Answer to comments of the referees

Compound extremes in a changing climate - a Markov Chain approach

K.Sedlmeier, S. Mieruch, G. Schädler and C. Kottmeier

In the following, we have answered all the comments of the three referees. The comments are highlighted in red color with our answers in black below. In these answers, we give a reference to the line of the latex diff file (attached below) in green where the changes in the manuscript can be seen.

1 Referee 1

Section 3 discussion of results (approx L250 on), it would be good to see some comparison with other research on the persistence of extremes in different regions and possible causes. e.g. Sillmann & Croci-Maspoli 2009, Furrer et al 2010, Photiadou et al 2014. Furrer, E.M., R.W. Katz, M.D. Walter, and R. Furrer, 2010: "Statistical modeling of hot spells and heat waves." Climate Research, 43, 191-205 Photiadou, C., Jones, M., Keellings, D., Dewes, C., 2014. Modeling European hot spells using extreme value analysis. Clim. Res. 58, 193–207. doi:10.3354/cr01191 Sillmann, J., Croci-Maspoli, M., 2009. Present and future atmospheric blocking and its impact on European mean and extreme climate. Geophys. Res. Lett. 36, L10702. doi:10.1029/2009GL038259

We have added a reference to some of the above mentioned research in the discussion sections (L 571-578). However a direct comparison is difficult as most papers refer to absolute univariate extreme events. Nevertheless follow up studies to analyze the interdependence between atmopsheric drivers and the here discussed dynamical aspects of relative extremes would be very interesting as they most likely also have an influence on the latter.

Similarly a sentence or two comparing the reliability of different models and observations would be good - e.g. CFSR and ERA-40 can be very different. This could be in the data section.

Thank you for this comment, we included a sentence in the methods section (Sect. 2.1.,L 106 and 2.2 L 132) and at the end of Sect. 4 (L 432) of the revised versionas well as the comment that the detection of differences in observational/reanalysis datasets and models concerning the dynamical behavior of extreme events is an additional interesting application of the method.

Did you test the significance of the changes in the reference period as well as the future? How did you account for uncertainty in the results?

Regarding the **uncertainty**, we took advantage of the applied ensemble approach. In Figs. 4 and 5 (of the revised manuscript) we show the results of the ensemble for the reference period, where we use a box plot for the ensemble: box = ensemble median and interquartile range, whiskers = ensemble minimum/maximum, gray bars: ensemble mean. This information is given in the text caption, to make it clear, we included it in the text under Sect. 4 (L 374ff). Similar box plots accounting for the ensemble uncertainty have been used in Figs. 10 and 11 showing the change signal of the descriptors. Here the changes were calculated for each ensemble member individually- the ensemble mean change is shown by the bar and the interquartile range of the change signal by the whiskers. As can also be seen from the figures we did not account for the uncertainties in the observational E-OBS dataset and consider the observations approximately as the truth. Nevertheless we included an additional section in the revised version where we calculated the error of the descriptors by a FT-resampling algorithm (Sect. 3.3, L 353-372). For this 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. 2 (page 12). In comparison to differences between regions and time periods, the error is small but we will include it in the discussions of Sect. 4 and 5. Regarding the **significance** we use the ensemble uncertainty, as mentioned above. In Sect. 5.2 we use the nonparametric Mann-Whitney-Wilcoxon test for the change signal (Figs. 10, 11). The p-values are shown below the bars in the respective figures.

L338 note about relative extremes - This should really be mentioned in the method section along with how you selected the extremes (e.g. thresholds, and at which level). Possibly a table of extremes would be informative for comparison?

We introduced a new subsection in the methods section of the revised manuscript where the thresholds and the partitioning of the data are described. This section 2.1. is called The Markov descriptors for two compound extremes. (L 258-273)

Minor corrections

We corrected these mistakes, the line numbers of the diff file where the changes were made are given in green.

L3 "the number of occurrences" L 3

L9 types L 9

L11 replace "which are" with "including" L 12

L12 rogue comma before fullstop. L 13

L26 occurrences L 39

L36 changes in the number of L 39

L46 should this be chaotic attractor? No

L107 please put into present tense to match the rest of the text. L 116

L115 ditto L 117

L145 unnecessary comma at start of line. L 160

L180 "number of states" L 195

L189 and 192 "Thus in the sense of successive compound..." L 207

L216 should this be per 100 days? No, 1000days is correct because this number refers to the total number of days, not only the compound extreme states.

L245 maybe say very rare? There are a lot of extremes in that sentence. L 380

L273 highest persistence is L 412

Figure 9 caption rogue fullstop before Percentages. now Fig. 7

2 Referee 2

In order to address the first comments of referee 2, we introduced a new section in the revised version called “Sensitivity analysis” (Sect. 3, L 274-371) 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 is included and discussed in the revised version of the paper in the newly introduced section (L 279-312)

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. (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 are included in Sect. 3, L 313-352) in the revised version of the paper. The stations used are listed in the data section (Tab. 1). 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 a 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. 10 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 in the newly introduced section (Sect. 3.3,L 353-371). 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. 2. 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 refer to these MIAAFT estimated errors when discussing the results throughout the paper.

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

Yes thank you, it should be France we have changed that. (L 440)

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 (bar), as well as the interquartile range (whiskers). We added a sentence at the beginning of Sect. 5.2 (L 469ff) of the revised version so it becomes clearer and have clarified it in the captions of Figs. 10 and 11.

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). Most changes are larger then the there derived errors which is an additional indicator of significance.

We additionally mentioned this in the text of the revised manuscript. 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.
The explanation has been expanded in the revised version, see L 482ff)

3 Referee 3

The authors should refer other approaches like the geostatistical analysis of spatially distributed extremes (Neves 2015). That is important because extremes have themselves some spatial organization.

We included this in the introduction of the revised version. (L 36ff)

There is no clear justification for the choice of the 6 box-regions and their size (6x6 grid points). Why they are representative of the PRUDENCE regions? Some minimal study about the spatial robustness of the Markov diagnostics should be presented. For example, does the results keep similar or change substantially when contiguous boxes are considered? The ideal should be to present maps of the diagnostics throughout Europe.

We thank the referee for that comment, because indeed, we have performed tests on the robustness of the Markov descriptors, which are the basis for the decision to use the actual 6 box-regions. The crucial point, why we have used the 6 box-regions is to achieve that each region contains the exact same amount of grid points / data points. This is of utmost importance for the comparison of the regions, otherwise, if the regions have been chosen with differing sizes no consistent comparison would be possible due to the fact that the Markov descriptors depend on the underlying sample size of the used data. To account for the spatial robustness we calculated the Markov descriptors for every grid point in Europe and visualised the results on a maps. From these maps we have seen that the Markov descriptors vary in general not strongly within the prudence regions. Accordingly we have chosen the 6 box-regions within the Prudence regions, which are representative for the respective region, based on the results of the grid point maps. In the revised we included the maps showing the grid point results (see Fig. 1 in the revised version.)

In the entropy definition H (eq. 7), log(1/m) must be replaced by log(m) so that H equals 1 for a random system without memory (all probabilities pij=1/m).

Thank you for the comment but our definition corresponds to those of other papers (see eg. Hill et al., 2004). Maybe you have missed the - sign at the beginning or the "‘/"’ sign in the equation (log(1/m) is the same as -log(m))? By using log(m) we would get negative entropies with our formula.

Line 189: Authors claim that H between 0 and 1 is an identification of deterministic chaotic behavior. However that condition is necessary but not a sufficient condition for chaos. Authors shall carefully rephrase the paper by taking that into account.

We agree with this comment and have rephrased the follwing sentence:

The dynamics of complex chaotic systems lie in between these limits, thus the entropy can be used to identify and characterize complex dynamics like deterministic chaos, which is not possible with standard linear methods

by

The dynamics of complex chaotic systems lie in between these limits, thus the entropy can give a hint to underlying complex dynamics like deterministic chaos, which is not possible with standard linear methods. To really test for deterministic choas other methods, based on state space reconstruction (e.g. estimating the correlation dimension, Lyapunov exponents etc.) to find strange attractors, are more suitable. (L 202ff)

and references to the chaotic behavior accordingly throughout the revised version of the paper.

Line 197: Authors say The reason for this is that the CO2 forcing is the only difference. . .. In fact, decadal variability is also likely. That sentence must be weakened by replacing the only by the main difference beyond the natural decadal variability.

No, because the crucial point is that this sentence (L 197) refers to the **model runs** (cf. line 198). The decadal variability of the model is not intrinsically changing with time. The only difference between the model runs in the past and in the future is the CO_2 forcing. Thus, changes of the decadal variability are of course possible, but the only reason is a changing CO_2 forcing. See L 15ff and the results in Sect. 3.2

Eq. 8 explain the meaning of the bar and subscripts rm.

Yes, we will do so and have also included a more detailed explanation of the EDI in the revised version (also see next comment). The bar in equation 8 stands for the climatological mean - $\overline{EP_{d,rm}}$ refers to the climatological mean state of EP corresponding to day d, where the climatological mean is calculated by a running mean of rm days over the 30 years of the respective time period. (Sect.2.4., L 245ff)

Line 234: Droughts may have different time scales from months to years. That is the reason for defining the SPI (Standard Precipitation index) (McKee et al. 1993). The presented EDI is appropriate for annual scaled droughts. Add this comment to the text. Moreover the EDI has its own annual cycle since the precipitation weights contributing to EDI are larger near the Julian day d. Does the annual cycle of EDI was removed?

The EDI does not have an annual cycle as this is intrinsically removed by the method (e.g. http://atmos.pknu.ac.kr/~intra2/eng.calculation.htm). In the equation:

$$EDI_d = \frac{EP_d - \overline{EP_{d,rm}}}{\sigma\left(EP - \overline{EP}\right)_d} \tag{1}$$

$\left(\overline{EP}\right)_d$ refers to the climatological mean state of EP and is calculated for each day as the 5day running mean over the 30years of the respective time period. Thus, by subtracting $\left(\overline{EP}\right)_d$ from EP, the annual cycle is removed. We are sorry that this did not become clear and have include a more thourough explanation of the EDI in the methods section of the revised version (Sect.2.4, L 251ff) and clearly state that the annual cycle is removed by the method.

Furthermore, the EDI is not only appropriate for annual scaled droughts. Since it is calculated from daily values, it is also able to detect droughts of shorter lengths. It is highly correlated to the soil moisture. For example if there is heavy rain on August 1st and September 30th, the EDI can detect a water deficit in between these two dates, whereas a monthly indice would not detect this (see Byun and Wilhite, 1999).

L235-238 Does temperature anomalies (Ta) and precipitation anomalies (Pa) refer to daily Ta and daily Pa with respect to the respective annual cycle. Please clarify. Add a sentence about the number of categories of the Markov chain and what categories of the compound attractor were considered? I suppose that authors have considered 2 parameters with a partition of 2 categories each. Confirm that at this stage for the sake of the paper understanding.

Yes, the Ta and Pa refer to daily temperature and precipitation anomalies with respect to the annual cycle. And we have considered 2 parameteres with a partition of 2 categories each which we then combined to a 4 state symbolic sequence. We added a new subsection to the methods section (

Fig. 3 In the recurrence plot I cannot see the black triangle for region 1.

thank you for the notice, we have changed that. (Fig. 5)

Fig. 4 In the caption, descriptors changes refer to changes in the period 1981- 2010 with respect to 1951-1980? Rewrite it in a clearer way.

We have replaced *1951-1980 vs 1981-2010.* by *changes between the time periods 1951-1980 and 1981-2010* (see Figs. 6, 7,10,11)

4 Additional changes

We have recalculated the entropy as we detected an error in the original manuscript. Therefore the entropy values are changed in the revised version. Furthermore we have changed the results for the summer extremes from TA > 95th percentile to TA>90th percentile. This way the summer and winter extremes yield the same number of univariate compound extreme events and the behavior of the two compound events can be better compared.

References

Byun, H.-R. and Wilhite, D. A.: Objective quantification of drought severity and duration, J. Climate, 12, 2747–2756, 1999.

Hill, M., Witman, J., and Caswell, H.: Markov chain analysis of succession in a rocky subtidal community, Am. Nat., 164, E46–E61, doi:10.1086/422340, 2004.

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.

Manuscript prepared for Nonlin. Processes Geophys.
with version 2015/09/17 7.94 Copernicus papers of the LaTeX class copernicus.cls.
Date: 25 July 2016

Compound extremes in a changing climate - a Markov Chain approach

Katrin Sedlmeier, Sebastian Mieruch, Gerd Schädler, and Christoph Kottmeier

Institute for Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany

Correspondence to: Katrin Sedlmeier
(katrin.sedlmeier@kit.edu)

Abstract. Studies using climate models and observed trends indicate that extreme weather has changed and may continue to change in the future. The potential impact of extreme events such as heat waves or droughts does not only depend on their number of ~~occurrence~~ occurrences but also on “how ~~the~~ these extremes occur”, i.e. the interplay and succession of the events. These quantities are quite unexplored, for past changes as well as for future changes and call for sophisticated methods of analysis. To address this issue, we use Markov chains for the analysis of the dynamics and succession of multivariate or compound extreme events. We apply the method to observational data (1951-2010) and an ensemble of regional climate simulations for Central Europe (1971-2000, 2021-2050) for two ~~type~~ types of compound extremes, heavy precipitation and cold in winter and hot and dry days in summer. We identify three regions in Europe, which ~~are probably~~ turned out to be likely susceptible to a future change in the succession ~~or dynamics~~ of heavy precipitation and cold in winter, ~~which are a region~~ including a region in south western France, northern Germany and in Russia around Moscow~~,~~.~~The~~A change in the succession of hot and dry days in summer ~~will probably affect~~ can be expected for regions in Spain and Bulgaria. The susceptibility to a dynamic change of hot and dry extremes in the Russian region will probably decrease.

1 Introduction

Multivariate extreme events (in this paper used in the sense of extremes of two or more climate variables occurring simultaneously) are likely to impact society greater than their univariate counterparts. For agriculture for example, the impact of a heat wave and a drought occurring at the same time is higher than for a univariate extreme where the other variable is in a normal state. These multivariate or so called compound events (IPCC, 2012) have received more and more attention in the scientific literature over the past years although still not to the extent of extremes of only one variable. Methods to analyze them include simple threshold analysis, multivariate distribution functions using copulas (e.g. Schoelzel et al., 2008; Durante and Salvadori, 2010), Bayesian approaches (e.g. Tebaldi and Sansó, 2009) or indices which are derived from multiple variables (e.g. the wildfire index KBDI (e.g. Keetch et al., 1968) or the revised CEI Gallant et al. (2014)). Furthermore, methods of multivariate extreme models have been used for the geostatistical analysis of spatially distributed extremes (Neves, 2015). All these methods focus mostly on the linear climate change signal - the absolute change in the number of ~~occurrence~~ occurrences or the calculation of return periods. The succession, i.e. the temporal ordering of the compound events is in most cases ~~mostly~~ not the main objective. For instance, the IPCC (IPCC, 2012) states : "*A changing climate leads to changes in the frequency, intensity, spatial extent, duration, and timing of extreme weather and climate events, and can result in unprecedented extreme weather and climate events.*" What is implicitly addressed with "duration and timing", but not explicitly stated is the succession of extreme events, which is quite unknown for past as well as future extremes.

The method proposed here, which is based on Markov chains, concentrates on the dynamical behavior or succession of these compound extreme events and studies an aspect of climate change which has not received much attention up to now, but is nevertheless important. We investigate a behavior of extremes which cannot be determined by simply analyzing the ~~change of~~ changes in the number of extremes. We can, for example, reveal changes in the entropy of the succession of compound extremes which is connected to the chaotic behavior of the climate variable. Thus an observed increase of this measure could be connected with an increase in the chaotic, intermittent or irregular nature of the system. On the other hand, a decrease of entropy corresponds to a slow-down of these dynamics. Knowledge about such developments for future climate, which rarely exists, could be important for many sectors e.g. agriculture, economy and society.

Previous studies on model dynamics have concentrated more on overall dynamical behavior such as Steinhaeuser and Tsonis (2014) who have conducted a model intercomparison study focusing on dynamical aspects based on a climate networks framework. The method introduced in this paper is ~~modified from a work by~~ inspired by the work of Mieruch et al. (2010). The idea is to understand climate time series as trajectories on a complex, possibly strange attractor (Lorenz, 1963). We partition the time series or state space into a finite number of states. This yields a coarse-grained description of the system, which can then be analyzed in the framework of symbolic dynamics (Ebeling et al., 1998; Daw et al., 2003). We apply a Markov Chain analysis on these symbolic sequences representing compound extremes, and characterize their dynamical or successional behavior using a small set of descriptors.

In this paper we study two different kinds of compound extreme events which are likely to have an impact on society, namely cold and heavy precipitation in winter, and heat and drought in summer. The Markov method is applied to E-OBS observational data (1951-2010) (Haylock et al., 2008), and an ensemble of regional climate simulations with the regional climate model COSMO-CLM driven by different global climate model data and ERA-40 reanalysis (Uppala et al., 2005). The time periods considered are the recent past (1971-2000) and the near future (2021-2050).

We identify regions in Europe, where the dynamical behavior of the analyzed compound extremes is prone to change. These findings highlight that it is not only the (simple)

linear increase of the occurrence of extremes (due to an increase in mean and variability), which is a challenge for adaption and mitigation. On top of these changes the regions also have to struggle with changes in the succession of compound extremes (defined as relative to a new normal state with changed mean and variability).

The strategy of this study is first to show that the Markov method is able to extract different dynamics of compound extremes for different regions in Europe, based on observational data and model data. Thus, on the one hand we see that the method yields meaningful information and on the other hand we show that the climate models are able to reproduce these dynamics in the frame of acceptable uncertainties. Additionally, we extract temporal change signals of the dynamics of compound extremes based on observations between the periods 1951-1980 and 1981-2010. This information is new and if used as supplementary information to other analyses, could lead to a better understanding of changes of extremes in Europe. For this paper, the magnitude of the observed past changes have been assessed, because it is important for a better interpretation and classification of future changes which are calculated by using the simulated regional climate model data. A comparison of the change signals between 1971-2000 and 2021-2050 to the observed past changes shows that they are of the same order of magnitude.

The paper is divided into the following sections. In Sect. 2, data and method will be introduced, followed by a sensitivity analysis of the method with respect to spatial and temporal variability as well as the error of estimation using FT surrogates in Sect. 3. A validation of the model ensemble is shown in Sect. 4. The change signal is analyzed in Sect. 5. Summary and outlook will be given in Sect. 6 and some areas discussed where the application of this method might be of value.

2 Data and Methods

2.1 Regional Climate Ensemble

For our analysis, we use a 12-member ensemble of regional climate simulations for Central Europe at a resolution of 50 km. The ensemble has been generated by downscaling different global climate model outputs with the regional climate model COSMO-CLM (COnsortium for Small scale MOdelling model - in CLimate Mode, Doms and Schättler (2002); Rockel et al. (2008)), further referred to as CCLM. The CCLM is a non-hydrostatic climate model coupled to the soil vegetation model TERRA and is the climate version of the numerical weather model of the German weather service. Data from six different global climate models (GCMs) has been used as initial and boundary data. Two of the GCMs have used the emission scenario A1B (Nakicenovic and Swart, 2000) as external forcing : CCCma3 (Scinocca et al., 2008) and three realizations of ECHAM5 (Roeckner et al., 2003). The other four, ECHAM6 (Stevens et al., 2013), CNRM-CM5 (Voldoire et al., 2013), HadGM3 (Collins et al., 2011) and EC-EARTH (Hazeleger et al., 2010) have used the emission scenario RCP8.5 (Riahi et al., 2011; Van Vuuren et al., 2011). Additionally the Atmospheric Forcing Shifting method (Sasse and Schädler, 2014) was applied to the ECHAM6 data. For this method the global climate data interpolated to the 50 km grid is shifted by two grid points in all cardinal directions before being used as boundary data. This accounts for the uncertainty in positioning of synoptic systems when interpolating the GCM data to the required resolution for forcing the RCM simulations. As all five ECHAM6 driven simulations obtained this way exhibit a high correlation, they are all weighed with a factor of 1/5 when calculating the mean. All other models receive a factor of one which leads to an effective ensemble size of eight. Additionally we use a COSMO-CLM run driven by ERA-40 (Uppala et al., 2005) boundary conditions. The ERA-40 reanalysis boundary conditions are assumed to be close to the "‘true"’ observed state. Nevertheless, they depend on the model, observational data and assimilation tequnique, among others and are not free from biases (see e.g. Hagemann et al., 2005; Simmons et al., 2004).

The simulation time periods are the recent past (1971-2000) and the near future (2021-2050). An analysis of the temperature trends of different ensemble members showed that the distribution of trend depends more strongly on the chosen global climate model than on the emission scenario. We therefore combine simulations with boundary data from GCMs with different emission scenarios to set up our ensemble.

We ~~chose~~ choose six regions, each comprising 6×6 grid points for our analysis. The regions ~~were~~ are chosen based on the PRUDENCE regions (Christensen and Christensen, 2007) which could not be used because of the necessity of the same amount of grid points for each area, and due to test results which show a different behavior for these regions. We investigate 30 year periods of daily data, thus each time series consists of $\approx 11,000$ data points, yielding $\approx 36 \times 11,000 \approx 400,000$ points in time for each region and ensemble member. The model domain and the six investigation areas, which are located in Spain, France, Germany, Scandinavia, Bulgaria, and Russia are shown in Fig. 3. These roughly match the PRUDENCE regions which are not applicable for the analysis since equal sized areas are a requirement for comparison among regions.

2.2 Observational data

For the comparison of our regional climate ensemble with observations, we use temperature and precipitation data from the gridded E-OBS dataset (Haylock et al., 2008). This dataset was produced as part of the ENSEMBLES project by interpolating station data from the ECA&D station dataset (European Climate Assessment, Klok and Klein Tank, 2009) to a 25 km grid. The station density is highest in Switzerland, the Netherlands and Ireland and rather low in Spain and the Balkans which leads to an over-smoothing in these areas. This especially affects extremes and has to be taken into account when validating our ensemble against E-OBS data. Furthermore it should be noted that a comparison

Table 1. ECA&D Station data

	Station (Station number)
1	Bamberg (40)
2	Hamburg Fuehlsbuette (47)
3	Hohenpeissenberg (48)
4	Potsdam (54)
5	Hamburg-Botanischer-Garten (4180)
6	Hamburg Sankt-Pauli (4184)
7	Hamburg-Wandsbek (4186)
8	Quickborn Kurzer Kamp (4536)

of E-OBS and another gridded dataset, namely Hyras (Rauthe et al., 2013) (only Central Europe), with respect to the dynamical behavior that we analyze in this paper, revealed differences between the two datsets (Sedlmeier, 2015). A comparison of dynamical aspects of different observational datasets yields an interesting application of the method which however will not be addressed within this paper. We additionally use blended temperature and precipitation time series starting from 1900 of eight stations (all in Germany) of the ECA&D dataset for a sensitivity analysis described in Sect. 3. The eight stations are listed in Tab. 1.

2.3 Compound extremes with Markov Chain descriptors

The method used in this paper consists of describing temperature and precipitation time series by a Markov Chain and subsequently calculating descriptors, which characterize the dynamical (successional) behavior of the compound extreme states. The method has been used in biology (Hill et al., 2004) to describe dynamics of succession of species in a rocky subtidal community. It has been introduced to atmospheric science by Mieruch et al. (2010) who used it for climate classification and a comparative study of two regions. In this section, a short introduction to Markov chains is given, followed by a step by step description of the method.

A first order, m state (m= number of discrete states of the Markov Chain), homogeneous Markov Chain is a time discrete, state discrete stochastic process which fulfills the Markov property:

$$P(x_t|x_{t-1}, x_{t-2}, ..., x_{t-n}) = P(x_t|x_{t-1}) \quad (1)$$

meaning that the present state x_t is only dependent on the preceding state x_{t-1}. From the Markov chain, a transition probability matrix **P** of the order $m \times m$ can be calculated which consists of all possible conditional probabilities $P(x_t|x_{t-1})$ between the m different states of the Markov chain.

For a homogeneous (≡ stationary) Markov chain, the transition probability matrix is time independent. A stationary distribution $\boldsymbol{\pi}$ is a vector that fulfills the following equation

$$\boldsymbol{\pi} = \mathbf{P}\boldsymbol{\pi}. \tag{2}$$

To test for homogeneity one must solve the eigenvalue problem of equation 2 to calculate the stationary distribution $\boldsymbol{\pi}$. If this is identical to the empirical distribution

$$\hat{\pi}_j = \frac{n_j}{\sum_j n_j}\,. \tag{3}$$

~~,~~the time series is considered stationary. The entries (transition probabilities) of the transition matrix **P** are estimated by

$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_i n_{ij}}\,. \tag{4}$$

In the following, the main steps of the Markov analysis are explained:

a) **Partitioning and combining of univariate time series to a multivariate symbolic sequence**
To represent the univariate time series (here daily mean temperature anomalies and daily precipitation anomalies) by a Markov chain, each time series is partitioned into a symbolic sequence of extreme and non-extreme regimes. These univariate symbolic sequences are then combined into a multivariate symbolic sequence of $m = 2^v$ different states (v number of variables). In this paper, $v = 2$, thus there are four possible states.

b) **Calculation of the transition probability matrix**
From the 2^v-state Markov chain, a transition probability matrix **P** of dimension $2^v \times 2^v$ can be calculated. Two conditions have to be met when calculating the descriptors. No entry of the transition probability matrix should be equal to zero and the time series needs to be stationary for the transition probability matrix to be time independent (see equations 2, 3).

c) **Calculation of the descriptors**
Following Mieruch et al. (2010), we focus on only three of the descriptors mentioned in Hill et al. (2004): persistence, recurrence time and entropy. These descriptors can be estimated for single states of the symbolic sequence or for the whole system. As the focus of this work lies on the compound extreme state, only the single-state definition of the descriptors is considered.

Persistence:

$$P_j = \hat{p}_{jj} \tag{5}$$

The persistence gives the probability that the system will stay in an extreme state in the following time step if it resides in an extreme state at the current time step. The limits are 0 (the system will never stay in the extreme state) and 1 (the system will always stay in the extreme state). Regarding the succession of the compound extremes, the persistence tells us how long the extremes last.

Recurrence time:

$$R_j = \frac{1-\hat{\pi}_j}{(1-\hat{p}_{jj})\,\hat{\pi}_j} \tag{6}$$

The recurrence time describes the number of days the system needs to get back to the extreme state. The limits are 0 (the system never leaves the state, corresponding to a persistence of 1) and ∞ (the system never comes back to the extreme state). The recurrence time is connected to the persistence. If the persistence increases, the recurrence time will also increase and vice versa, except if a change in the number of states $\hat{\pi}_j$ occurs. Thus, it is important to include the absolute ~~numbers~~ number of the states for the interpretation of the results.

Entropy:

$$H\left(p_j\right) = -\sum_i \hat{p}_{ij} \log \hat{p}_{ij} / log\left(\frac{1}{m}\right) \tag{7}$$

According to Shannon (1948), the entropy is an inverse measure of the predictability of the Markov Chain. Its limits are 0 (deterministic system) and 1 (random system). The dynamics of complex chaotic systems lie in between these limits, thus the entropy can ~~be used to identify and characterize~~ give a hint to underlying complex dynamics like deterministic chaos, which is not possible with standard linear methods. To really test for deterministic chaos other methods, based on state space reconstruction (e.g. estimating the correlation dimension, Lyapunov exponents etc.) to find strange attractors, are more suitable. Thus, in the sense of ~~the succession of~~ successive compound extremes a change in entropy tells us if the succession of extreme states gets more chaotic or more regular.

d) **Data pre-processing**

In order to extract the information on ~~the succession of~~ successive compound extremes, we have to remove linearities (e.g. trends) and cycles, which would bias the results. Thus, we remove the external solar forcing by subtracting the mean annual cycle. A long-term trend is removed by a linear regression. Although, e.g. the temperature trend due to the anthropogenic CO_2 emissions is removed from the data, we hypothesize that all changes in the succession of extremes are linked to the CO_2 increase. The reason for this is that the CO_2 forcing is the only difference between the model runs for the periods 1971-2001 and 2021-2050.

We use percentiles to partition our datasets, and keep the number of univariate extreme events the same for different time periods and regions as well as for all ensemble members. By this, the results can be compared among each other, differences are only due to different dynamical behavior. For partitioning dry days, we did not use precipitation anomalies but the effective drought index (EDI). The EDI (see Sect. 2.4) is related to soil moisture and is therefore a much better measure for describing dry extremes than precipitation itself, since all percentiles below the percentage of dry days will lead to the same partitions.

In order to get a better feeling for the descriptors and understand how they relate with each other, we will do a small thought experiment. We take a Markov chain consisting of a time series of 1000 symbols of which 10% are extreme, the rest are normal. In this case a persistence of 0.5 would mean that in half of the 100 extreme cases, the next case is also extreme, there are 50 transitions from the extreme state to the extreme state. The maximum episode length in this case is thus 51 extreme states in a row (with all others randomly distributed). The recurrence time and entropy are inversely related to how these 50 extreme transitions are ordered. Recurrence time depends on the number of episodes (fewer episodes lead to a larger recurrence time, more episodes to a shorter recurrence time) and entropy additionally on the mean episode length. In this paper, we also look at changes in the descriptors. A change in persistence of 0.05 in the above case would mean 5 more extreme-extreme transitions per 1000 days, and an increase from 50/100 to 55/100 (extreme-extreme transitions/extreme-normal transitions) is surely a noticeable change. The range of actually probable values of the descriptors is smaller than the whole possible range. A persistence of 0.99 for example, would mean that there is only one extreme episode in the whole time period, all 100 extreme states occur after each other. In a climate system, this is unlikely to happen. Thus, for climate one cannot expect to observe a change of the daily persistence from e.g. 0.5 to 0.8, because such a change would be catastrophic.

2.4 Effective drought index: EDI

The effective drought index (EDI) is an index for detecting drought conditions by calculating daily deviations of precipitation from a climatological mean state. It was proposed by Byun and Wilhite (1999)~~and is calculated by the following formula for a given day d:~~

$$EDI_d = \frac{EP_d - \overline{EP_{d,rm}}}{\sigma\left(EP - \overline{EP}\right)_d}$$

. An important concept of the EDI is the use of effective precipitation EP, rather than precipitation P itself. EP describes the depletion of water sources by a weighted summation over the 365 days preceding a given day d:

$$EP_d = \sum_{n=1}^{365} \left(\frac{\sum_{m=1}^{n} P_{d-m}}{n} \right) \tag{8}$$

By this, the memory effect of the soil is taken into account. EP therefore strongly correlates with soil moisture and the EDI is thus a good measure when considering droughts. Using the effective precipitation EP, the EDI is calculated by the following formula for a given day d:

$$EDI_d = \frac{EP_d - \overline{EP_{d,rm}}}{\sigma\left(EP - \overline{EP}\right)_d} \tag{9}$$

where $\overline{EP_{d,rm}}$ is the climatological mean corresponding to a given day d calculated as the 30-year average over a 5 day running mean (rm=5). By subtracting this climatological mean of EP from the daily value, the yearly cycle is removed from the EDI time series.

3 ~~Markovian descriptors for the reference period 1971-2000~~

2.1 The Markov descriptors for two compound extremes

To calculated the Markov descriptors we first calculated temperature and precipitation anomalies using the mean annual cycle of the respective time period and ensemble member/observation. We calculate the Markovian descriptors for two types of extremes

- cold and heavy precipitation (temperature anomaly (Ta) < 10th percentile and precipitation anomaly (Pa) > 75th percentile) in winter (DJF) and
- heat and drought (Ta >90thpercentile and EDI < 25th percentile) in summer (JJA),

and for the six regions shown in Fig. 3. As an example, we show how we constructed the Markov chain for the cold/heavy precipitation extreme at a single grid point. First we identify temperature values below the 10th percentile $T_{\ell,t}$ and above $T_{h,t}$ (t is the time index). Similarly we identify low and high precipitation values $P_{\ell,t}$ and $P_{h,t}$. Following we combine these symbols and find the following possible states: ($T_{\ell,t}$,$P_{\ell,t}$), ($T_{\ell,t}$,$P_{h,t}$), ($T_{h,t}$,$P_{\ell,t}$) and ($T_{h,t}$,$P_{h,t}$). Now we can rename these states to e.g. $S_{\ell,\ell,t}$, $S_{\ell,h,t}$, $S_{h,\ell,t}$ and $S_{h,h,t}$ and then a Markov chain could look like: $S_{\ell,h,1},S_{\ell,\ell,2},S_{\ell,\ell,3},S_{h,h,4},S_{h,h,5},S_{h,h,6},S_{h,\ell,7},\ldots,S_{h,h,N}$, where N is the total number of data points. From such a sequence we calculate the transition probability matrix and from this, the descriptors.

3 Sensitivity analysis

Before applying the method to the observational data and the model ensemble, we tested the applicability of the method by several sensitivity tests using the above defined descriptors. Therefore we consider the gridded E-OBS data and additionally ECA&D station data.

3.1 Spatial variability

In order to test the spatial variability of the descriptors we calculated them for the entire E-OBS dataset for the time period 1971-2000 for the two types of extremes mentioned above. The descriptors were calculated over a moving window of 9 gridpoints. The time series for each grid point were detrended and partitioned separately before the 9 partitioned time series were merged to calculate the descriptors. The value is assigned to the center grid point.

For both type of extremes, the descriptors show smooth spatial patterns (see Fig. 1), nevertheless variations between different regions can be identified.

The persistence for the winter extremes (left side of Fig. 1) is lower than for summer extremes, especially in northern and Central Europe compound cold and wet events are most likely events of a short duration and rather rare (with recurrence times of up to 400 days). Along the Mediterranean coast and south eastern Europe the values are higher and probabilities of residing in a compound extreme state of over 50% are observed. The recurrence time for these events is also comparatively low (around 100 days). Interpreting the results, one has to keep in mind that we are always referring to relative compound extremes. The entropy is around 0.9 for most of the area with small regions showing lower entropies down to 0.5. These high values can be explained by the low persistence - as compound winter extremes are grouped in very short episodes (low persistence), they are very hard to predict. The highest persistences for summer events (right side of Fig. 1) are observed in Scandinavia and the eastern part of the E-OBS domain and lowest in Central Europe and the northern coast of Spain. The persistence is above 50% for the whole domain, which means that the probability of the system residing in a compound extreme state is high and these events are grouped in episodes of long duration. The recurrence time lies between 40 and 100 days and is as such also lower than that for compound winter events. Lowest values are observed in the Balkan region. The entropy lies between 0.4 and 0.65 which means that the extreme events are not so easy to predict, especially for parts of Central Europe where the entropy is highest. However, according to our definitions, summer extremes can better be predicted than winter extremes.

For the main analysis in this paper we apply the method to 6 regions which we chose in rough agreement with the Prudence regions. The crucial point for being able to compare the descriptors of different regions is that each region contains the same amount of grid/data points. Since the descriptors do not vary strongly within the Prudence regions, we chose regions consisting of 6x6 gridpoints from within these widely used regions. The regions which will be analyzed in the further sections of this paper are shown in Fig. 3.

Note that the results shown in Fig. 1 can only be qualitatively compared to those of the regions considered later or the station data in the next section as the number of grid points (or stations) contributing to the analysis differs.

3.2 Temporal variability

To assess the temporal variability of the descriptors we calculated the descriptors for 30-year moving windows of observational station data from the ECA&D station dataset Klein Tank et al. (2002). Since we are interested in the daily values of temperature and precipitation, only stations were chosen with a continuous daily record (with an allowance of 50 missing values at most). Using these criteria there are eight stations with temperature and precipitation time series from 1900-2015 of which all are in Germany (see Tab. 1). One station has 15 missing values for temperature. These days were excluded from the analysis, considering the 30year time windows consisting of 10950 days, this amounts to roughly 0.1% of the values and does not alter the value of the descriptors. Of these eight stations, 5 are in the vicinity of Hamburg and have the same values for the first 17-22 years. The records of two stations in Hamburg are identical throughout the whole time period, therefore only one of them is included in the analysis which leaves a total of seven stations.

The descriptors were again calculated for both types of compound events. Linear temperature trends were removed separately for each of the 30year time-windows and in order to fulfill the criteria of the Markov method (stationarity and non-zero entries of the transition probability matrix, see Sec. 2.3), the partitioned data of these seven stations were combined to one time series to calculate the descriptors.

The results are shown in Fig. 2 for both winter (black) and summer (gray) extremes. 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 an observed shift in the global increase in CO_2 around 1950 (see e.g. Pachauri et al., 2014, Fig. SPM.1 (d)). 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 a lower level of persistence and to a higher level of recurrence time. Although CO_2 is still increasing after 1950, the recurrence time e.g. remains constant. Hence, the recurrence time seems not to be dependent on the absolute CO_2 concentration, but on the increase of latter.

2. Thus we can conclude that the natural variability of the descriptors 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 the natural variability of the descriptors in preindustrial times. But we could show that the approximate natural variability (before and after the shift in 1950) is smaller than the shift, which is probably due to the change in CO_2 increase. Just for a rough estimation: the mean level shift of the persistence for winter extremes is about 50% (from 0.2 to 0.1) and for the recurrence time it is about 20% (from 180 to 140 days). Regarding our results of changes of the descriptors (1971-2000 vs. 2021-2050) presented below (see see Sect. 6), we find changes of the persistence larger than 50% and changes of the recurrence time larger than 20%. We additionally perform significance tests on our results which show that these changes are indeed significant, excluding natural variability as the source for the observed changes.

3.3 Error of estimation using Fourier Transform (FT) surrogates

To assess the estimation error of the descriptors we used the Multivariate Iterated Amplitude Adjusted Fourier Transform (MIAAFT) algorithm as described by Venema et al. (2006); Schreiber and Schmitz (2000). By this algorithm, the data is shuffled and thus the original distribution is preserved. In addition, the auto and cross-correlation of the temperature and precipitation time series are approximately preserved. We constructed 100 MIAAFT surrogates for the temperature and precipitation anomalies (or the EDI time series for summer events, respectively) for the E-OBS dataset for the reference period (1971-2000). We then estimated the standard deviation of the descriptors calculated from these surrogate time-series. It is important to note that this standard deviation, under the framework of such a bootstrap test, already represents the standard error of the mean, which corresponds to the normal standard deviation devided by $\sqrt{N}$. The errors for both types of extremes and the six regions are listed in Tab. 2. The errors do not vary much between the different extremes and regions, the error of the persistence is in the order of 0.01 or lower, the one of the recurrence time between 1 and 2.6 and the error of the entropy in the order of 0.005. Adopting these errors to the values of the E-OBS descriptors for the reference period (shown in Figs. 4 and 5 in Sect. 4) the error of the persistence is about 2-10%, for the recurrence time about 2% and for the entropy about 1-2% (cf. Tab. 2). This estimation error is much smaller than the ensemble uncertainty and can approximately be neglected. This shows that the estimation of the descriptors is robust. Further, we will consider the E-OBS data approximately as truth and we will use the ensemble uncertainty as the error for our main analysis.

Table 2. Estimation of the error of the descriptors by using MIAAFT surrogates for winter and summer extremes. The values were calculated using the E-OBS dataset for the reference period (1971-2000). In parentheses the percentage of the error with respect to the value of the E-OBS descriptors for the same time period and region are given.

	DJF			JJA		
	P	R	E	P	R	E
reg1	0.010 (7.9%)	1.701 (2.1%)	0.004 (0.5%)	0.007 (1.2%)	1.183 (1.8%)	0.009 (1.6%)
reg2	0.011 (8.2%)	2.182 (1.5%)	0.010 (1.0%)	0.010 (1.7%)	2.055 (3.5%)	0.010 (1.7%)
reg3	0.010 (7.9%)	2.563 (1.5%)	0.005 (0.6%)	0.009 (1.6%)	0.923 (1.4%)	0.007 (1.1%)
reg4	0.008 (12.3%)	1.150 (1.0%)	0.005 (0.6%)	0.008 (1.3%)	0.990 (1.8%)	0.011 (1.9%)
reg5	0.010 (3.9%)	2.450 (1.0%)	0.010 (2.2%)	0.008 (1.3%)	1.103 (1.7%)	0.009 (1.5%)
reg6	0.007 (1.8%)	0.797 (1.4%)	0.004 (0.4%)	0.009 (1.6%)	1.150 (2.0%)	0.009 (1.7%)

4 Markovian descriptors for the reference period 1971-2000

Fig. 4 shows the descriptors for cold extremes and heavy precipitation in winter from 1971-2000. The As for all boxplots in this chapter, the boxes show the 25th and 75th quantile of the ensemble , (interquartile range) and the whiskers the minimum and maximum value of the ensemble. The colored line marks the ensemble median and the gray line the ensemble mean. Crosses mark the descriptors of the observations. The observed persistence for the different regions lies between 0.06 and 0.37. This means that the system does not stay in this extreme state for a very long time, the lowest observed persistence is in region 4 (Scandinavia) where extreme-extreme transitions are extremely very rare. The recurrence times vary strongly between the regions, the values are between 64 and 314 days. Regions 1 and 6 (Spain and Bulgaria) show the lowest recurrence times. In region 6 (Bulgaria) the compound cold and wet episodes have the longest duration and occur with the highest frequency. The entropy of the observations lies between 0.23 0.86 in region 3 (Germany) and 0.25 0.96 in region 1 (Spain) and between 0.19 0.74 in region 3 (Germany) and 0.26 in region 5 (Russia0.98 in region 1 (Spain) for the CCLM ensemble. Thus, the deduced entropy (both, observations and model) covers a rather small portion of the range of theoretically possible values from 0 to 1. As mentioned in Sect. 2.3 the range in which we actually expect the values of the descriptors is smaller. Therefore, when comparing the descriptors, the values have to be interpreted relative to the regions. One must be careful, however, because the descriptors do not permit to draw any conclusions about the absolute predictability of the states as long as the total numbers of states are not considered.

Focusing on the descriptors for the CCLM ensemble (box plots and gray bars in Fig. 4), we can see that with this method we are able to detect significant differences in dynamical behavior between some of the regions. In comparison to the descriptors of the observations (crosses in Fig. 4), the ensemble is able to capture the differences between the regions fairly well except for the persistence in region 5 where the ensemble shows a much lower persistence and the recurrence time of region 4 (Scandinavia) which is lower for the observations. However, these are regions where the the density of station data underlying the E-OBS dataset is not very high and the E-OBS results may not be as reliable. The highest persistence is again in region 6 (Bulgaria) which also shows the lowest recurrence time and therefore has comparatively long events which occur more frequently than in other areas. The triangles mark the descriptors of the reanalysis driven simulations. They fit well for some regions, for others they are farther away from the observations than the CCLM-ensemble. such that the number of univariate extremes is the same for hot and dry extremes and cold and wet extremes

Fig. ?? 5 shows the descriptors for hot and dry extremes in summer. Crosses again mark the descriptors of the observations. All descriptors are higherPersistence and recurrence time are higher, entropy is lower for hot and dry summer extremes than for cold and wet extremes in winter, this also holds when partitioning the data. A direct comparison can be made because the data were partitionedsuch that the number of univariate extremes is the same for hot and dry extremes and cold and wet extremes (not shown). This might partly be due to the lower variability of EDI compared to precipitation anomalies but one would also expect the dynamical behavior of these extremes to be different. By our definition, hot and dry episodes in summer are longer and not as frequent as cold and wet extremes in winter. The highest ~~persistences are in regions 2,~~persistence is in regions 4 and 5 (~~France,~~ Scandinavia and Russia), the lowest in region 3 (Germany). The entropy lies between ~~0.24 and 0.25 and is fairly similar for all regions considered. This is a higher value~~ 0.53 and 0.60 and is highest in region 2 (France) and lowest in region 6 (Bulgaria). The values are lower than for the cold and wet extremes, ~~so this state probably~~ the winter compound extreme state exhibits more complex dynamics and is harder to predict (caution: this is also influenced by the total number of extremes). The CCLM ensemble (box plots) again captures the tendencies of the observed descriptors fairly well but shows a large spread and differences between the regions are ~~not significant~~mostly not significant for persistence and recurrence time. The ERA-40 driven CCLM simulations (triangles in Fig. ~~??)~~ 5) again fit well to the observations for ~~most regions~~ some regions and show very different behavior for others.

For both types of compound extremes the ensemble mean and median seem to be able to capture the differences between regions shown by observations although not always in absolute numbers. An interesting result is that reanalysis driven CCLM data is sometimes farther away from the observational descriptors than the model data, especially for the cold and wet extremes in winter. This leads to the question whether the dynamical behavior of the driving GCM is greatly altered by the RCM downscaling and errors in both models compensate during the downscaling process. A further cause of this deviation of the ERA-40 driven simulations could be a misrepresentation of the dynamics by the reanalysis dataset. A follow up study comparing dynamical behavior of both RCM and ~~GCM~~ GCMs is planned for the future. Additionally it would be interesting to also compare different reanalysis datasets using this method as there have been studies showing differences in their variability (e.g. Hagemann et al., 2005)

5 Climate change signal of the Markovian descriptors

5.1 Change signal within the reference period

In order to get an idea about the order of magnitude of the change signal, the observational E-OBS dataset was split into two equal parts of 30 years, 1951-1980 and 1981-2010. The descriptors were calculated for both time periods and a change signal derived.

For cold and wet extremes (see Fig. ~~??~~6) all regions except ~~Bulgaria~~ France show a decrease in persistence, region 5 and 6 (Russia and Balkan) show the strongest absolute decrease (≈ 0.15) and Germany the highest relative decrease of -72 % (relative changes are shown above the respective bars). The recurrence time does not change much for all regions except region 5 (Russia) where it decreases by 150 days. In this region, compound cold and wet extremes occurred more frequently but were of shorter duration in 1981-2010. The entropy ~~only shows changes greater~~ shows a decrease of more than 5~~% in Scandinavia where it decreases and~~ % in Spain and Germany where the system becomes more regular. In Spain an increase of entropy is observed and the compound extremes are harder to predict in 1981-2010 with respect to 1951-1980. The change signal for all descriptors and seasons (except for the entropy of France and Russia) are greater than the estimated error by FT-surrogates (see Tab. 2), thus these changes are robust.

Changes for hot and dry extremes in summer (see Fig. ~~??)~~ 7)~~show a decrease in all descriptors~~ are below 10% for most regions~~, region 5 (Russia)is the only one with an increase in persistence and recurrence time. The order of magnitude of the change signal is slightly lower than for cold and wet extremes in winter (maximum change persistence: -0.07, recurrence time -7 days) and here, regions 3 (Germany) and 4 (Scandinavia) are the ones with the largest changes~~. Nevertheless for most regions these changes are still greater than the estimated errors by FT-surrogates (see Tab. 2). In Scandinavia, both persistence and recurrence time show a decrease, the extreme episodes are of shorter duration but occur more often. In Spain and Germany, both descriptors show an increase - especially in recurrence time, thus episodes of compound extremes occur less frequently. An increase in recurrence time can also be seen in Russia. The entropy increases in regions 3-6 (Germany, Scandinavia, Russia and Balkan), in these regions the system becomes less regular with respect to compound hot and dry events and harder to predict, whereas in Spain the Entropy shows a decrease - these compound events are easier to predict.

5.2 Projected changes in the near future

In a second step we calculate the change signal between 1971-2000 and 2021-2050 for all members of the CCLM-ensemble. An additional information of interest for the interpretation of the results is the change in the number of compound extreme days. The number of univariate extreme days are kept constant when partitioning the data (see Sect. 2.3) but the combination can change.

The climate change signal is calculated separately for each ensemble member and then the mean climate change signal (bar in the following plots) as well as the interquartile range (marked by the whiskers) of the individual change signals are calculated and pictured. The number of compound cold and wet extreme days increases in all regions except region 5 (Russia) between the two time periods 1971-2000 and 2021-2050 and the number of compound extreme days differs between the regions. Regions 1 and 6 (Spain and Bulgaria) show the highest number of compound extreme events. (see Fig. 8). The ensemble mean values of the descriptors for cold and wet extremes in winter are shown in Fig. 10, whiskers give the interquartile range. The significance of the change signal was calculated using the nonparametric MannWhitneyWilcoxon test ~~, the~~ which tests for a difference in location of the values of the ensemble for the two different time periods. The p-values are shown below the bars in the respective figures. ~~Most of results of this chapter are not~~ About one third of these p-values are smaller than 0.5, thus significant at the 5% significance level, e.g. region 5 (Russia) ~~however~~ shows a significant change signal for the persistence and some changes are significant at the 10 % or 20% significance level (~~P-value~~ p-value ≤ 0.1 or ≤ 0.2). ~~Following~~ Nevertheless we follow von Storch and Zwiers (2013), who ~~propose to~~ question hypotheses testing on future climate ensembles and instead propose to better use "*a simple descriptive approach for characterizing the information in an ensemble of scenarios*"~~instead of the ensemble significance, we also~~. Being conscious about the difficulties which may arise during hypotheses testing, we look at the ensemble spread in form of the interquartile range to assess the robustness of the results, and consult the significance test to support our findings. In many cases, the majority of ensemble members show a change signal in the same direction and the change signal is of a similar order of magnitude as the observed past changes in the preceding section (Figs. ~~?? and ??)~~ 6 and 7). In addition, a comparison to the results of the error estimation using FT-surrogate time series (Tab. 2) yields that the changes are higher then the estimated error. Therefore we conclude that ~~these results do show that there might be possible changes~~ future changes of the succession of cold and wet extremes in winter in some regions ~~and in the following we will discuss the changes of the ensemble mean values~~in Europe can be expected. These changes are, for the significant cases, larger than 50 % for the persistence, larger than 20 % for the recurrence time and larger than 5 % for the entropy. Regarding the findings from our sensitivity analysis (Sect. 3.2) such changes are larger than the natural variability of the descriptors, which hence can be ruled out as the cause. Further, the sensitivity study has shown that such changes in the past occurred concurrently with a strong increase in CO_2 emissions. As explained in Sect. 2.3, the only difference between the model runs for the periods 1971-2001 and 2021-2050 is the CO_2 forcing, thus the most probable reason for these changes in the future is the increase in CO_2 emissions.

Fig. 10 reveals three regions which seem to be particularly susceptible to changes of the dynamics / succession, namely regions 2 (France), 3 (Germany) and 5 (Russia). The persistence changes for all regions and cold and wet episodes are likely to be of longer duration in the future. In regions 2 and 3 (France and Germany) the recurrence time decreases. The consequences of these changes are that these regions will probably experience more and longer cold and wet events in winter. Furthermore, these are less predictable (increase of entropy). The situation is different for region 5 (Russia), here the duration of cold and wet periods probably increases ~~but the events will be fewer (decrease~~ as well, but the number of events stays constant. Thus the system resides for longer times in the non-extreme states (increase in recurrence time)~~but more predictable (decrease of entropy)~~.

The change in number of compound hot and dry extreme days is depicted in Fig. ~~??~~9. Here, the number of compound extreme days varies with the region (although the number of univariate extremes are kept the same). Region 1 (Spain) shows a relatively low number of compound hot and dry days (note: all extremes in this paper are relative), regions 5 and 6 (Russia and Bulgaria) have a high number and also the highest decrease between the two time periods. Except for region 3 (Germany), which shows a slight increase, the number of compound extremes decreases in all regions. However, the change is generally small, $< 10\,\%$. Thus, the observed changes of the descriptors can mostly be attributed to the change in the dynamics and not to a change in the numbers of events, except maybe for regions 4 and 5 (Russia and Bulgaria).

The change signal of the descriptors is pictured in Fig. ~~??~~11. Two regions are most probably susceptible to changes in the dynamics of the hot and dry state, namely regions 1 (Spain) and 6 (Bulgaria). Region 1 shows a ~~quite strong~~ small increase in persistence and ~~recurrence time~~ a quite strong increase in recurrence time (in the order of 20 %) of the hot and dry state~~and a corresponding decrease in entropy~~, the entropy does not change. The hot and dry periods get longer ~~and the system gets more regular, as indicated by the entropy decrease~~but less frequent. Regarding again the sensitivity study (Sect. 3.2) it can be seen that a change of 20 % of the recurrence time in summer (JJA) is at least twice as large as the variability of the recurrence time (about 10 %) from 1900 to 2015 and constitutes a fairly large jump. The situation for region 6 is similar to that of region 1, with an increase in persistence and recurrence time and only a very small change in entropy. In addition, Region 3 (Germany) shows an increase in persistence and a decrease in entropy. This means the episodes will be longer and more regular whereas in region 5 (Russia) the persistence slightly decreases and the recurrence time increases. This implies changes towards shorter and less frequent events.

6 Conclusions and Outlook

The changing climate leads to a change in extreme weather, which comprises several aspects like frequency, duration, intensity etc. On top of these rather linear changes, modifications of the complex succession of extremes can be expected. However, information on the succession or dynamical behavior of climate extremes is rare. Therefore, to extract such information from climate time series we applied a Markov chain analysis on compound extremes, namely cold and wet in winter and hot and dry in summer. We have shown that our climate model ensemble is able to reproduce past dynamics of compound extremes fairly well within acceptable uncertainties. Thus, we have reasonable confidence in the future simulations of this model ensemble. We identified three regions in Europe, which are probably susceptible to a future change in the succession and dynamical behavior of cold and wet extremes in winter. In region 5 (Russia) we detected an increase of the persistence and recurrence time, which means that the probability of staying in the cold and wet state from one day to the next will increase, but the system will take longer to approach this state again. In regions 2 (France) and 3 (Germany), cold and wet episodes become both longer and more frequent. The entropy in these regions also ~~decreases~~ increases in the future, which is counterintuitive, because one would expect that an increase in persistence is related to a decrease in entropy (cf. Eqs. 5 and 7). However, since the entropy (Eqs. 7) does not only consider the compound extreme state but also transitions from this state to the normal state and univariate extreme states, complex interactions can be extracted with the entropy. The impacts of these calculated changes are beyond the scope of this study, and it can only be speculated about possible effects. One could imagine that longer and less predictable cold and wet periods could lead to larger snow chaos regarding traffic and other human life, especially in regions which already experience extreme cold temperatures in winter. Again, these findings suggest that a reordering of the succession of compound extremes could be happening on top of the observed linear changes, as e.g. the temperature increase.

For hot and dry states in summer, the Markov method identified two regions where changes are probable, Spain and Bulgaria. The persistence and recurrence time in regions 1 and 6 (Spain and Bulgaria) both increase in the future, which means that the system resides longer in the extreme state. The entropy ~~decreases, which is expected, because it is easier to predict that an extreme state will follow an extreme state. In this light, the systems are getting more regular. However, any~~ does not change significantly. Any reordering of the succession of extremes has an impact. For instance such changes could be harmful for the local agriculture, because, as explained above, these dynamic changes would occur on top of the known linear increase of e.g. temperatures. Interestingly, in region 6 (Bulgaria) the absolute number of compound hot and dry extremes (Fig.~~??~~9) decreases in the future, but the extreme periods become longer. The changes for region ~~5~~ 3 (Russia) are small ~~for persistence and entropy but larger for recurrence time, which increases. This is probably connected with the decrease of the number of compound events in the future. Thus, it seems~~ but indicate that the region in Russia near Moscow will be less susceptible to dynamical changes of the succession of compound extremes and will additionally experience less compound extremes in the near future.

A number of studies have shown an influence of atmospheric drivers (mostly NAO) and atmospheric blocking patterns on summer as well as winter temperature extremes and generally the temperature variability in Europe (e.g. Photiadou et al., 2014; Sillmann and Croci-Maspoli, 2009). Although the extremes analyzed in these studies were mostly of absolute nature, an analysis of the influence of the same factors on the relative extremes studied in this paper would be very interesting. Using a similar methodology as described in this paper to calculated persistence, recurrence and entropy of time series of e.g. the NAO index in a certain regime could be linked to the descriptors of the compound extreme events.

Areas to apply this method are manifold. Besides the analysis of different dynamical behavior varying on the region and extreme considered, it can be used as a model validation tool. As extremes and especially compound extremes are an important quantity that we want to assess with climate model data, it is necessary for the models to capture the dynamical behavior of these extreme events. As shown in this paper, the models can also project changes of the future dynamical behavior which is ~~are~~ an interesting supplementary information to changes in mean and variability. An example where this could be useful is the decision whether to apply simple or more sophisticated bias correction techniques.

Follow up studies using simulations of other regional climate models and regional climate ensembles for time periods further in the future (e.g. ENSEMBLES, http://ensembles-eu.metoffice.com/, or CORDEX, http://www.euro-cordex.net/, data for the end of the century) would be interesting. For one, this would allow an analysis of whether or not there are significant differences depending on the regional climate model used. In addition, data for the end of the 21st century is available where changes in the descriptors could possibly be larger because the influence of the ~~CO2~~ CO_2 forcing plays a more important role. In this sense, the Markov chain analysis could be useful to identify possible future regime shifts (Scheffer and Carpenter, 2003; Scheffer et al., 2009). Of further interest is an analysis of the dynamical behavior of the driving GCMs as well as the ERA-40 reanalysis dataset since for parts the ERA-40 driven CCLM model runs performed worse in comparison to observations than the CCLM ensemble. This leads to the question whether or not the CCLM model runs compensate for errors in the driving GCMs and are right for the wrong reasons. Comparison of the E-OBS dataset to other regionally defined datasets would also be helpful to evaluate the observational data.

Acknowledgements. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu).Figs.1 and 3 were made using the GMT (Generic Mapping Tools) web-application www.piece-of-earth.net. All other graphics were made using R (R Development Core Team, 2008). We also thank P.Berg and R.Sasse for their contributions to the CCLM-Ensemble. The authors thank members of the IMAGE and RCR sections at NCAR for the fruitful discussions and the three anonymous referees for their helpful comments and suggestions.

References

Byun, H.-R. and Wilhite, D. A.: Objective quantification of drought severity and duration, J. Climate, 12, 2747–2756, 1999.

Christensen, J. H. and Christensen, O. B.: A summary of the PRUDENCE model projections of changes in European climate by the end of this century, Clim. Change, 81, 7–30, 2007.

Collins, W., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C., Joshi, M., Liddicoat, S., et al.: Development and evaluation of an Earth-system model–HadGEM2, Geosc. Model Devel. Disc., 4, 997–1062, 2011.

Daw, C. S., Finney, C. E. A., and Tracy, E. R.: A review of symbolic analysis of experimental data, Rev. Sci. Instrum., 74, 916–930, 2003.

Doms, G. and Schättler, U.: A description of the nonhydrostatic regional model LM, Part I: dynamics and numerics, Tech. rep., Consortium for small-scale modelling, Deutscher Wetterdienst, Offenbach, Germany, 2002.

Durante, F. and Salvadori, G.: On the construction of multivariate extreme value models via copulas, Environmetrics, 21, 143–161, doi:10.1002/env.988, http://dx.doi.org/10.1002/env.988, 2010.

Ebeling, W., Freund, J., and Schweitzer, F.: Komplexe Strukturen: Entropie und Information, B. G. Teubner, 1998.

Gallant, A. J., Karoly, D. J., and Gleason, K. L.: Consistent Trends in a Modified Climate Extremes Index in the United States, Europe, and Australia, J. Climate, 27, 1379–1394, 2014.

Hagemann, S., Arpe, K., and Bengtsson, L.: Validation of the hydrological cycle of era 40, ERA-40 Project Rep. Series, 24, 42 pp, 2005.

Haylock, M. R., Hofstra, N., Tank, A. M. G. K., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006, J. Geophys. Res., 113, doi:10.1029/2008JD010201, 2008.

Hazeleger, W., Severijns, C., Semmler, T., Stefanescu, S., Yang, S., Wang, X., Wyser, K., Dutra, E., Baldasano, J. M., Bintanja, R., et al.: EC-earth: a seamless earth-system prediction approach in action, Bull. Amer. Meteor. Soc., 91, 1357–1363, 2010.

Hill, M., Witman, J., and Caswell, H.: Markov chain analysis of succession in a rocky subtidal community, Am. Nat., 164, E46–E61, doi:10.1086/422340, 2004.

IPCC: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change, Cambridge University Press,Cambridge, UK, and New York, NY, USA, 2012.

Keetch, J. J., Byram, G. M., et al.: A drought index for forest fire control, 1968.

Klein Tank, A., Wijngaard, J., Können, G., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., et al.: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, International journal of climatology, 22, 1441–1453, 2002.

Klok, E. and Klein Tank, A.: Updated and extended European dataset of daily climate observations, Int. J.Clim., 29, 1182–1191, 2009.

Lorenz, E. N.: Deterministic Nonperiodic Flow, J. Atmos. Sci., 20, 130–141, 1963.

Mieruch, S., Noel, S., Bovensmann, H., Burrows, J., and Freund, J.: Markov chain analysis of regional climates, Nonlin. Proc. Geoph., 17, 651–661, 2010.

Nakicenovic, N. and Swart, R.: Special report on emissions scenarios, Special Report on Emissions Scenarios, Edited by Nebojsa Nakicenovic and Robert Swart, pp. 612. ISBN 0521804930. Cambridge, UK, Cambridge University Press, July 2000., 1, 2000.

Neves, M. M.: Geostatistical Analysis in Extremes: An Overview, in: Mathematics of Energy and Climate Change, pp. 229–245, Springer, 2015.

Pachauri, R. K., Allen, M. R., Barros, V., Broome, J., Cramer, W., Christ, R., Church, J., Clarke, L., Dahe, Q., Dasgupta, P., et al.: Climate change 2014: synthesis Report. Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change, IPCC, 2014.

Photiadou, C., Jones, M. R., Keellings, D., Dewes, C. F., et al.: Modeling European hot spells using extreme value analysis, Climate research, 58, 193–207, 2014.

R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org, ISBN 3-900051-07-0, 2008.

Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., and Gratzki, A.: A Central European precipitation climatology–Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS), Meteor. Z., 22, 235–256, 2013.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., and Rafaj, P.: RCP 8.5 – A scenario of comparatively high greenhouse gas emissions, Clim. Change, 109, 33–57, 2011.

Rockel, B., Will, A., and Hense, A.: The Regional Climate Model COSMO-CLM(CCLM), Meteor. Z., 17, 347–348, doi:10.1127/0941-2948/2008/0309, 2008.

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornblueh, L., Manzini, E., et al.: The atmospheric general circulation model ECHAM 5. PART I: Model description, Tech. rep., MPImet/MAD Germany, 2003.

Sasse, R. and Schädler, G.: Generation of regional climate ensembles using Atmospheric Forcing Shifting, Int. J. Climatol., 34, 2205–2217, doi:10.1002/joc.3831, 2014.

Scheffer, M. and Carpenter, S. R.: Catastrophic regime shifts in ecosystems: linkingtheory to observation, Trends Ecol. Evol., 18, 648–656, 2003.

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., van Nes, E. H., Rietkerk, M., and Sugihara, G.: Early-warning signals for critical transitions, Nature, 461, 53–59, doi:10.1038/nature08227, 2009.

Schoelzel, C., Friederichs, P., et al.: Multivariate non-normally distributed random variables in climate research–introduction to the copula approach, Nonlin. Proc. Geoph., 15, 761–772, 2008.

Schreiber, T. and Schmitz, A.: Surrogate time series, Physica D: Nonlinear Phenomena, 142, 346–382, 2000.

Scinocca, J., McFarlane, N., Lazare, M., Li, J., Plummer, D., et al.: The CCCma third generation AGCM and its extension into the middle atmosphere, Atmos. Chem. Phys. Discuss, 8, 7883–7930, 2008.

Sedlmeier, K.: Near future changes of compound extreme events from an ensemble of regional climate simulations, Ph.D. thesis, Karlsruhe, Karlsruher Institut für Technologie (KIT), Diss., 2015, 2015.

Shannon, C. E.: A Mathematical Theory of Communication, Bell System Technical Journal, 27, 623–656, 1948.

Sillmann, J. and Croci-Maspoli, M.: Present and future atmospheric blocking and its impact on European mean and extreme climate, Geophysical Research Letters, 36, 2009.

Simmons, A., Jones, P., da Costa Bechtold, V., Beljaars, A., Kållberg, P., Saarinen, S., Uppala, S., Viterbo, P., and Wedi, N.: Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature, Journal of Geophysical Research: Atmospheres, 109, 2004.

Steinhaeuser, K. and Tsonis, A. A.: A climate model intercomparison at the dynamics level, Climate Dyn., 42, 1665–1670, 2014.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., et al.: Atmospheric component of the MPI-M Earth System Model: ECHAM6, J. Adv. Model. Earth Syst., 5, 146–172, 2013.

Tebaldi, C. and Sansó, B.: Joint projections of temperature and precipitation change from multiple climate models: a hierarchical Bayesian approach, J. Roy. Stat. Soc., 172, 83–106, 2009.

Uppala, S. M., Kållberg, P., Simmons, A., Andrae, U., Bechtold, V., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G., et al.: The ERA-40 re-analysis, Quart. J. Roy. Meteor. Soc., 131, 2961–3012, 2005.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., et al.: The representative concentration pathways: an overview, Clim. Change, 109, 5–31, 2011.

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.

Voldoire, A., Sanchez-Gomez, E., y Mélia, D. S., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., et al.: The CNRM-CM5. 1 global climate model: description and basic evaluation, Climate Dyn., 40, 2091–2121, 2013.

von Storch, H. and Zwiers, F.: Testing ensembles of climate change scenarios for atistical significanceClim. Change, 117, 1–9, 2013.

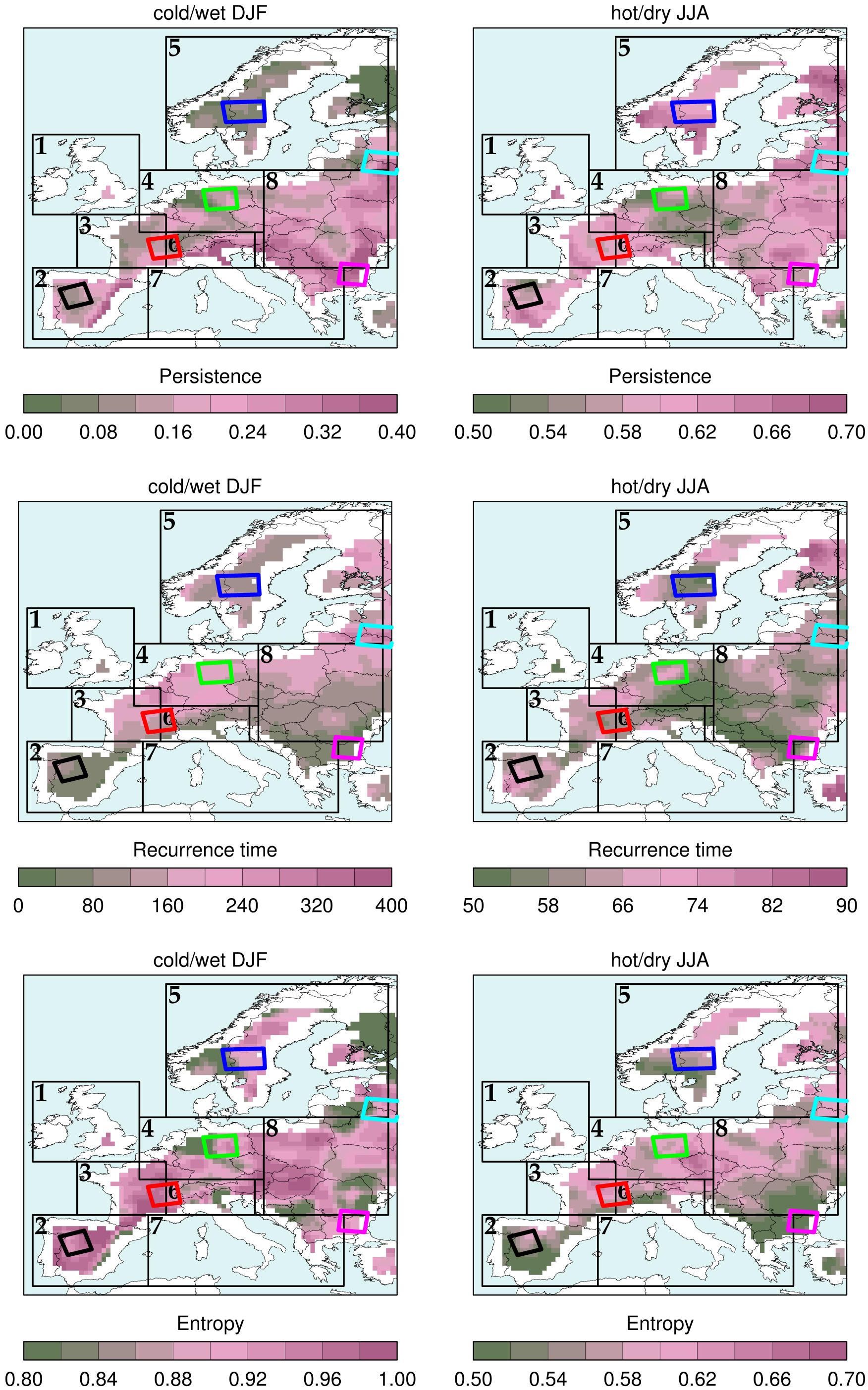

Figure 1. E-OBS descriptors for the reference period (1971-2000). Left side: Descriptors for cold and wet extremes in winter (DJF) (Ta < 10th percentile and Pa > 75th percentile). Right side: Descriptors for hot and dry extremes in Summer (JJA) (Ta > 90th percentile and EDI < 25th percentile). Descriptors were calculated for a moving window over 9 gridpoints and values assigned to the center grid point (see text). Boxes show the Prudence Regions (http://ensemblesrt3.dmi.dk/quicklook/regions.html)

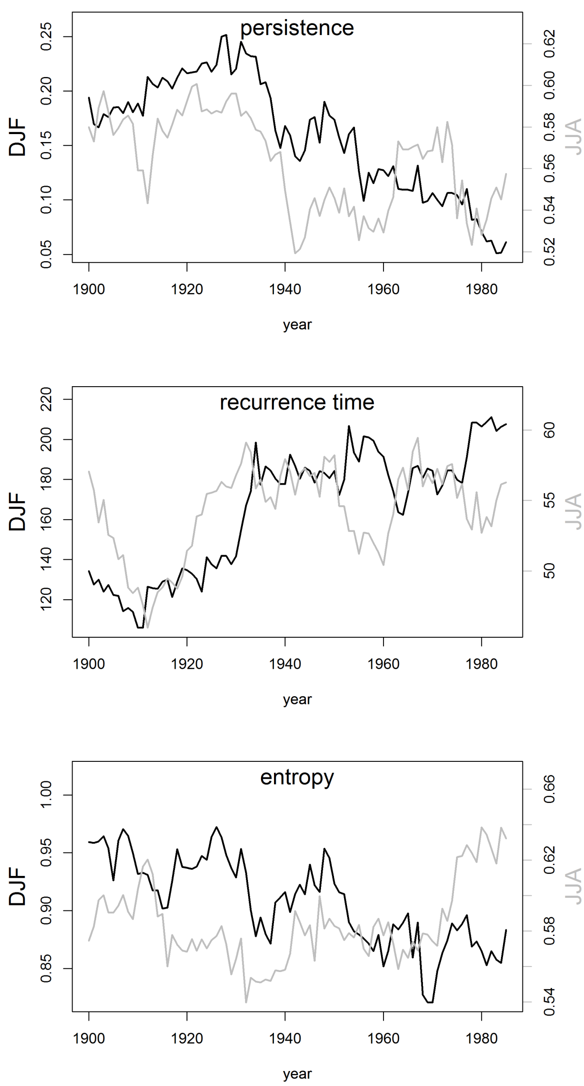

Figure 2. Descriptors for ECA& D station data for running windows over 30-years (values are assigned to the first year of the 30-year time period.) from 1900-2015. Black curve: cold and wet extremes in winter (DJF) (Ta < 10th percentile and Pa > 75th percentile). Gray lines: hot and dry extremes in Summer (JJA) (Ta > 90th percentile and EDI < 25th percentile)

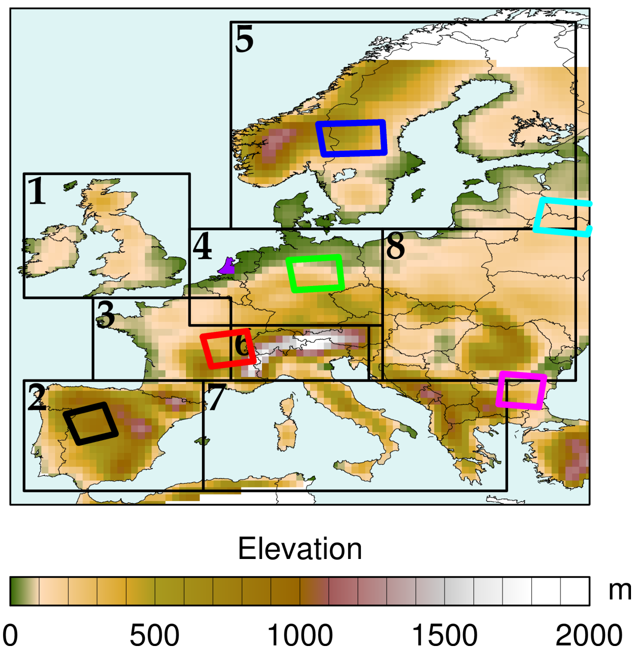

Figure 3. Elevation of the CCLM 50 km Model domain [m]. Boxes mark the six investigation areas, 1:Spain (black), 2:France (red), 3:Germany (green), 4:Scandinavia (blue), 5:Russia (cyan) and 6:Bulgaria (magenta).

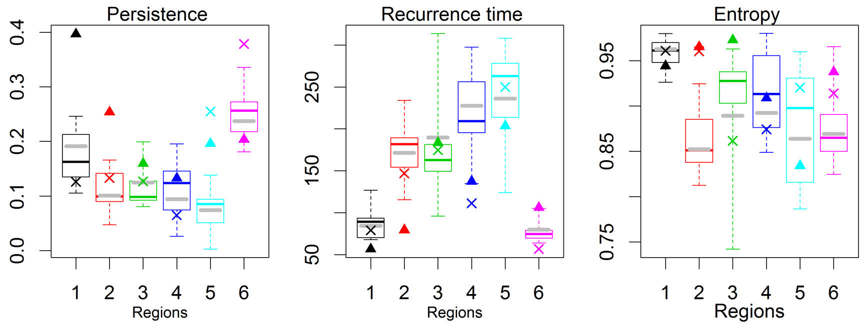

Figure 4. Descriptors for cold and wet extremes in winter (DJF) (Ta < 10th percentile and Pa > 75th percentile) in the reference period 1971-2000 for the 6 investigation areas. Box plots of the CCLM ensemble: box = ensemble median and interquartile range, whiskers = ensemble minimum/maximum, gray bars: ensemble mean, triangles :reanalysis driven CCLM runs, crosses: E-OBS observations. The coloring corresponds to the regions in Fig. 3.

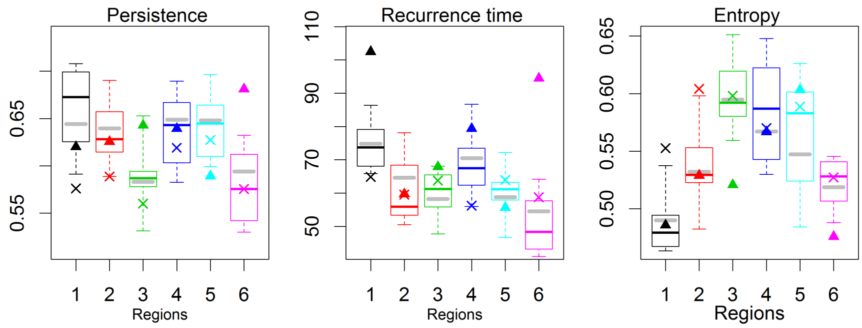

Figure 5. Descriptors for hot and dry extremes in Summer (JJA) (Ta > ~~95th~~ 90th percentile and EDI < 25th percentile) in the reference period 1971-2000 for the 6 investigation areas. Box plots of the CCLM ensemble: box = ensemble median and interquartile range, whiskers = ensemble minimum/maximum, gray bars: ensemble mean, triangles :reanalysis driven CCLM runs, crosses: E-OBS observations. The coloring corresponds to the regions in Fig. 3.

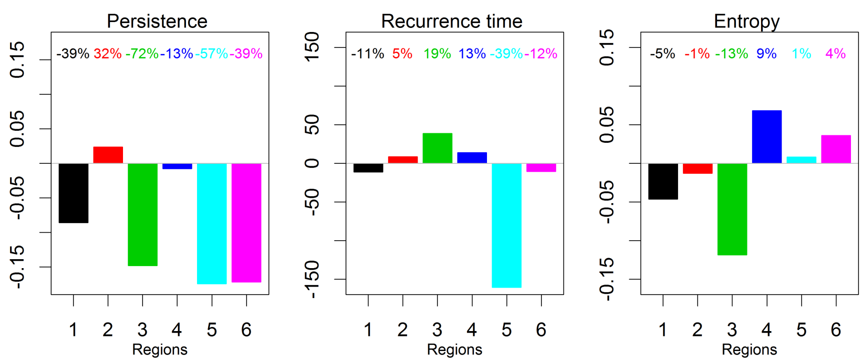

Figure 6. Change signal ~~for~~ of descriptors for E-OBS observations: Cold and wet extremes in winter (DJF) (Ta < 10th percentile and Pa > 75th percentile)~~;~~. Changes between the time periods 1951-1980 ~~vs~~ and 1981-2010. Percentages denote the relative change. The coloring corresponds to the regions in Fig. 3.

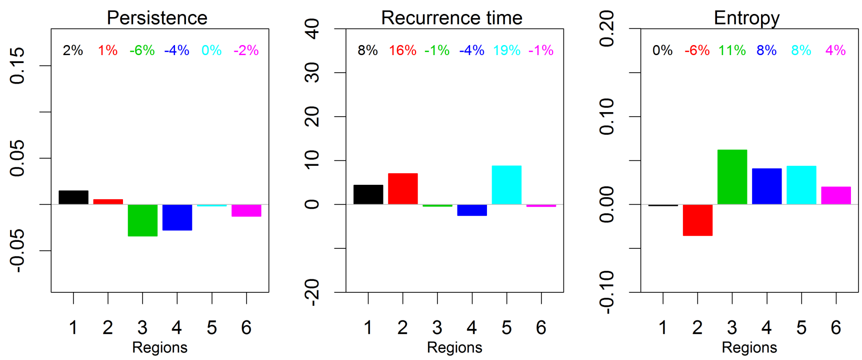

Figure 7. Change signal ~~for~~ of descriptors for E-OBS observations: Hot and dry extremes in Summer (JJA) (Ta > ~~95th~~ 90th percentile and EDI < 25th percentile)~~,~~. Changes between the time periods 1951-1980 ~~vs~~ and 1981-2010. Percentages denote the relative change. The coloring corresponds to the regions in Fig. 3.

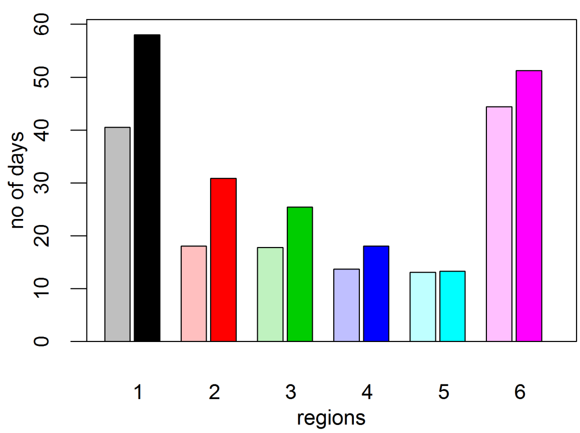

Figure 8. Number of compound cold and wet extremes in winter (DJF) (Ta < 10th percentile and Pa > 75th percentile) , 1971-2000 (light colors) and 2021-2050 (dark colors), ensemble mean. The coloring corresponds to the regions in Fig. 3.

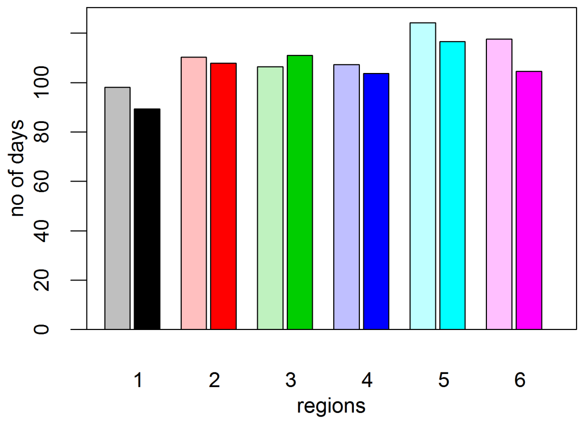

Figure 9. Number of compound hot and dry extremes in summer (JJA) (Ta > ~~95th~~ 90th percentile and EDI < 25th percentile), 1971-2000 (light colors) and 2021-2050 (dark colors), ensemble mean. The coloring corresponds to the regions in Fig. 3.

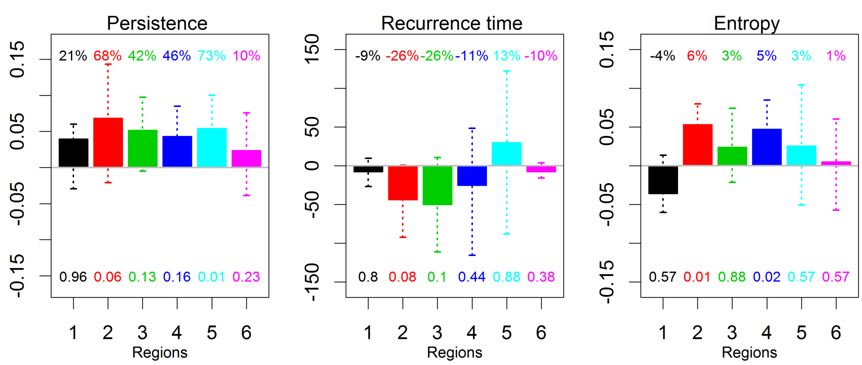

Figure 10. ~~Ensemble mean climate~~ Climate change signal of descriptors for cold and wet extremes in winter (DJF) (Ta < 10th percentile and Pa > 75th percentile)~~,~~. Changes between the time periods 1971-2000 ~~vs~~ and 2021-2050. ~~Points~~ Bars show the ensemble mean (of the individual change signals), whiskers the 75th and 25th quantile, respectively. Percentages above the bars denote the relative change of the ensemble mean, the numbers below the p-value. The coloring corresponds to the regions in Fig. 3.

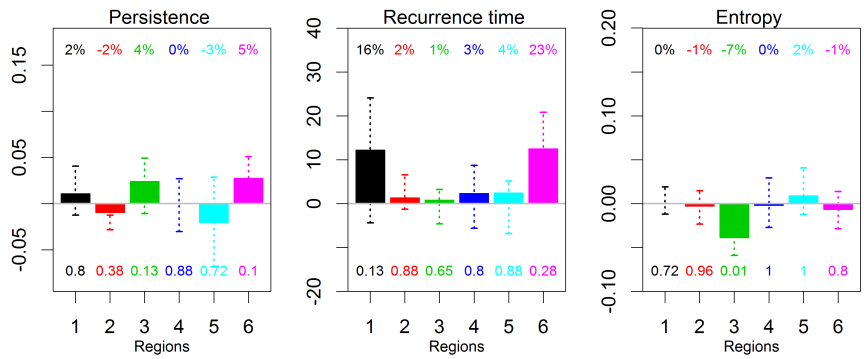

Figure 11. ~~Ensemble mean~~ Climate change signal of descriptors for hot and dry extremes in Summer (JJA) (Ta > ~~95th~~ 90th percentile and EDI < 25th percentile)~~,~~. Changes between the time periods 1971-2000 ~~vs~~ and 2021-2050.. ~~Points~~ Bars show the ensemble mean (of the individual change signals), whiskers the 75th and 25th quantile, respectively. ~~.~~ Percentages above the bars denote the relative change of the ensemble mean, the numbers below the p-value. The coloring corresponds to the regions in Fig. 3.