's and m's than with respect to the unrescaled quantities?

In principle yes. However, normally, model data are recalibrated, which is a post-processing step. Based on past model data and past observations the bias of the mean of the model and possibly the bias of the variance of the model is corrected, i.e. rescaled to the observations. And yes, the model performs better applying the recalibration step.

p1490l23: So the rescaled o is compared to the rescaled W_τ here?

This will be clearer if the notation of rescaled and not rescaled quantities is different. But yes, these are the rescaled quantities here and we will clarify this in the revised version.

p1491l12: Well, the lines in Fig. 4 do not fully match, so the two estimators are obviously not "equivalent". Moreover, is Fig. 4 your "best" example, i.e., how does this figure look like in the other regions?

We will rephrase this sentence. We decided to add the figures of all Prudence regions as additional material to the paper if this possible at this stage, so you will be able to see that this is the best example, but others are good as well.

p1491l26: "based on normalized time series": Which time series were normalized how?

Add to the sentence: "according to Eq. 5".

p1492l5-7: It would be interesting to see where the error of the network approach to heat wave detection comes from in the different $(r; d)$ cases. There are two possibilities, as far as I can see:

1. the network approach to heat wave detection does simply not work
2. the network approach to heat wave detection does work but the CCLM simulations are bad

You can check which possibility applies to which case by comparing Mdr(W) with Wtau based on CCLM data (let me call this MCCLM) to Mdr(W) with Wtau based on E-OBS data (as in Fig. 4; let me call this MEOBS). If MCCLM is similar to MEOBS we have possibility 1 while if MCCLM MEOBS we have possibility 2, right?! Please do this analysis and discuss your findings.

First, we have to say that Fig. 4 shows that the network approach works well. Second, it is true that the decadal prediction skill is rather weak, but this is exactly the reason, why we have performed the study, i.e. improving the prediction of heat wave dynamics.
During our study we actually analysed what you called MCCLM and MEOBS. We will add this figure (similar to Fig. 4) and discuss the results.

3 Comments on the figures

Fig. 2: The difference between subplots a) and c) is not well visible. I'd recommend you make the difference more pronounced.

We will do that.

Fig. 4: Explain in the caption what the $M = \dots$ are.

We will do that.

Figs. 4-6: Why do the o time series jump between decades?

The running mean is applied to each decade separately. Therefore, the number of heat periods at the end of a decade is the mean of the 6 previous years whereas at the beginning it is the mean of the 6 years afterwards. In addition, CCLM is initialized at the beginning of the decade and runs freely. So, the model is close to the initialization in the first years unlike in the last years of the decades. That is why jumps could appear.

Fig. 5: Are these the original or the rescaled o's and m's? (See my comment on p1490l18; introduce new symbols for the rescaled quantities and use them where applicable.) Why do you use different scales for o and m?

We will adapt the figure caption with respect to this. The values are not rescaled.

We do not expect the model to resolve the exact number of heat periods. In our opinion, the dynamics of both EOBS and CCLM are different and thus different scales should be used for both of them.

Figs. 4, 6: Same question as for Fig. 5 (with W_τ in place of m)

W_τ is not rescaled. It is the link strength which ranges between 0 and 1 whereas the number of heat periods is in the order of 3 or so. In order to compare these quantities and since we think that the whole dynamic is registered within these five decades, we use different scales for the axes. As mentioned above, the conversion of link strength to number of heat periods is based on the different axes.

Fig. 7: Could you please make this comparison more quantitative, e.g., by writing the values of $M_d^r(m)$ and $M_d^r(W)$ into the respective $(r; d)$ boxes.

We will do that, at least by putting the differences between the M's in the boxes.