Answer to comment of referee #2

Compound extremes in a changing climate - a Markov Chain approach

K.Sedlmeier, 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

In order to address your first comments, we will introduce a new section in the revised version called "Sensitivity analysis" where we address the spatial and natural variability and analyze the error by means of Fourier-Transform surrogate time series. Detailed comments can be found below.

one should demonstrate that new descriptors reasonably reflect underlying physical mechanisms. Before using any new measure for characterization of ongoing and expected climate change, one should investigate their variability in natural conditions. The authors use the gridded E-OBS data set, however, they unfortunately chose just a few grid points in six different areas. It is a pity, since the E-OBS data set gives an excellent opportunity to study spatial variability of any descriptor which has an ambition to characterize the temporal evolution of a physical quantity attributed to each grid point. I think the model is reasonably simple to compute full coverage for Europe for all three descriptors and map them. The simple visual evaluation would indicate if the descriptors reasonably reflects physical reality in the case the maps show interpretable smoothly changing patterns. Or, if the maps show just a colored grains or a sort of Pollocks paintings, than there is a problem with the descriptor and its connections to physical reality.

We thank the referee for that comment and totally agree that new descriptors must be tested for revealing a connection to physical reality. Indeed, we did these tests prior to our analysis, which were also the basis for choosing the regions discussed in this paper. We have calculated a full coverage for the descriptors averaging over 3x3 grid points for the whole area and these maps show interpretable smoothly changing patterns as you can see in Fig. 1. This figure will be included and discussed in the revised version of the paper in the newly introduced section.

As to Pollock's painting: a map like a Pollock's painting might not be achieved easily for the Markov descriptors. Pollock's paintings are not random and not noise, rather they are in between determinism and noise, they are fractal (Taylor et al., 2007, , and citations therein). Thus, due to their fractal geometry they have deep underlying mechanisms in common with natural patterns and hence also with our atmospheric time series.

While E-OBS data set can be used to test spatial variability, ECA&D station data set offers a number of long-term records in which temporal variability can be tested. So one can relate the change of the introduced descriptor due to climate change to their changes due to natural variability in preindustrial era. Real long-term records would reflect natural variability due to natural nonstationarity.

This is a good suggestion. Unfortunately, there is only one station with a **continuous** (without missing values) temperature and precipitation record (starting in 1887) available from the ECA&D data set. Further, only a few stations within Germany have available **continuous** time series starting in 1900. Nevertheless we calculated the descriptors for a combined time series of the available 7 stations in Germany for running windows of 30 years starting in 1900. The combination of the time series is necessary in order to fulfill the stationarity criteria explained in Section 2.3. of our original manuscript (non zero entries of the transition probability matrix and stationarity of the time series). It is important to note that we removed all linear trends for each 30 year section seperately as it has been done in the rest of the paper. The resulting time series of the descriptors are shown in Fig. 2 for both winter (black) and summer (gray) extremes. These results will be included in the "Sensitivity analysis" section in the revised version of the paper. The stations used will be listed in the data section. Especially for the persistence and recurrence time, a clear shift is visible between 1930 and 1950. This time range is not preindustrial, but the crucial point is that the observed shift coincides with a globally oberved shift in the increase in CO2 around 1950 (http://www.ldeo.columbia.edu/~spk/Research/AnthropogenicCarbon/images/ddic_uptake_hist.png). Thus from this finding we observe two main points:

1. The descriptors (especially persistence and recurrence time) seem to be sensitive to changes of the CO_2 increase. That means a stronger increase of CO_2 (e.g. from 1950 on) yields to an decrease of the persistence and increase of the recurrence time. Again it is of utmost importance to note, that we removed the linear trends from each 30 year section of the temperature and EDI data.

2. Thus we can conclude that the natural variability can be approximated by the variability observed before and after the shift. This natural variability is smaller than the shift of the mean.

Concluding, due to the non-availability of preindustrial data we could not really test natural variability vs. natural nonstationarity. But we could show that natural variablity (before and after the shift in 1950) is smaller than the shift, which is probably due to the change in CO_2 increase. The mean level shift for the winter extremes of the persistence is about 50% (from 0.2 to 0.1) and for the recurrence time it is about 20% (from 180 to 140 days). Regarding Fig. 8 in the original manuscript we see that changes of the persistence above 50% have been observed (red and cyan regions) and changes of the recurrence time above 20% (red and green). Thus, according to the sensitivity tests natural variability can most probably be excluded as the sole cause for these changes. Interestingly our significance test also states that these changes are significant with very small p-values. These findings strongly support the results found in our study that changes of the succession of compound extremes are likely to occur in the future due to the increasing CO_2 emissions, whereas natural variability plays a minor role.

One can test numerical variability of the descriptors by constructing appropriate surrogate data. E.g., FT surrogate data generation averages dynamics over whole record randomized, so one can get ranges for random variability of the descriptors in a stationary data.

We have done this as part of our analysis and will now include the results in the revised version. To construct FT surrogates of our data, we used the MIAAFT algorithm (Venema et al., 2006) which in addition to preserving the original distribution of the data also preserves the auto and cross-correlation of the temperature and precipitation time series. 100 surrogate data sets for the 6 regions used throughout the paper were calculated for the E-Obs data set in the reference period (1971-2000) and their standard deviation taken as the error (by using the exact same regions the values are transferable to later chapter which would not be possible had we chosen a different number of data points). An overview of the errors can be seen in Tab. 1. The errors are fairly similar for all regions and do not differ largely between the two seasons. As in the original manuscript, we will keep on using the ensemble approach for estimating the uncertainty of the descriptors and their climate change signal, but will refer to these MIAAFT estimated errors when discussing the results throughout the paper.

	DJF			JJA		
	P	R	E	P	R	E
reg1	0.010	1.701	0.004	0.007	1.183	0.009
reg2	0.011	2.182	0.010	0.010	2.055	0.010
reg3	0.010	2.563	0.005	0.009	0.923	0.007
reg4	0.008	1.150	0.005	0.008	0.990	0.011
reg5	0.010	2.45	0.010	0.008	1.103	0.009
reg6	0.007	0.797	0.004	0.009	1.150	0.009

Table 1: Estimation of the error of the descriptors by using MIAAFT surrogates for winter (DJF) and summer(JJA) extremes. Values are calculated for the 6 regions of Fig. 1.

2 Technical corrections

P. 9, last para ....Fig. 4) all regions except Bulgaria... Should not it be France?

Yes thank you, it should be France

p 10, 4.2 The statistical treatment should be described in more details: Differences of the ensemble means are plotted, i.e. one get the mean and percentiles for each ensemble, then the difference of means is clearly defined, but what are the percentiles?

The climate change signal is calculated for each ensemble member separately. What is shown in the plot is the mean difference (gray bar), as well as the median and interquartile range (box) and the minimum/maximum difference (whiskers). We will add a more detailed description in the text so it becomes clearer.

Is this an appropriate way to evaluate the significance of changes?

The significance of the changes is determined by the ensemble approach and we think that this is an appropriate way of analyzing significance in this context. Furthermore the changes can be compared to the errors as derived by the MIAAFT algorithm (of the newly added Chapter). Changes are larger then the there derived errors which is an additional indicator of significance. We will mention this additionally in the text. Furthermore, as explained above, the significance test is in accordance to our sensitivity tests. These sensitivity tests have shown that changes of the persistence in the order of 50% (for recurrence time 20%) cannot be achieved by natural variability, but by a shift of the increase in CO_2 emissions. Similarly, the significance test states that changes of the persistence in the order of 50% (recurrence 20%) are significant.

References

Taylor, R. P., Guzman, R., Martin, T., Hall, G., Micolich, A., Jonas, D., Scannell, B., Fairbanks, M., and Marlow, C.: Authenticating Pollock paintings using fractal geometry, Pattern Recognition Letters, 28, 695–702, 2007.

Venema, V., Meyer, S., García, S. G., Kniffka, A., Simmer, C., Crewell, S., Löhnert, U., Trautmann, T., and Macke, A.: Surrogate cloud fields generated with the iterative amplitude adapted Fourier transform algorithm, Tellus A, 58, 104–120, 2006.

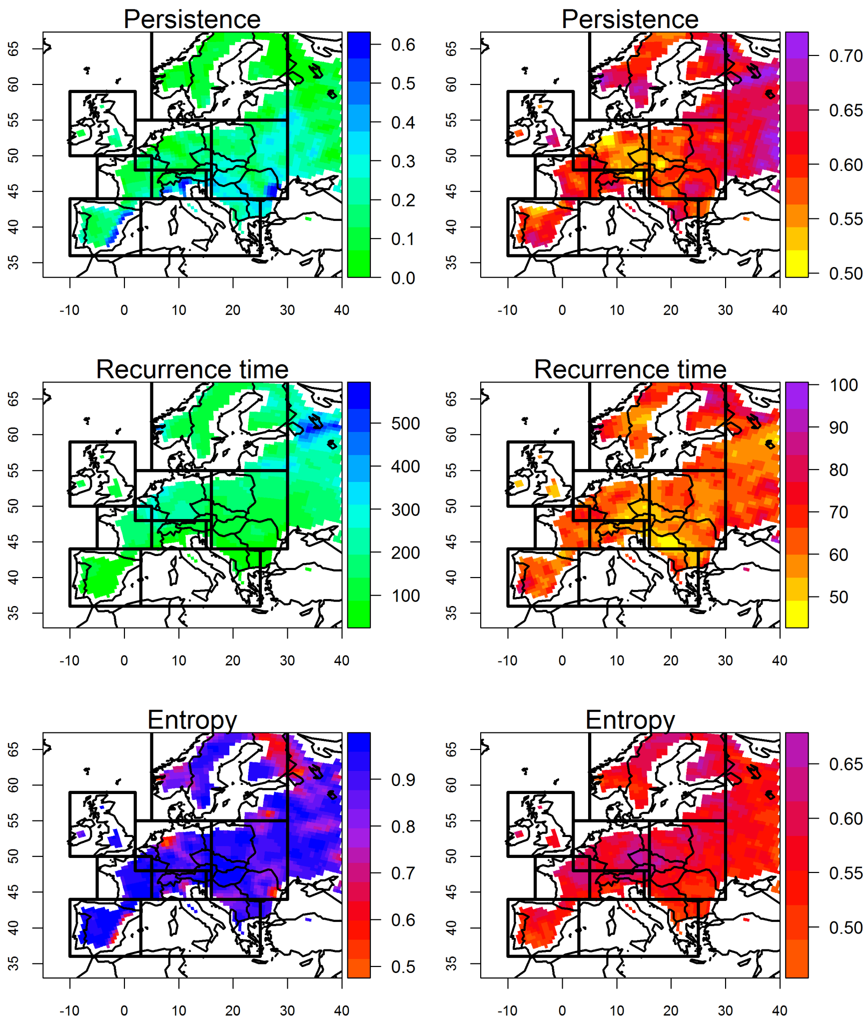

Figure 1: E-Obs descriptors for the reference period (1971-2000). Left side: Descriptors for cold and wet extremes in winter (DJF) (Ta < 10th percentile and Pa > 75th percentile). Right side: Descriptors for hot and dry extremes in Summer (JJA) (Ta > 95th percentile and EDI < 25th percentile). Descriptors were calculated for a moving window over 9 grid points and values assigned to the center grid point. Boxes show the Prudence Regions (http://ensemblesrt3.dmi.dk/quicklook/regions.html).

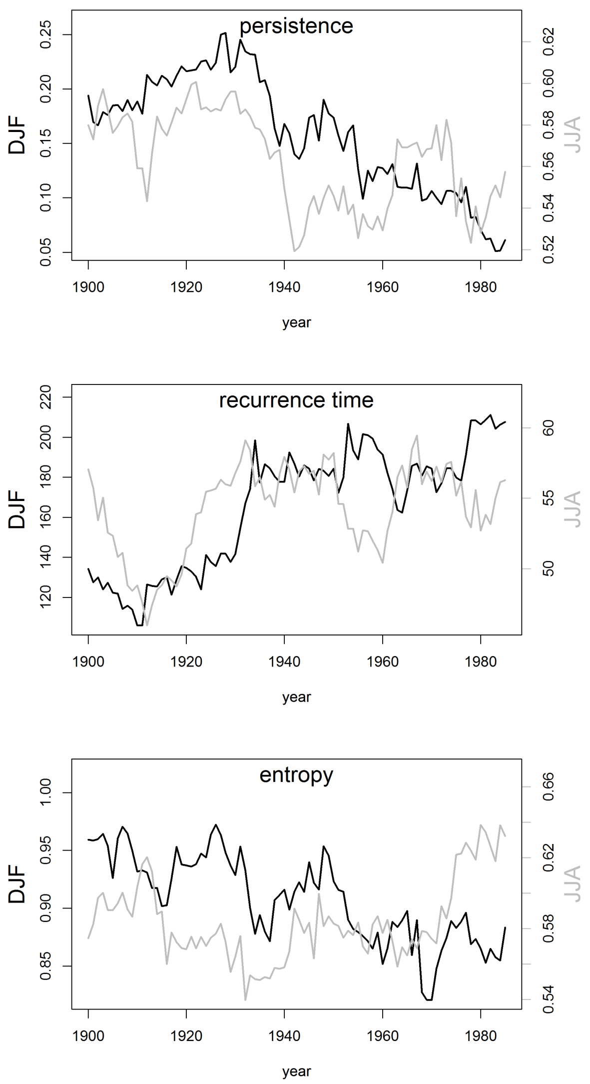

Figure 2: Descriptors for ECA& D station data for running windows over 30-years (values are assigned to the first year of the 30-year time period.) from 1900-2015. Black lines: cold and wet extremes in winter (DJF) (Ta < 10th percentile and Pa > 75th percentile). Gray lines: hot and dry extremes in Summer (JJA) (Ta > 95th percentile and EDI < 25th percentile).