List of Major Changes

- 1) For the validation of LAERTES (Sect. 4), we changed:
 - (a) the time period TP1b (old: 1950-2017) to the HYRAS period (new: 1951-2006)
 - (b) the investigation areas ME and AL are reduced to the corresponding HYRAS grid cells, indicated by ME* and AL*
 - (c) the estimation of the percentiles is newly done considering only wet days with R>0.1mm
 - (d) we included additional quantities like the linear error in probability space L and a frequency analysis
 - (e) Figure 5 was re-done using perturbed sequences of concatenated ensemble members to reduce the influence of member size on the shape of the curves.
- 2) Figure 8a was wrong (old version), we replaced it with the correct one
- 3) More precise captions of figures and tables in both the manuscript and the supplemental material
- 4) We rearranged the Section 2.2 describing the ensemble data to be more precise and consistent
- 5) We rearranged the introduction
- 6) We adjusted the conclusions accordingly

Point-by-point response to Reviewer #1

Florian Ehmele on behalf of all co-authors October 31, 2019 Update: March 9, 2020

Dear reviewer No. 1,

Thank you very much for your work and the useful and valuable comments that will help to improve the scientific quality of our manuscript. Below you will find your comments given in gray and our responses to the individual points in black. Please also consider our comments to Reviewer 2 as there is some coincidence of the comments and the corresponding answers.

This manuscript addresses the issue of heavy precipitation in RCM simulations. This is a very timely issue with importance for many sciences, which rely on RCM simulations. My fundamental concern with this study is the first conclusion ("Extreme precipitation is well represented in LAERTES-EU."). The same is expressed in the authors' short summary ("The simulations show a good agreement with observations for both statistical distributions and time series of heavy precipitation."). I am sorry, but I just can not see enough support for this crucial statement in the manuscript.

You and also RC2 have the same concerns about this first conclusion. Thinking about this a second time we came along that this statement might be too general. It was meant that heavy precipitation is consistent in all parts of LAERTES-EU and that our results fit in the range of previous studies (e.g. Früh et al., 2010) and also in the range of observations knowing that the used observational data sets have uncertainties as well. We will rewrite this to be more precise what was meant to be stated here. However, we do think that, for instance, the IPCs do support the statement in terms of the statistical distribution of precipitation values. The time series of LAERTES-EU (ensemble mean) is within the range of both analyzed observational data sets, and the ensemble spread covers the observed variability. Please note that it was never intended that LAERTES-EU shows a one-by-one agreement with historical events.

1) The authors state that E-OBS underestimates precip by almost a third. To me, this means that these data are not useful to evaluate the performance of extreme value simulations. As E-OBS is only available for land surfaces, I also find it surprising that the ME box includes parts of the North Sea.

E-OBS has some limitations, like a certain underestimation especially for extremes, but these have been mentioned by several previous studies like Haylock et al (2008) or Hofstra et al. (2009), and mainly appear in a grid point comparison with measurement sites. As these and other studies already used and analyzed E-OBS, we follow their conclusions and did not perform a further analysis on data quality. Keeping the limitations in mind, E-OBS can be useful for evaluation. Unfortunately, there is no other high-resolution daily precipitation data set available that covers entire Europe for a quit long time period. As the focus of this study is on intensive areal precipitation, we think it does not make sense to use single ground based observations that potentially are available for longer time scales, or in terms of the focus on longterm evolution other products like satellite data with a very limited time frame are not helpful and also have limitations. We will add a comment on this situation to the revised manuscript.

The prudence region are defined as regular lat/lon-boxes and therefore cover ocean areas as well. But, in every case ocean grid cells have been set to a missing value in every dataset and therefore, they are not in the results. We will add a sentence on that in the method section for clarification.

2) The evaluation using IPCs is good, but doing this on a highly aggregated level seems to limit the opportunity to really test the simulated precip. Here I would like to see more creative tests such as IPCs for smaller areas and/or IPCs for certain seasons. As the analysis is done now, there is a risk for error compensation.

For the IPCs no aggregation was done. We take all grid point values within the investigation area (e.g. ME) and at all timesteps into account and group them into a histogram giving the probability of occurrence (=IPC). On purpose, we only use all year data and no seasonal differences as the paper would have become to long doing a seasonal analysis for every part of it. This study was meant to be an introduction to LAERTES-EU and some long term investigations on the upper part of the precipitation distribution.

For a more appropriate evaluation of LAERTES-EU, we will follow Reviewer 2 and include some further analysis using other methods and quantities. In particular, we will add a frequency analysis.

3) I am sorry, but I do not see how the Q-Q plots help to evaluate the performance for heavy precip. If anything, the total precip is evaluated. But even then, comparing cumulative values introduces a spurious correlation, and on top of that, R2 is no suitable measure as a value of one does not ensure a 'perfect' model. May be I miss something here, but I find this analysis not convincing.

We will restructure the evaluation part and will introduce additional skill measures in the evaluation (see also answers to RC2), for example, the linear error in probability space L (e.g. Potts et al., 1996)

4) The 99% of precip (=around 10 mm) is not really 'heavy precipitation.'

Technically, this is correct. However, the focus of this study is intensive areal precipitation, which is related to widespread flooding along the great major river networks. For a single grid cell, 10 mm is no big deal but 10 mm on average over a large area such as the Rhine catchment or the entire PRUDENCE region is quite a lot. Maybe the term 'heavy precipitation' somehow is irritating at this point. We will include a clarification on that at the beginning of the revised manuscript. Furthermore, please see also the report of Reviewer 2. We will change the percentile calculation to wet days only, as currently dry spells are included.

Minor comments:

P4L115: "more or less independent simulations". This needs to be clarified. In some respect, these simulations might be independent, but as the same RCM is used, the simulations obviously are dependent!

We agree that this formulation is inept and we will remove it. What was meant is that the temporal evolution of the day-to-day weather in hindcasts is independent after a few weeks. The statement did not refer to the model system. The ensemble does not cover the full range of uncertainty, namely the model uncertainty. But, in the context of the paper we regard this as an advantage, since the data set is homogeneous over time due to the consistent model setup.

P6L149: does this mean there was a bias correction? Were extreme precipitation simulations affected by this at all? I assume not but would like to get some clarification.

The dry-day adjustment only corrects the number of days without precipitation (R<0.1mm/day) in the model as RCMs tend to produce too much days with very small rainfall amounts (drizzle effect; Berg et al., 2012). The absolute values (R>=0.1mm) are not affected. A bias correction, meaning an adjustment of the absolute precipitation values by e.g. a quantile mapping, was not performed at this stage. In a consecutive study (Kautz et al., planned submission in summer 2020), a specific application of LAERTES-EU for hydrological issues will be presented for which such a bias correction is mandatory. For any other application, a reduction of the drizzle effect has to be done anyway.

P7L197ff: I can see the argument that GCMs underestimate heavy precip, but the same argument should, although to a smaller degree, apply to RCMs. So, what is the physical reason that RCMs 'tend to overestimate precipitation intensities'?

Two effects are of relevance at this point and which act together. The limited time period of observations results in unknown distributions, especially at the heavy tail. In a dataset of 65 years, extreme events with return periods of 100 years or more are not represented in a statistically robust way. The RCM has a physical background when calculating precipitation amounts which makes it possible to reach higher than observed values. Furthermore, the huge number of simulations allows for a more robust estimate of the high-intensity tail of the distribution, whereas the observations display only a few single events in this range.

Point-by-point response to Reviewer #2

Florian Ehmele on behalf of all co-authors October 31, 2019 Update: March 9, 2020

Dear reviewer No. 2,

Thank you very much for your work and the useful and valuable comments that will help to improve the scientific quality of our manuscript. Especially your suggestion on how to implement the comments to the paper are very useful. Below you will find your comments given in gray and our responses to the individual points in black. Please also consider our comments to Reviewer 1 as there is some coincidence of the comments and the corresponding answers.

This paper analyzes long-term trends of heavy precipitation in multiple dynamically downscaled simulations for the historical period and the near future over Europe. The different sets of simulations are validated against gridded observations and tested whether they can be combined to a large ensemble for the detection of trends in the historic time period. This paper is relevant in terms of assessing the possibility of combining various simulations from the same RCM with varying driving data. As well as, the detection of trends within the historic time period.

General comments:

1) Please be more clear about what you are showing in the figures. In most cases it wasn't clear to me if you are showing the ensemble mean or a metric with pooled data from the entire ensemble.

Thank you for this feedback. We will change the figure caption to be more precise and accurate to become more clear. Therefore we will also include the related minor comments you wrote below.

2) Several sections need more clarification on what was analyzed and for what spatial extend and aggregation.

Going through your major and minor comments below and include them into the new version of the manuscript, we think this will clarify a lot of points within the text. Please see our detailed comments to the specific points below.

Major comments:

... Model evaluation:

1) I have concerns with the comparison of E-OBS, CCLM and HYRAS over the sub-region AL. It is not clear to me whether the comparison was only performed for the HYRAS grid cells, which cover a substantially smaller area than E-OBS and CCLM, or whether E-OBS and CCLM represent the entire AL domain compared to a much smaller area in HYRAS. On P8-L204f you state this concern yourself '[...] which might be a reason for the vanished differences between E-OBS and HYRAS and the resulting specious deviations to the RCM'. Did you compare the three datasets for the HYRAS grid cells only? If not please do so.

This is a crucial comment, thank you for that. Checking our data we found out that the analysis for the AL region indeed was performed for the entire region for CCLM and E-OBS but HYRAS only for available grid cells. We will fix this in the new manuscript

version and also double check our results for the ME region to be done on HYRAS grid cells only. We will name the sub-areas of ME and AL, in which HYRAS is available, with an asterisk (ME* or AL*), to be clear.

2) Further, you state that '[...] by taking into account all grid points and all time steps within the investigation are (P6-L151)', does this mean that for both ME and AL you have included ocean grid cells in the spatial average of the RCM data? For both domains gridded observational datasets are only available over land. Please clarify this, and in any case ocean grid cells were included remove them from the comparison.

In every case, ocean grid cells have been set to a missing value in every data set and therefore, they are not in the results. We will add a sentence on that in the method section for clarification.

2) In P8-L212f you state that '[. . .] HYRAS was aggregated to the E-OBS/RCM grid [. . .]'. However, you first mentioned this here for the Q-Q plots, so can I assume that the IPC's in Figure 2 are also based on aggregated HYRAS data? Please clarify this and if the aggregation of HYRAS applies to all related analysis then please move this detail to the methods section.

The HYRAS data have first been aggregated to the E-OBS/RCM grid of 0.22° resolution for all type of analysis in this study. We will clarify this by moving the corresponding statement more to the front of the manuscript into the method section as you have requested.

3) Further, if I understand correctly the evaluation is based on the TP1b time period (1950-2017), however the HYRAS data is only available for the period 1951-2006. Please comment on why the analysis wasn't based on the shorter HYRAS time period. I would recommend doing the analysis for 1951-2006.

In this case, you are right with the different time periods. We assume that there will be only small changes when reducing TP1b to the HYRAS period 1951-2006, but nevertheless, we will fix this for all analyses in terms of consistency.

4) The evaluation on such a highly aggregated level poses a risk of error compensation. It might be better to do the evaluation for each grid cell first (e.g. calculating the RMSE) and afterwards averaging the error metric.

We will restructure the evaluation part and will introduce additional skill measures in the evaluation. However, as the focus of this study is on intensive areal precipitation, we do not want to add too detailed grid point based analyses and take a deeper look in the spatial mean precipitation statistics. Therefore, we add additional quantities like the linear error in probability space (e.g. Potts et al., 1996), and we perform a frequency analysis on different time scales.

5) Based on the concerns above, I don't really agree with your first point in the conclusion 'Extreme precipitation is well represented in the LAERTES-EU [P20-L352]'.

You and also RC1 have the same concerns about this first conclusion. Thinking about this a second time, we came along that this statement might be too general. It was meant that heavy precipitation is consistent in all parts of LAERTES-EU and that our results fit in the range of previous studies (e.g. Früh et al., 2010) and also in the range of observations knowing that the used observational data sets have uncertainties as well. We will rewrite this to be more precise.

. . . Added-Value:

6) Regarding your conclusions on the IPCs showing '[. . .] a clear added value of RCM data compared to coarser global models'. From that one figure I don't really see the added value, since you haven't compared the driving GCM with RCM simulation. You have compared the IPC's to the 20CR reanalysis dataset. Because of the spatial averaging over such a large area, it might be that the trends in the GCM and RCM might be very close to each other.

To substantiate this statement we will include the IPCs of the MPI-ESM model in Fig. 2. At least we will include the IPC of data block 1 which used the LR version of MPI-ESM, and data block 3 which used the HR version as global forcing.

... CC-scaling

7) Your conclusion on the trends following the CC-scaling in your conclusion [P21-L379ff] are flawed. If you make a statement like this, please perform the temperature scaling with the LAERTES-EU temperature data and not by relating the precipitation change to a temperature approximation from another study. Please see Kröner et al (2017) and Pfahl et al (2017) for other effects than thermodynamics. Kröner et al (2017), Climate Dynamics, https://doi.org/10.1007/s00382-016-3276-3 Pfahl et al (2017), Nature Climate Change, https://doi.org/10.1038/nclimate3287

We agree with the reviewer that a relationship should be established using LAERTES-EU temperature data. We will do some brief analysis with the block 1 & 3 temperature data and put them into the context of the already cited studies on 20th century temperature changes. Then the argumentation should be more consistent and reasonable. This will be concentrated in the conclusions.

Minor comments:

P2-L25f: see also Zhang et al (2017) for a discussion on CC scaling Zhang et al (2017), Nature Geosciences, DOI: 10.1038/NGEO2911

P2-L27-34: Please add some more recent literature on this topic, e.g. Fischer and Knutti (2016), Nature, DOI: 10.1038/NCLIMATE3110 Alexander (2016), Weather and Climate Extremes, http://dx.doi.org/10.1016/j.wace.2015.10.007 Barbero et al (2017), GRL, doi:10.1002/2016GL071917

P2-L45f: Please add a view more recent studies on trends in European floods. E.g.: Blöschl et al (2017), Science, DOI: 10.1126/science.aan2506; Blöschl et al (2019), Nature, https://doi.org/10.1038/s41586-019-1495-6

P2-L49f: Connection of Heavy Precipitation over central Europe and cyclones, see also Hoffstätter et al (2017), Int. Journal of Climatology, https://doi.org/10.1002/joc.5386

P2-L55f: Also see van der Wiel et al (2019) and Martel et al (2019) for the added value of large ensembles for flood risk or return periods of heavy precipitation van der Wiel et al (2019), GRL, https://doi.org/10.1029/2019GL081967 Martel et al (2019), Journal of Climate, https://doi.org/10.1175/JCLI-D-18-0764.1

Thank you very much for these recent references. We will go through them and decide which one and where to include them in the new version of the manuscript.

P4-L114f: Please elaborate more on what you mean by '[. . .] more or less independent simulations'

We agree that this formulation is inept and we will remove it. What was meant is that the temporal evolution of the day-to-day weather in hindcasts is independent after a few weeks. The statement did not refer to the model system. The ensemble does not cover the full range of uncertainty, namely the model uncertainty. But, in the context of the paper we regard this as an advantage, since the data set is homogeneous over time due to the consistent model setup.

P5-L139: What do you mean by un-initialized? Please clarify this for the reader, that by initialized you mean initialized by observational(-like) salinity and other variables, whereas the un-initialized data originate from a normal CMIP5 historical simulation. I had to go to Marotzke et al (2016) to understand what was meant by this.

We agree that this has been explained insufficiently and will state it more clearly in the revised manuscript.

P6-L157f: Are the 99th and 99.9th percentile based on all days or wet-days only? If you want to look at heavy precipitation it might be better to look at wet days only. Like this the values would not be affected by the dry-day adjustment as much. Further, it is not clear to me if you have first spatially aggregated and then calculated the percentiles, or the other way around. Please comment on whether you think that this will make a difference to your results. This could maybe also solve your concerns on P15-L282f '[. . .] an overestimation of precipitation [...], could be a result of missing data for the applied dry-day correction.'

The percentiles were estimated using all days including dry ones. As we focus on heavy precipitation we agree with you that using wet days only would be more appropriate. Nevertheless, the uncertainty in the first half of the century will remain. The dry-day correction adjusts the number of days without precipitation (R<0.1mm) solely and not the values themselves (R>=0.1mm). This means that the dry-day correction effects the percentiles anyway in some case. But, in order to get a more thorough analysis of heavy precipitation, we will change to a wet days only calculation. Regarding your second point, we first did a spatial aggregation of precipitation to receive the areal precipitation and than calculated the percentile of this spatial mean values which are of deep interest in this study. The other way round we would get a spatial mean value of the percentile which is more relevant for a 2D analyses and related spatial variability giving local effects. We will include some sentences on that in the manuscript.

P7-L189: Could you briefly comment on why you chose the old 1961-1990 period as your reference climatology.

A couple of studies (e.g. Cahill et al., 2015 or Folland et al., 2018) showed that the climate change signal at least for global mean temperature is significantly increased since the early 1980s, which is to a lower degree applicable for Europe, too (Folland et al., 2001). Therefore, using the time period 1981-2010 would possibly include a strong changing signal to the analysis. Using 1961-1990 reduces the influence of these effects as this period shows more stable conditions to a certain degree. Doing so, there is more room for the interpretation of the future projection instead of comparing them to the directly preceding time period.

References:

Cahill et al. (2015), DOI: 10.1088/1748-9326/10/8/084002 Folland et al. (2018), DOI: 10.1126/sciadv.aao5297 Folland et al. (2001), DOI: 10.1029/2001GL012877 P9-L218f: Your conclusion to Table 2 stating that there is a higher correlation when driven by MPI-ESM-HR versus lower resolution MPI-ESM-LR is technically correct, however the differences are so marginal that I find it difficult to attribute the differences to resolution of driving data. Especially, when not only the resolution is different in the HR and LR simulations, but also the initialization. Maybe add a short sentence ['However, differences are only marginal.'].

Thanks for that comment. Yes, the differences are marginal but the differences between the LR and HR blocks are larger than those within the LR blocks or within the HR blocks. We will include a statement such as the suggested one to the manuscript.

Chapter 4.3: This is a nice analysis that shows the benefit of large ensembles, however since you are not looking at return values afterwards it could be nice to highlight another strength of large ensemble namely isolating the forced response from internal variability. Since you are looking at trends and variability this could be a better fit. But this is just a suggestion to improve the flow of the paper. Like a said it is a nice analysis as is.

We decided to use return values in this case as it easy to estimate and on a statistical perspective the amount of data has a significant influence on the estimates. Although you like the presented analyses, we would like to change Fig. 5 a little bit. For this particular figure the simulations were put together starting with block 1 simulation 1 and ending up with the last simulation in block 4. Doing so the shape of the curves strongly depend on the length of the single simulation runs. Therefore, we want to change the figure using the mean values of 100 random combinations of the simulation runs. The values of the signal-to-noise ration will not change that much and the given statements remain valid.

Figure 8a: Shouldn't there be also some more positive anomalies in the climTP period? Did I miss something? Because if you base the annual anomalies on this period, shouldn't you be having positive and negative anomalies within this period?

We are sorry, but unfortunately there was a wrong figure included at this point. Of course there should be and there are positive and negative anomalies in the climTP period. We replace the current plot with the correct one.

Figure 9: Nice plots! Thank you!

P22-L396f: '[. . .] can be used as input for hydrological modeling'. In general, yes and especially when looking at higher return levels of floods. However, as mentioned a few lines above this ensemble is restricted by temporal homogeneity, which can play a very important role in hydrology.

In general, we agree with that and it definitely makes sense when investigating historical events, trends, flood frequency, and so on. In a particular application and as mentioned in the following sentence, LAERTES-EU serves as stochastic weather generator which leads to quasi stochastic hydrological simulations covering the internal climate variability and also a wider range of values occurring. For such statistical applications, LAERTES-EU can be used to get robust hydrological statistics, too. For this specific case, it is necessary to do a bias correction to avoid too high discharges as a consequence of an overestimation of precipitation. This application as

well as the bias correction will be part of a consecutive study (Kautz et al., planned submission in spring 2020).

Technical Corrections:

Table 1: Projections for the period 2020-28 are missing

The projections are included in block 4 as the given year stand for the initialization years of the decadal simulations meaning an initialization in 2018 includes data for 2019-2028. We will change Table 1 accordingly so that it is clear that 'period' means the covered years.

Figure 2, 3: Please add the years of the period ([. . .] TP1b (1950-2017)). I had to go back and look for the TP1b definition. But I would anyway suggest changing the period to 1951-2006 (see major comments).

Both comments will be implemented in the revised version.

P3-L69: Typo ('Regionla', 'Regional') P5-L128: grammar (replace 'it' with 'the') P20-L342: grammar ('estimate' instead of 'estimated') Thanks! We will fix that.

Figure 4: Please clarify what the RCM spread is. I assume it to be Min-May, right?

That's true. The RCM spread means the range between the minimum and maximum occurred value of the displayed variable. We will include a short clarification into the text.

Long-term Variances of Heavy Precipitation across Central Europe using a Large Ensemble of Regional Climate Model Simulations

Florian Ehmele¹, Lisa–Ann Kautz¹, Hendrik Feldmann¹, and Joaquim G. Pinto¹

¹Institute of Meteorology and Climate Research, Department Troposphere Research (IMK–TRO), Karlsruhe Institute of Technology (KIT), Hermann–von–Helmholtz–Platz 1, 76344 Eggenstein–Leopoldshafen, Germany. **Correspondence:** Florian Ehmele (florian.ehmele@kit.edu)

Abstract. Widespread flooding events are among the major natural hazards in <u>Central central</u> Europe. Such events are usually related to intensive, long-lasting precipitation <u>over larger areas</u>. Despite some prominent floods during the last three decades (e.-g. 1997, 1999, 2002, and 2013), extreme floods are rare and associated with estimated long return periods of more than 100 years. To assess the associated risks of such extreme events, reliable statistics of precipitation and discharge are required. Comprehensive observations, however, are mainly available for the last 50–60 years or less. This shortcoming can be reduced

using stochastic data sets. One possibility towards this aim is to consider climate model data or extended reanalyses.

This study presents and discusses a validation of different century-long data sets, a large ensemble of decadal hindcasts, and also projections predictions for the upcoming decade - Global reanalysis combined to a new large ensemble. Global reanalyses for the 20th century with a horizontal resolution of more than 100 km have been dynamically downscaled with

- 10 a regional climate model (COSMO–CLM) towards a higher resolution of 25 km. The new data sets are first filtered using a dry–day adjustment. The simulations show a good agreement with Evaluation focuses on intensive widespread precipitation events and related temporal variabilities and trends. The presented ensemble data is within the range of observations for both statistical distributions and time series. Differences mainly appear in areas with sparse observation data. The temporal evolution during the past 60 years is well captured. The results reveal some long-term variability with phases of increased
- 15 and decreased heavy precipitation precipitation rates. The overall trend varies between the investigation areas but is mostly significant. The projections predictions for the upcoming decade show ongoing tendencies with increased precipitation for upper percentiles areal precipitation. The presented RCM ensemble not only allows for more robust statistics in general, in particular it is it is also suitable for a better estimation of extreme values.

1 Introduction

5

20 Ongoing climate change affects not only the global scale but also impacts the regional climate. Regarding air temperature, there is a more or less clear trend in the recent past, which reveals a clear anthropogenic signal. However, various climate simulations show distinct spatial differences for precipitation trends, especially for heavy precipitation (e. g. Moberg et al., 2006; Zolina et al., 2008; Toreti et al., 2010). What is known is a theoretical increase of the water vapor eapacity according to the Clausius-Clapeyron (CC) equation of about 6–7% per degree of temperature increase

- 25 (e. g. Trenberth et al., 2003; Berg et al., 2009). For instance, Lenderink et al. (2011), Berg et al. (2013), or O'Gorman (2015) showed that this CC rate can be surpassed up to a factor 2 (Super-Clausius-Clapeyron scaling). In contrast, Stephens and Ellis (2008) found a change of precipitation below the theoretical CC rate. Nevertheless, the CC rate generally thought to be a good proxy for future precipitation projections (Westra et al., 2013).
- Easterling et al. (2000) showed that a linear trend in heavy precipitation varies for different countries and depends also on
 the considered time period. Moberg and Jones (2005) evaluated observational data from about 80 rain gauges in central and western Europe for the time period 1901–1999 revealing an increase in extreme winter precipitation. A recent (e.g. Moberg et al., 2006; Zolina et al., 2008; Toreti et al., 2010). A review of observed variability and trends in extreme climate events states that it is difficult to find significant relations between the greenhouse gas-enhanced climate change and increases or decreases in extreme precipitation events (Field et al., 2012). This is attributed to their rare occurrence, the general
- 35 high spatial variability of precipitation, and due to a lack of long-term high-quality observations. Feldmann et al. (2013) found an increase of both areal mean precipitation

Magnitude and sign of heavy precipitation trends strongly depend on various factors such as the regarded area or the considered time period (e.g. Easterling et al., 2000). Global tendencies towards more intense precipitation throughout the 20th century were revealed, for example, by Donat et al. (2016). Varying regimes between summer and winter season also account

- 40 into precipitation trends. For example, Moberg and Jones (2005) found an increase in winter precipitation across central and western Europe between 1901 and extremes in central Europe in order of 5–10% which will continue with almost same magnitude for the next decades. Moreover, the 1999, while Pal et al. (2004) found a decrease in summer precipitation for the period 1951–2000. Dittus et al. (2016) found an increasing trend between 1951 and 2005 in extreme total precipitation amounts for Europe in GCM simulations (CMIP5). Similar trends were found in global reanalyses (e.g. ERA–20C, Poli et al., 2016),
- 45 but not in observations. In contrast, Primo et al. (2019) found positive trends for two ground-based observational stations in Germany using extreme precipitation indices.

Model resolution is another crucial factor. The use of high resolution regional climate models (RCM) instead of global data sets revealed a more detailed and orographically related spatial structure of the precipitation fields and trends - Global tendencies towards more intense precipitation throughout the 20th century were also revealed by Donat et al. (2016).

- 50 In summary, these studies partly document contrasting results. Following Field et al. (2012), this can have different reasons. One major point are the underlying choice of data sets (model runs, reanalysis, and/or observations). The definition-(e.g. Feldmann et al., 2013). An increase of both areal mean precipitation and extremes in central Europe in order of 5–10% was found in RCM simulations by Feldmann et al. (2013), which will continue with almost same magnitude for the next decade. Differences in precipitation trends also stem from varying definitions of extreme events varies between such as cer-
- 55 tain thresholds, percentile-based indices, or return periods (e. g. Maraun et al., 2010). Other crucial points are that different time periods and areas were investigated as well as different model resolutions(e.g. Maraun et al., 2010). While most of these studies show trends in daily precipitation, just a few deal with sub-daily trends. Barbero et al. (2017), for instance, compared trends in sub-daily and daily extremes. Although significant increasing trends were found for both time ranges, trends in daily extremes are better detected than in sub-daily extremes.

- 60 Spatially extended intensive rainfall events are frequently related to widespread flooding along the main river networks of central Europe causing major damage in the order of several billion euro (EUR) per event (e.g. Uhlemann et al., 2010; Kienzler et al., 2015; Schröter et al., 2015; MunichRe, 2017). Mudelsee et al. (2003) investigated the trends in the occurrence of extreme floods related to heavy precipitation events along the Oder and Elbe rivers. They found a decrease for winter floods in both river catchments, while there seems to be no significant trend for summer floods. In contrast,
- 65 Dittus et al. (2016) found an increasing trend between 1951 (e.g. Uhlemann et al., 2010; Kienzler et al., 2015; Schröter et al., 2015; MunichRe, 2017). A prominent example of such an extreme and 2005 in extreme total precipitation amounts for e.g. Europe in global climate model simulations (CMIP5). Similar trends were found in reanalyses (e.g. ERA–20C, Poli et al., 2016), but not in observations. Moreover, Mudelsee et al. (2004) and Nissen et al. (2013) highlighted a strong dependency of central European flood events on the specific weather pattern of cyclone
- 70 pathway "Vb" like the severe flood event of 2002 devastating event is the flood in 2012 along the rivers Elbe and Danube (Ulbrich et al., 2003a, b). Such outstanding events are by definition extremely rare, which makes the risk estimation difficult or althe limited most impossible due to time period with available area-wide observations (e.g. Pauling and Paeth, 2007; Hirabayashi et al., 2013). Nevertheless, the estimation of flood risk and related trends for (e.g. Pauling and Paeth, 2007; Hirabayashi et al., 2013). However, trend analyses of such extreme events and the related risks
- 75 <u>during</u> the past and <u>for</u> the future are of great importance for insurance purposes or flood protection (e.g. Merz et al., 2014; Schröter et al., 2015; Ehmele and Kunz, 2019).
 (e.g. Merz et al., 2014; Schröter et al., 2015; Ehmele and Kunz, 2019).

A possible way of dealing with the unsatisfactory data availability are century-long simulations using climate models (e. g. Stucki et al., 2016) (e. g. Stucki et al., 2016) or stochastic approaches 80 (e. g. Peleg et al., 2017; Singer et al., 2018; Ehmele and Kunz, 2019).

Several previous studies have investigated long-term trends and variability of extreme precipitation using century-long reanalysis data sets. For instance, (e.g. Peleg et al., 2017; Singer et al., 2018; Ehmele and Kunz, 2019). The currently used GCMs were found to be in good agreement with the available but limited observations (Fischer and Knutti, 2016). Brönnimann et al. (2013) or Brönnimann (2017) analyzed historical extreme events using century-long reanalysis data sets and concluded that

- 85 the quality of the reanalysis reanalyses strongly depends on the number and type of the assimilated observations, mainly sea level pressure and monthly mean sea surface temperature. The investigated historical events were reproduced, but the magnitudes were underestimated. A possible reason is the decreasing number and quality of observations in the early century and therefore, a lack of assimilation data. The suitability of reanalysis data to investigate extreme precipitation for England and Wales was investigated by Rhodes et al. (2015). While time series of daily precipitation totals are well represented in both data
- 90 sets, timing errors of heavy precipitation events were identified as one of the major problems. Stucki et al. (2012) investigated historical flooding events in Switzerland and indicate that the reanalyses underestimate precipitation in Switzerland which may result from the insufficient representation of the alpine topography. In addition, The timing and the exact location of heavy precipitation were also found to be inaccurate.

As shown by van der Wiel et al. (2019) or Martel et al. (2020), large ensembles can have an added values for flood risk
estimation and for the calculation of return periods of heavy precipitation. van der Wiel et al. (2019) found a clear benefit in using an ensemble approach for the estimation of changes in hydrological extremes including compound events compared to traditional approaches. Martel et al. (2020) found similar results, namely a reduction in the projected return period of 100-year annual maximum precipitation with the different ensembles, albeit having different model structures and resolutions. Furthermore, it was emphasized that a higher resolution is advantageous to predict climate change signals over complex terrain.
Other studies also highlighted the improvements of using high resolution RCMs for the investigation of climate extremes

100 Other studies also highlighted the improvements of using high resolution RCMs for the investigation of climate extremes (e.g. Feser et al., 2011; Feldmann et al., 2008, 2013; Schewe et al., 2019), especially over complex terrain (e.g. Torma et al., 2015).

The studies mentioned above document partly contrasting results and demonstrate the challenges arising when dealing with extreme precipitation and related phenomena. In this study, a set of different realizations with one RCM is used and

105 combined to the new ensemble LAERTES-EU (LArge Ensemble of Regional climaTe modEl Simulations for EUrope), which can be used for more profound statistical analyses. Basis is the global reanalysis data set 20CR (Compo et al., 2011), which was dynamically downscaled for Europe. Several studies highlighted the improvements of using high resolution RCMs for the investigation of climate extremes (e. g. Feser et al., 2011; Feldmann et al., 2008, 2013; Schewe et al., 2019), especially over complex terrain (e. g. Torma et al., 2015). LAERTES-EU consists of a handful of 20th century reanalysis data sets and

- 110 a large ensemble of decadal hindcast simulations mainly for the second half of the century. Although all simulations were performed with the same RCM version and set-up, LAERTES-EU is a combination of different external forcings, boundary conditions, and/or assimilation. Projections Predictions for the upcoming decade will round up our analysis. The investigative focus lies on heavy precipitation daily values of intensive areal precipitation which can be associated with major flood events in central Europe. As demonstrated for example by Schröter et al. (2015), severe flood events along the major river networks
- 115 in central Europe are related to long-lasting and widespread precipitation events of mainly stratiform origin with embedded convective precipitation. Typically, intensities do not reach the most extreme rates of the distribution but are characterized by high spatial mean values.

LAERTES-EU is validated in terms of coincidence with observations, regarding temporal variability, statistical distributions, and possible long-term trendsand temporal variability.

- 120 \therefore The following research questions will be addressed.
 - (1) How well is extreme areal precipitation represented in the RCM ensemble LAERTES-EU?
 - (2) What is the added value are potential benefits of LAERTES-EU compared to other available data sets?
 - (3) Which temporal evolution and variability of extreme areal precipitation over central Europe manifest during the pastand what are the differences between the simulations and observations?
- 125 (4) Which tendency is expected for the upcoming decade?

A better interpretation of RCM data and a more profound understanding of extreme <u>areal</u> precipitation may have several applications such as risk assessments. However, Although being relevant, we do not handle the potential mechanisms behind temporal variances and trends as well as spatial and seasonal differences are not part of this paper and will be discussed in continuative studies as this goes beyond the scope of this study.

130

This paper is structured as follows: The data sets which were used in this study are introduced in Sect. 2. Section 3 sums up the methods used for the analysis and the validation. In Sect. 4 LAERTES-EU is validated with observations for a reference period. The investigation of temporal variabilities and trends is given in Sect. 5. Finally, Sect. 6 gives a summary and lists our main-conclusions.

2 Data sets

135 Two different types of data sets are applied in this study: gridded precipitation data based on observations and , partly centurylong , climate model simulations (LAERTES-EU). The observational data sets are primarily available for the second half of the 20th century and serve as reference data for the validation of the ensemble. For validation Furthermore, we compare LAERTES-EU with the global forcing global model and also with the global reanalysis data set of 20CR (Compo et al., 2011)as well, which were used as initial data for some of the RCM-simulations.

140 2.1 Observations

The main reference for this study is the European observational data set E–OBS version v17 for including daily precipitation (Haylock et al., 2008; van den Besselaar et al., 2011) is a gridded data set with a horizontal resolution of 0.22° (≈ 25 km), covering the years 1950 to 2017. This version shows some improvements towards older versions, since updated algorithms and new stations have been included in some areas (e.-g. for Poland). The E–OBS algorithm interpolates observa-

- 145 tions from weather stations to a regular grid using geostatistical methods (e. g. Journel and Huijbregts, 1978; Goovaerts, 2000) (e.g. Journel and Huijbregts, 1978; Goovaerts, 2000). Note that E–OBS is a land-only data set , and ocean grid points are set to a missing value. Haylock et al. (2008) stated that rainfall totals in E–OBS are reduced by up to almost one third compared to the raw station data at the corresponding grid cells. Regarding extremes, the deviation of E–OBS is even more pronounced (Hofstra et al., 2009). Nevertheless, both studies stated that the spatial mean precipitation in E–OBS is very close to other
- 150 observations.

Additionally Although E–OBS has some limitations, we use it as main reference for this study as there is no other comparable high-resolution daily precipitation data set available that covers entire Europe for a long time period. Other products like satellite data with a very limited time frame are not helpful and also have limitations. There are single ground-based observations with very long time series but as the focus of this study is on intensive areal precipitation this data is of limited usefulness for

155 validation.

Additionally to E–OBS, we compare the RCM simulations with the high-resolved HYRAS data set provided by the German Weather Service (DWD; Rauthe et al., 2013). HYRAS is a gridded precipitation data set with a horizontal resolution of up to

1 km for the time period 1951–2006 and covers Germany and the surrounding river catchments. The HYRAS algorithm also uses ground based measurements and interpolates the point observations to the regular grid. For this study, the HYRAS data was first aggregated to the E–OBS/RCM 25 km grid. HYRAS hereafter means this aggregated 25 km data set.

2.2 Regional climate model simulations

160

185

LAERTES-EU consists of a combined large downscaling ensemble of simulations with one RCM. There are two different types: long-lasting simulations of 45–110 years and simulations over one decade. In the latter, only a period of 10 years (e.g. 1961–1970) was simulated with a specific number of ensemble members. Then, the initialization point was shifted by one

- 165 year (e.g. 1962–1971) and so on until the end of the covered time period. In total, LAERTES-EU consists of 1183 more or less independent simulations (sample size) with approximately 12.500 simulated years. The number of ensemble members at a specific time varies from 6 at the beginning of the century to a maximum of 188 members between 1970 and 2000 (see Fig. S1 in Supplementary).
- LAERTES-EU is divided into four different data blocks (Table 1). All regional simulations used combines a large number of
 regional dynamical downscaling simulations for Europe performed with a single RCM. The used RCM is the non-hydrostatic model of the Consortium for Small-scale Modeling Modelling (COSMO) in climate mode model version 5 (CCLM5; Rockel et al., 2008)and have, which has a spatial resolution of 0.22° (≈ 25 km). The model covers the EURO-CORDEXEURO-CORDEX¹ domain (Jacob et al., 2014). All the simulations were Overall, the simulations use the same domain, model version and set-up, which was adapted from EURO-CORDEX (Kotlarski et al., 2014). According to Feldmann et al. (2008), a dry–day correction
- 175 is important as climate models tend to overestimate the number of wet days with low intensities below 0.1 mm, known as the drizzle effect (Berg et al., 2012). In order to reduce this typical bias, a dry-day adjustment was first applied to LAERTES-EU. The E-OBS data were used for this correction, as they have the same spatial extension and resolution as the CCLM simulations. All simulations are performed within the BMBF (Federal Ministry of Education and Research of Germany) project MiKlip II² (Marotzke et al., 2016) For all simulations the same domain, model version and set-up, adapted from EURO-CORDEX, were
 180 used.

The boundary forcing was to create and test a decadal prediction system including a regional downscaling component for Europe.

For all downscaling simulations the boundary conditions were derived from the Max–Planck Institute of Meteorology coupled Earth System Model (MPI–ESM). This global model consists of the atmospheric component ECHAM6 (Stevens et al., 2013), the ocean component MPI–OM (Jungclaus et al., 2013), and the land-surface model JSBACH (Hagemann et al., 2013).

LAERTES-EU is divided into four different data blocks (Table 1) depending on the setup of the forcing MPI–ESM ensemble simulations. The differences between the four different data blocks stems from the setup, external forcing and initialization of the MPI–ESM simulations. The data blocks 1 and 2 of the RCM ensemble (cf. Table 1) obtained it the boundary values from the MPI–ESM–LR simulations using a T63 resolution and 47 vertical layers. Data block 3 and 4 used the MPI–ESM–HR

¹http://www.euro-cordex.net

²https://www.fona-miklip.de/

Table 1. Overview of the RCM ensemble LAERTES-EU with the name of the simulation within the MiKlip project, the classification into data blocks, the underlaying set-up (experiment), the covered time period, and the number of simulation years. For data blocks 2 and 4, period means the range of the initialization years; XX stand for the ensemble number and YYYY for the initialization year.

name	block	experiment	period	years	comment
as20ncepXX	1	20CR via MPI-ESM-LR	1900–2009	330	3 members of 110 years each
decXXoYYYY	2	MPI-ESM-LR DROUGHTCLIP	1910–2009 <u>1911–2019</u>	3000	3 members with 100 decades each
historical_rXi1p1-HR	3	MPI-ESM-HR HISTORICAL	1900–2005	410	run 1–3 each with 106 years,
					run 4–5 each with 46 years (1960–2005)
preop	4	MPI-ESM-HR CMIP5	1960-2016-<u>1961-2026</u>	2850	5 members with 57 decades each
dcppA-hindcast	4	MPI-ESM-HR CMIP6	1960-2018- 1961-2028	5900	10 members with 59 decades each

- 190 version (Müller et al., 2018) as their driving model. In this version, the horizontal resolution is T127 and 95 vertical layers are applied. Three types of forcing ensembles can be distinguished:
 - (I) MPI–ESM assimilates reanalysis data for long-term simulations (data block 1).
 - (II) Long-term historical-type simulations, according to the CMIP5 specifications (data block 3; Taylor et al., 2012).
 - (III) Initialized decadal (10-year) hind- and forecast simulations (data blocks 2 and 4).
- 195 The MPI-ESM forcing data used for the three long-term simulations in data In data block lassimilated the, the first type (I) is applied. Here the 20th Century Reanalysis (20CR; Compo et al., 2011; Müller et al., 2014) over the period 1900–2009data (20CR; Compo et al., 2011) are assimilated into the MPI-ESM-LR (Müller et al., 2014). 20CR has a spatial resolution of approximately 2° (T62) and was generated using the Global Forecast System (GFS; Kanamitsu et al., 1991; Moorthi et al., 2001) of the National Centers for Environmental Prediction (NCEP)³. It used a 56 member Ensemble Kalman Filter approach
- to assimilate surface pressure, monthly sea surface temperature and sea-ice observations. From these simulations the starting conditions for a decadal hindcast ensemble (data block 2) has been derived (Mieruch et al., 2014; Müller et al., 2014; Reyers et al., 2019; Feldmann et al., 2019). Each year three initialized decadal simulations were started, to study the long-term predictive skill on decadal time scales. Three of the 20CR members are assimilated into MPI-ESM to provide long-term (110 years each) climate reconstruction simulations over the period 1900–2009 (Müller et al., 2014).
 Afterwards, a downscaling with CCLM uses these global simulations as boundary conditions (e.g. Primo et al., 2019).
- Afterwards, a downscaling with CCLM uses these global simulations as boundary conditions (e.g. Primo et al., 2019). Data block 3 contains the downscaling of five un-initialized (historical)consists of the second type (II), were five so called historical simulations of MPI–ESM–HR with CMIP5 observed natural and anthropogenic external forcing (Taylor et al., 2012).

³http://www.ncep.noaa.gov/

Data block climate forcing (Taylor et al., 2012) are used as boundary conditions for CCLM. The ensemble was generated by starting the MPI–ESM from arbitrary dates in a pre-industrial control simulation (Müller et al., 2014). Three of the five CCLM

210 members cover the period 1900–2005 (106 years each). The two additional simulations cover the period 1960–2005 (46 years each).

Data block 2 and 4 encompasses two sets of decadal hindcasts over the period since 1960 (Müller et al., 2012; Marotzke et al., 2016). The preop-ensemble consist of initialized decadal simulations (type III). The starting conditions are derived from an observed state (Müller et al., 2012; Marotzke et al., 2016). For each starting year, an ensemble of decadal simulations is generated and

- 215 then, the initialization point is shifted by one year (e.g. 1961–1970, 1962–1971, and so on). Due to the overlap, a specific calendar year may be covered by several decadal hindcasts with different starting years. These decadal hind- and forecasts thus represent the current state of the major modes of climate variability compared to the so-called un-initialized historical simulations (data block 3). The downscaling procedure, the skill, and the added value are described in Mieruch et al. (2014), Feldmann et al. (2019), and Reyers et al. (2019).
- 220 In data block 2, the starting conditions of the three decadal hindcast members with MPI–ESM–LR are derived from the assimilation experiments in data block 1. The starting years of the CCLM downscaling range from 1910 to 2009. This means the last simulated year is 2019.

Data block 4 consists of two parts. Both of them use the MPI–ESM–HR version. The so-called preop-ensemble has five memberseach year. The climate forcing for these simulations stems also. The external climate forcing is derived from CMIP5;

- 225 whereas for the 10 member per year deppA-ensemble the CMIP6 external forcing was applied (Eyring et al., 2016; Boer et al., 2016). . The starting years range from 1960 to 2016 (last simulated year 2026). The so-called dcppA-hindcast ensemble has ten members and uses the external forcing for CMIP6 (Eyring et al., 2016). The global simulations are a contribution to the Decadal Climate Prediction Project of CMIP6 (DCPP; Boer et al., 2016). The starting years are 1960 to 2018 (last simulated year 2028).
- The In total, LAERTES-EU consists of 1183 simulation runs (sample size) with approximately 12.500 simulated years. The number of ensemble members for a specific year varies from six at the beginning of the century to a maximum of 188 members between 1970 and 2000 (see Fig. S1 in the supplemental material). The simulation in all four data blocks are affected by the observed external climate forcing, but they differ with respect to the representation of the observed climate variability, whereas data block 1 uses assimilated 20CR reanalysis data, data block 2 and 4 contain initialized hindcastsand, which to some
- 235 degree follow the observed low frequency variability, and data block 3 only uses the external forcing information. Nonetheless, the four groups of downscaling simulations can be grouped into a large ensemble, since the regional simulations were all performed with the same setup of the RCM. Despite the same initial conditions and model setup, the temporal evolution of the day-to-day weather is (statistically) independent between the members after a few weeks. This is an advantage, since the data set is homogeneous over time but also covers uncertainties in the observations including unknown and not yet observed events. The validity of this combination approach is tested within Sect. 4.
 - In order to reduce well-known limitations of climate model simulation, the ensemble data first were filtered using a dry-day adjustment. According to Feldmann et al. (2008), a dry-day correction is essential as climate models tend to overestimate the

number of wet days with low intensities below 0.1 mm (Berg et al., 2012), known as the drizzle effect. The dry-day correction was performed using the E-OBS data, as it has the same spatial extension and resolution.

245 3 Methods

The capability of LAERTES-EU to simulate realistic precipitation amounts and distribution is an important requirement. Moreover, temporal variability and possible trends should also be well represented for trustworthy data sets. The methods were applied to different investigation areas and time periods. Equations and additional information can be found in Appendix A–C. As the focus of this study is heavy-intensive areal precipitation, we concentrate on high percentiles of spatially aggregated

250 daily rainfall totals, namely 99 %, and 99.9 %. The percentiles are based on wet days only. First, a spatial aggregation of daily precipitation values was applied. Afterwards, the percentile of these areal precipitation were calculated for each year separately. In all data sets, ocean grid cells were set to a missing value and therefore neglected.

3.1 Validation methods

- LAERTES-EU is analyzed and validated using various methods. The intensity spectrum gives the statistical probability of each precipitation amount by taking into account all grid points and all time steps within the investigation area and without any aggregation. Therefore, the range of occurred values is divided into evenly spaced histogram classes, which then are normalized with the total sample size. The resulting intensity–probability–curve (IPC) is a good indicator ,-if the model is capable to simulate realistic precipitation intensity distributions.
- As an extension to the IPCs, the linear error in probability space L (cf. Eq. A1–A3 in Appendix A) is analyzed (e.g. Ward and Folland, 1991; Potts et al., 1996). Therefore, empirical cumulative density functions (ECDF) are calculated for each simulation run and for the observations. The data basis is the same as for the IPCs. The value ΔC_r (Eq. A1) is defined as the difference between the ECDF of a model run r and that of the observation (difference of probabilities) up to a specific precipitation intensity. It is therefore a measure for the over- or underestimation of the model. Using ΔC_r , the linear error in probability space (L_r ; Eq. A2) is the mean of the absolute values $|\Delta C_r|$ over the entire precipitation range as defined by Déqué (2012) or Wahl et al. (2017). The better both density function coincide, the lower the value of L_r . According Eq. A2.
- 265 Déqué (2012) or Wahl et al. (2017). The better both density function coincide, the lower the value of L_r . According Eq. A2, L_r is always positive. The ensemble mean is given by \overline{L} (Eq. A3).
 - The internal variability of LAERTES-EU on different time intervals is compared to that of the observations. Given that the focus of this study is on intensive widespread precipitation, this analysis is performed using spatial mean precipitation amounts averaged over the investigation areas. First, the time series of daily spatial means are aggregated over different intervals, namely
- 270 monthly, seasonal, and yearly precipitation sums as well as 5, 10, or 30-year running means. In a second step, the standard deviation of a gamma distribution σ_{Γ} is calculated for each of these interval series (see Appendix A; Eq. A4), for every single member of LAERTES-EU, and for the observations. Finally, the ensemble mean of the four data blocks and of the complete ensemble is built. This method enables the analysis of how well the internal variability on different time scales is captured by LAERTES-EU.

275 The quantile–quantile (Q–Q) plot compares the simulated distribution with the observed one using different percentiles of daily spatial mean precipitation. The Q–Q distributions are used to calculate the coefficient of determination R^2 with R being the Pearson correlation coefficient (Eq. A5in Appendix A).

The added value of the ensemble size is analyzed by using the signal-to-noise ratio S2N (Eq. A6). Therefore, we determine a Gumbel distribution (cf. Appendix A) for different sample sizes and the corresponding 90% confidence interval. The S2N , then, is is then the ratio of the return value of the Gumbel distribution divided by the 90% confidence interval (Früh et al., 2010).

3.2 Decadal variability and trend analysis

For the analysis of the temporal evolution of heavy precipitation, we use time series of different percentiles of spatial mean precipitation and quantities introduced and recommended by the Expert Team on Climate Change Detection and Indices (ETCCDI; Karl et al., 1999; Peterson, 2005). Currently, 27 indices for temperature and precipitation are defined by the ETCCDI.

- These indices can be used from local to global scales. Additionally, they combine extremes with a mean climatological state (Zwiers et al., 2013). In this study, we use the two indices R95pTOT and R99pTOT (Eq. B1–B2 in Appendix B), which indicate the amount of precipitation above the 95 % or 99 % percentile, respectively.
- In terms of trend analysis, a Mann–Kendall test (Mann, 1945; Kendall, 1955) is performed with related significance investi-290 gations (Appendix C). Regarding possible oscillations, the complete time series is split into sub-series with a minimum length of 10 years and up to 130 years (trend matrix). The Mann–Kendall test is applied to each of this these sub-series.

3.3 Investigation areas and time periods

The focus of this study is central Europe, implying the countries Germany, Switzerland, the Netherlands, Belgium, Luxembourg, and parts of France, Poland, Austria, the Czech Republic, and Italy. Following Christensen and Christensen (2007),
these countries are mostly coincident with two of the areas defined in the PRUDENCE project (prediction of regional scenarios and uncertainties for defining European climate change risks and effects), namely the PRUDENCE regions (PR) Mid–Europe (ME) and the Alps (AL; Fig. 1). Albeit these boxes contain both land and ocean, the latter was set to a missing value and neglected. During validation, ME and AL were reduced to the HYRAS grid cells lying within the corresponding box, hereafter referred to as ME* and AL*.

- The data sets are investigated on different time periods (TP): TP1 covers the past from 1900 to 2017, which is divided into a sub-period TP1b containing only the period with available observations from 1950 only containing the period 1951 to 2017 2006, with both observations (E–OBS) and HYRAS) being available. The time period TP2 is used for the predictions from 2018 to 2028. Note that the simulations were performed within the MiKlip project back in 2018 (using observations until 2017), which is the reason why the prediction period starts in 2018.
- 305 For climatological aspects, we use the time period 1961–1990, hereafter referred to as climTP. Topographic map of Europe at model resolution 0.22° with the PRUDENCE regions ME and AL (red boxes) and state borders (black contours). A couple of studies (e.g. Cahill et al., 2015; Folland et al., 2018) showed that the climate change signal for global mean temperature



Figure 1. Topographic map of Europe at model resolution 0.22° (in meters above mean sea level; m a.m.s.l.) with the PRUDENCE regions Mid-Europe (ME; dark red box) and the Alps (AL; gray box), state borders (black contours), and the HYRAS area (light red contour). Ocean grid cells are set to a missing value.

significantly increased since the early 1980s. Therefore, using the time period 1981–2010 as reference would possibly include
a strong changing signal to the analysis. Using 1961–1990 reduces the influence of these effects, as this period shows more
stable conditions to a certain degree. This also permits more room for the interpretation of the future predictions.

310

4 Validation of the RCM ensemble

In the following, the above described methods are applied in order to validate LAERTES-EU concerning its representativeness with observations. With this aim, data for the investigation period TP1b is used - and the boxes ME and AL (cf. Fig. 1) are limited to the HYRAS area (ME* and AL*).

315 4.1 StatisticsStatistical distributions and frequencies

The IPCs give the range of simulated (observed) precipitation intensities at any grid point in-within the investigation area and its corresponding probability (Fig. 2). For both investigation areas, the IPCs reveal a distinct added value of the RCM compared to the global model. Due to the coarse resolution, the GCM is incapable of simulating-intensities greater than approximately 60100 mm d^{-1} and underestimates are not found in the GCMs, which underestimate by a large degree the probability of a wide

320 range of intensities. the high intensities. The same applies for the global reanalysis 20CR. On the other hand, the RCM tend to



Figure 2. Intensity–probability–curves (IPCs) of daily rainfall totals of the RCM simulations (dry–day adjusted), observations (E–OBS and HYRAS), GCM simulations (forcing MPI–ESM data at two resolutions LR and HR), and global reanalysis data (20CR) for (a) Mid–Europe (ME*) and (b) the Alps (AL*), both limited to the HYRAS area during the investigation period TP1b (1951–2006). For the IPCs, every grid cell value at every time step was taken into account without any aggregation.

overestimate precipitation intensities and the IPCs lie above those of the observations tends to overestimate the probability for precipitation intensities above a threshold of approximately 50 mm d^{-1} , but cover the entire range of values . For Mid-Europe (Fig. 2a) as the observations. The wider range of intensities at the upper tail of the distribution may include possibly not yet observed events.

- 325 For ME*, the IPCs of the RCM are close to HYRAS, but there is a systematic difference between HYRAS and E–OBS (Fig. 2a). As already mentioned by Haylock et al. (2008), E–OBS has a certain negative bias up to –30 % when using grid point based quantities. The given deviation between of HYRAS and E–OBS is in between this range. For the Alpine region Similar results can be found for AL* (Fig. 2b), the IPCs of E–OBS and HYRAS are almost identical with values up to 200 mm d⁻¹. The difference. The differences between the RCM simulations and the observations at a given probability again is in order of 20%,
- 330 thus within the E-OBS uncertainty are slightly less than for ME*. For both investigation areas the range of simulated values is much higher with up to 470400 mm d^{-1} . Note that only a small part of AL is covered by HYRAS which might Naturally, higher intensities are more likely in the mountainous AL* region.

In contrast to the grid point based IPCs, Fig. 3 shows the mean standard deviation of a gamma distribution (cf. Sect. 3.1 and Appendix A) for the time series of spatial mean precipitation amounts aggregated over different time intervals. For both areas,

335 there is an expectable continuous decrease of internal variability towards longer periods for all data sets/data blocks. For ME*, LAERTES-EU is in good agreement with both observations at least up to a yearly perspective. For longer time periods, data block 1 shows a slightly different behavior compared to the other data blocks and observations. Nevertheless, data blocks 2–4



Figure 3. Intensity–probability–curve–Mean standard deviation σ_{Γ} in mm (IPCmean over data blocks) of spatially averaged precipitation aggregated over different time intervals: daily rainfall totals of the RCM simulations (dry–day adjustedday), observations-monthly (E–OBS and HYRASmth), seasonal (seas), yearly (yrs), and global reanalysis 5/10/30-year running mean (20CR5yr/10yr/30yr) for (a) Mid–Europe (ME)* and (b) the Alps (AL* (TP1b; 1951–2006)during–. The four data blocks of LAERTES-EU are considered separately; RCM mean stands for the investigation period TP1b complete ensemble mean (gray). The results for E–OBS and HYRAS are given in black and magenta. Note that it is not possible to estimate the 30yr values for the decadals of data blocks 2 and 4.

and the ensemble mean continue to match with the observations up to the 10-year running mean. Note that it is not possible to estimate the 30-year running mean for the decadal simulations of data block 2 and 4 given the data availability. For data block3,
only an external climate forcing was used meaning these so-called historicals are free runs in terms of daily weather evolution. Therefore, it is not expected that the multi-decadal variability is in phase to the observed circulation after a certain time, which can be a reason for the vanished differences between slightly higher differences of data block 3 compared to the observations at the longest time scale. Furthermore, note that the results of Fig. 3 do not indicate a perfect match of LAERTES-EU in terms of absolute values, but rather that the internal variability (spread) of spatial mean precipitation totals is well captured.
For the mountainous AL* region, the internal variability is higher and all data blocks have a higher standard deviation at all time intervals. This means that the spread of simulated precipitation amounts is increased compared to that of the observation. A necesible reason for this difference and means from sparse measurements in the region aparidared for both E ORS and

A possible reason for this difference can emerge from sparse measurements in that region considered for both E–OBS and HYRASand the resulting specious deviations to the RCM, especially for long-term observations. The more or less constant difference between LAERTES-EU and the observations can be an indicator of a possibly systematic bias in this region.

- A direct linkage between observed and simulated precipitation is given by a The Q-Q plot (Fig. ??). Therefore, plots of daily spatial mean precipitation fields for both investigation areas are used. Then, the distributions for these values are calculated shown in Fig. S2 in the supplemental material. Generally speaking, the distribution of the RCM is in better agreement with similar to those of the observations, at least with to E-OBS, with little deviations from the optimum (diagonal line) for most of the spectrum and differences at around 10% for the upper part of the distribution. In comparison to HYRAS, the
- 355 maximum deviation is higher with around 20%. For AL(Fig. S2), the RCM data differs more and over a wider range of the spectrum compared to HYRAS *, the differences between the RCM and HYRAS are larger than for ME. * (Fig. S2). Even though HYRAS was aggregated to the E–OBS/RCM grid, the more pronounced differences especially for the extremes might be a result of the higher resolution of the HYRAS data, which, in particular, is of greater relevance in the mountainous region of AL*.
- The findings of Fig. ?? S2 are confirmed by the determination coefficients R^2 (Table 2). For both E–OBS and HYRAS, the coefficient is very high with $R^2 > 0.98$. There is a slightly higher R^2 for E–OBS than for HYRAS, which is an artificial effect of the data resolution. The region AL* shows a minimal higher skill compared to ME* in E–OBS and slightly lower values in HYRAS. Table 2 also reveals higher correlations of the CCLM simulations driven by the high resolution high-resolution MPI–ESM–HR data compared to those driven by the lower resolved MPI–ESM–LR data. Quantile–quantile plot of spatial
- 365 mean daily precipitation for investigation period TP1b comparing the RCM simulations (data block 1–4) with E–OBS (solid lines) and HYRAS (dashed lines) for Mid–Europe (ME).

Even though this seems to be systematic, the differences are marginal.

Table 2 also contains the mean linear error in probability space \overline{L} for the different data blocks. Again, the differences between the data blocks are marginal with all cases being close to $\overline{L} = 0$ which stands for a good agreement of LAERTES-EU

370 with observations. In contrast to R^2 , \overline{L} has lower values for the simulations driven by MPI–ESM–LR. For all data blocks, \overline{L} is considerable higher for the mountainous AL* region. Note that both quantities being close to its optimum value does not indicate a perfect model. It rather means that the overall statistics regarding the entire range of intensities to a high degree coincide with the observations.

4.2 Time series

- 375 Beside overall statistics, other properties of LAERTES-EU like the temporal variability should cover the range of observations as well. Therefore, we analyze the time series of yearly values of different percentiles of the spatial mean precipitation for the investigation areas. In Fig. 4, the time series of the 99% percentile for ME^{*} is shown. Both observational data sets have a high year-to-year variability with similar shapebut the mean over TP1b is about 10% higher in HYRAS than in E-OBS. The ensemble mean value is very close to the E-OBS mean with relative deviations between -5of LAERTES-EU is higher.
- 380 with a relative deviation of 1–10 % and 4 %, and 0.6 % on average during (TP1b . Compared to HYRAS the differences are -14 to -5average is 7 % with -8 % on average). The spread of both observational data sets is covered by the ensemble spread (minimum to maximum values) of LAERTES-EU except for few extreme peaks (e.g. 1985in E-OBS or 1998 in HYRAS). In AL, the HYRAS *, the E-OBS mean is about 155 % higher than E-OBS HYRAS but both time series have again a similar

Table 2. Coefficients of determination R^2 between the RCM and observations (top number) for the quantile-quantile contemplation of Fig. ?? S2 and linear error in probability space \overline{L} (bottom number) between the RCM and both observations (E–OBS and HYRAS) for Mid–Europe (ME^{*}) and the Alps (AL^{*}), always using HYRAS grid cells only. Both quantities are based on daily spatial mean precipitation amounts.

RCM	E-OBS	HYRAS			
	${ m ME}^*_{\sim}$	AL^*_{\sim}	ME^*_\sim	AL*	
Data block 1	0.9914	0.9924	0.9876	0.9835	
	0.0016	0.0058	0.0027	0.0080	
Data block 2	0.9914	0.9925	0.9878	0.9848	
	0.0009	0.0037	0.0021	0.0058	
Data block 3	0.9963	0.9976	0.9936	0.9930	
	0.0017	0.0062	0.0029	0.0083	
Data block 4	0.9966	0.9981	0.9943	0.9938	
	0.0011	0.0038	0.0023	0.0059	

shape (Fig. S3). The ensemble mean in this area lies within both observation means and a little closer to HYRAS. The relative deviation is 6–15 again is higher with relative deviations of 12–23 % (1016 % on average) to E–OBS and -8 to 0.218-29 % (-521 % on average) to HYRAS. The ensemble spread also covers the observed variability.

Regarding more extreme values, namely the 99.9 % percentile, similar results can be found . E-OBS shows a certain bias to HYRAS of approx. 10% for ME and 25% for AL (Fig. S4 and S5). Again, E-OBS and HYRAS show a similar behavior for both areas with mean value differences of less than 1%. The ensemble mean is close to E-OBS with a deviations of -10 to
1shows a mostly positive bias with deviations of less than 10% (-36% on average) for ME, and 6 to 16during TP1b) compared to E-OBS for ME* and 6-18% (average of 10% on average) for AL. Compared to HYRAS, LAERTES-EU differs between -19 and -7% (-11% on average) in ME, and between -18 and -10% (-15% on average) in AL.*. Furthermore, there is are a distinctly higher spread and variability of the 99.9% for both, the observations and LAERTES-EU. Again, the minimum values of the ensemble spread seem to be constant over time, while there is an increase in the maximum values for ME but no clear signal for AL. Except for a few peaks, LAERTES-EU covers the spread of the observations.

4.3 Added value of the sample size

In order to demonstrate the added value of the presented LAERTES-EU, we use the signal-to-noise ratio (S2N, Eq. A6) for different sample sizes and return periods (cf. Appendix A). Sample size, in this case, means the number of data which is equivalent with the number simulation runs. Note that the simulations vary in length (number of years) with a minimum



Figure 4. Time series of the yearly 99% percentile (wet days and HYRAS area only) of daily spatial mean precipitation values for Mid-Europe (ME^{*}) during TP1b (1951–2006) of the LAERTES-EU ensemble mean (black), the ensemble spread (minimum to maximum; gray), E–OBS (red), and HYRAS (blue). The dotted lines symbolize the mean values of the observations throughout TP1b.

- 400 length of 10 years and a maximum of 110 years. In order to reduce the influence of the sample length on the results, the single simulation runs of simulation runsLAERTES-EU where randomly concatenated using a hundredfold permutation. Observations have a sample size of 1. Again, S2N is calculated for daily spatial mean precipitation amounts during TP1b only using the HYRAS area.
- For both MEand AL, the * and AL*, S2N increases with the sample size steadily increases with sample size for all calculated
 return values meaning a more statistically robust estimate of the return values (Fig. 5). At the beginning there is a strong increase of S2N until a sample size of approximately 10. Between a sample size of 10 to 100 the increase of S2N is weak. This range is typically used as ensemble size. For a sample size of 100 and more S2N increases rapidly. Signal-to-noise ratio for different return periods T (colored lines) dependent on the sample size for (a) ME and (b) AL.
- Furthermore, the S2N is lower for higher return periods which is a result of the increasing uncertainty of the best estimate 410 due to less or even no data points for very high return periods. However, S2N also increases with sample size for the very high return periods. The robustness of a 2-year return value estimate of a sample of size 20-1 is about the same as the 1000-year estimate for a sample of size 1000.-20. This means that even for extremes, which have not been observed yet, some robust statistical analysis can be carried out.



Figure 5. Signal-to-noise ratio S2N for different return periods T (colored lines) of daily spatial mean precipitation dependent on the sample size for (a) ME^{*} and (b) AL^{*}. The LAERTES-EU members were randomly stringed together permuting the order a hundred times. The shown S2N is the mean of this permutation.

5 Long-term variability and trends

415 The temporal evolution and variability of extreme precipitation throughout the <u>entire past</u> time period TP1 (<u>1900–2017</u>) and also for <u>predictions of the upcoming decade the predictions (TP2) is; <u>2018–2028</u>) are evaluated in this section. Beside time</u>



Figure 6. Boxplot of the distribution of daily spatial mean precipitation values (including dry days) for ME. Each decade during TP1 (blue) was considered separately. The centerline of a box marks the median; the lower and upper end of the box mark the 25th 25% and 75th 75% percentile (interquartile range); the whiskers represent approximately the 99.9% percentile; TP2 the prediction part is marked in green.

series of percentiles, we use climate change indices and statistical distributions. In this section, all land grid cells within the investigation areas ME and AL are used for calculating the daily areal mean precipitation amounts.

5.1 Precipitation distributions

- 420 Figure 6 shows the evolution of the distribution of spatial areal mean precipitation throughout TP1 and TP2 by treating each decade independently. For the core of the distributions, namely medians, interquartile ranges, and upper whiskers, only small variances can be found between the different decades which means that there is almost no change for the majority of the precipitation amounts. Nevertheless, a marked positive trend for the uppermost extremes of the distributions appears with maximum values around 18 mm d⁻¹ at the beginning of the 20th century and about 24 mm d⁻¹ in the 21st century. The distribution
- 425 for the upcoming decade 2020–2028 (Figure 6, green boxplot) shows only minimum shows only small differences to those of the present decade since 2010 with an almost equal median and interquartile range, but slightly higher maximum values (Figure 6, green boxplot). Note that the decade 2010–2019 contains the years 2018 and 2019 from the predictions, and that the last "decade" 2020–2028 is shorter with 9 years.

The boxplot for AL is shown in Fig. S6 and illustrates that not only the high percentiles reveal a decrease in the middle of the 430 century, but the entire distribution is shifted towards lower values. Nevertheless, there is no clear tendency for the maximum values. For TP2 (Fig. S6, green boxplot) the the upcoming decade the distribution is similar to that of the present decade in

Table 3. Overall trend of daily spatial mean precipitation during TP1 and TP2 (1900–2028) using a linear regression of the yearly series of the 99% and 99.9% percentile (pct; wet days only) for ME and AL; Given are absolute values and the relative changes (RC) compared to the climatological mean (climTP; 1961–1990) for the ensemble minimum (min), the ensemble mean, and the ensemble maximum (max) percentile values within LAERTES-EU, and the related significance (p–value; $\alpha = 0.05$).

area	pct	variable	trend	RC	climTP	p_{lpha}
			(mm)	(%)	(mm)	
		min	-0.4	-5.3 - <u>4.6</u>	7.5 -7 <u>.8</u>	0.9966- 0.9387
ME 99	99	mean	0.7- 0.8	7.0- 7.8	10.0 - <u>10.3</u>	1.0 - <u>1.0000</u>
		max	2.6	19.0	13.7 - <u>13.9</u>	1.0 - <u>1.0000</u>
		min	-0.71.0	-7.8 - <u>10.9</u>	9.0	0.9974
ME	99.9	mean	1.1	8.2 -8.4	13.4 - <u>13.5</u>	1.0 -1.0000
		max	7.2. 6.7	33.2 - <u>31.0</u>	21.7- 21.6	1.0 -1.0000
AL	99	min	-2.6	-17 <mark>.8</mark> .1	14.6 - <u>15.4</u>	1.0 -1.0000
		mean	0.3-0.1	-1.5 -0.4	20.2-21.0	0.9381- 0.7208
		max	4.4-5.4	15.9 - <u>18.9</u>	27.7- 28.4	1.0 -1.0000
		min	-3.8 -21 4 .3	-23.9	17.8	1.0 - <u>1.0000</u>
AL	99.9	mean	-0.0	-0.0	27.3	0.0- 0.0000
		max	8.3 9.0	18.9- 20.0	44.0 44.7	1.0 - <u>1.0000</u>

case of median and the upper part of the distribution - (Fig. S6, green boxplot). The interquartile range is reduced due to a increased lower boundary of the boxplot.

5.2 Temporal evolution of yearly percentiles

435 5.2.1 Overview

The overall trend during TP1 and TP2 using a linear regression for both areas and percentiles is given in Table 3. While the ensemble mean shows a significant positive trend for ME for both percentiles, a small but significant negative trend can be found for the 99 % of AL, while there is almost no change in the 99.9 % of AL. In all cases, the ensemble spread increases due to both a decrease of the minimum values and an increase of the maximum values both being highly significant. The change of

440 the maximums is stronger than the reduction of the minimums and more pronounced in AL than in ME.



Figure 7. Time series of the yearly 99% percentile of daily spatial mean precipitation (wet days only) for Mid–Europe (ME; land only) of the LAERTES-EU ensemble mean (solid line), and the ensemble spread (minimum to maximum; dots and shaded area) during TP1 (1900–2017; black/gray) and TP2 (2018–2028; reddish).

Analogous to Table 3 we analyze the trend for TP1b only (Table S1 in Supplementarythe supplemental material). The tendencies are the same for all cases but less pronounced except for the mean 99.9% of AL where the negative trend during TP1b is slightly stronger than for the whole time series.

445

Figure 7 shows the temporal evolution of the 99% percentile during the 20th and the beginning of the 21st century for the whole LAERTES-EU. As given in Table 3, the lower boundary changes are small, while there is a visible positive trend of the ensemble mean and the upper boundary of the ensemble spread. Note that the larger spread from the 1960s onwards might be artificial due to the decisively larger number of members of data block 4. Nevertheless, there is a clear consistency in the time series for ME.

Some differences emerge for the Alpine region AL (Fig. S7). At first, there is a distinct decrease of the ensemble mean between 1960 and 1970 which might reveal from the rising number of members. As the ensemble matches well with the observations, we presume an overestimation of precipitation in the first half of the 20th century in that region, which could be a result of missing data for the applied dry–day correction. Due to the more complex terrain, the structure of the precipitation fields is more complex, and therefore more sensitive for different types of effects such as the dry–day correction. Time series of the yearly 99% percentile of spatial mean precipitation for Mid–Europe (ME) of the LAERTES-EU ensemble mean (solid

455 line), and the ensemble spread (dots and shaded area) during TP1 (black/gray) and TP2 (reddish).

Table 4. Climatological mean (climTP; 1961–1990) of days per year exceeding the 99% and 99.9% percentile (pct; wet days only) for ME and AL, linear regression (LR) and relative change (RC) compared to climTP for different investigation periods (TP), and related significance (p-value; $\alpha = 0.05$).

area	<u>pct</u>	<u>climTP</u>	<u>TP</u>	LR_	RC	₽œ
ME	<u>99</u>	3.20	<u>1+2</u>	1.25	<u>39 %</u>	1.0000
			$\underbrace{1b}$	0.76	<u>24 %</u>	1.0000
	<u>99.9</u>	0.60	<u>1+2</u>	0.36	<u>60 %</u>	1.0000
			$\underset{\sim}{\overset{1b}{\sim}}$	0.19	<u>32 %</u>	1.0000
AL	<u>99</u>	3.11	<u>1+2</u>	-0.17	<u>-6 %</u>	0.8262
			$\underbrace{1b}$	-0.37	-12%	0.9251
	<u>99.9</u>	0.62	<u>1+2</u>	-0.02	-2%	0.7084
			<u>1b</u>	-0.04	<u>-6 %</u>	0.2973

The results for the 99.9% percentile are similar for both areas (Fig. S8 and S9). The positive trend for ME is even more pronounced, while the drop in the 1960s for AL is less visible and therefore, the time series is more constant.

Climatological mean 1961–1990 (climTP) of days per year exceeding the 99% and 99.9% percentile (pet) for ME and AL, linear regression (LR) and relative change (RC) compared to climTP for different investigation periods (TP), and related 460 significance (p-value), area pet climTP TP LR RC p_αME99 3.55 1+2 1.39 39 % 1.0 1b 1.11 31 % 1.0 99.9 0.57 1+2 0.36 63 % 1.0 1b 0.27 47 % 1.0 AL 99 3.78 1+2 -0.26 -7 % 0.9618 1b -0.58 -15 % 0.9964 99.9 0.61 1+2 -0.01 -2 % 0.6825 1b -0.01 -1%0.3775

Taking a look into For ME, the evolution of the number of days exceeding the climatological mean percentile, reveals a strong positive and significant trend appears for ME for for both the 99 % (Fig. 8, top) and 99.9 % percentile (Fig. S10). The exact values of the climTP mean, the linear regression, the relative change, and the significance can be found in Table 4 (top 465 numbers). For the Alpine regionAL, the year-to-year variability is higher and the overall trend is slightly negative (Fig. 8, bottom, and S11) and at least significant for the 99% percentile. Again, we analyze the trend for TP1b separately (Table 4, bottom numbers). the The tendencies for TP1b are the same but less pronounced except for the days exceeding the 99% percentile in AL, where there is a stronger trend signal in TP1b compared to the whole time series, which is also significant to a high degree.

470

5.2.2 Past trends and periodic oscillations

Trend analysis of the 99% percentile for ME with (a) the relative amount of members of LAERTES-EU with a positive (blue) or negative (red) trend; (b) the trend in mm per year averaged over the members from (a), and (c) relative amount of members



Figure 8. Deviation of the <u>LAERTES-EU</u> ensemble mean of the yearly number of days above the 99% percentile (wet days only) of daily spatial mean precipitation compared to the climatology (1961–1990; climTP; <u>1961–1990</u>) for <u>ME (topa) Mid-Europe (ME)</u>, and <u>AL</u> (bottomb) the Alps (AL). Red bars indicate negative anomalies (less days), blue bars positive anomalies (more days). The predictions (TP2; 2018–2028) are given in green. The black line indicates a linear regression.

from (a) that have a significant trend; cases with no distinct number (less than 60%) of members with same trend sign are marked in gray in (a)–(c). For a more detailed analysis of trends, the method Mann-Kendall test described in Sect. 3.2 is applied to the time series of daily spatial mean precipitation percentiles. Figure 9a shows the relative number of LAERTES-EU members (relative) with that show a positive or negative trend of the 99% percentile for ME. Only cases in which more than 60% of the complete ensemble members reveal the same tendency are then considered for further investigations. For these



Figure 9. Trend analysis of the 99 % percentile (wet days only) of daily spatial mean precipitation for ME with (a) the relative amount of members of LAERTES-EU with a positive (blue) or negative (red) trend; (b) the trend in millimeter per year averaged over the members from (a), and (c) relative amount of members from (a) with a significant trend; cases with no distinct number (less than 60 %) of members with same trend sign are marked in gray in (a)–(c).

casesthe, the ensemble mean trend is calculated (Fig. 9b) and the relative amount of significant members is displayed (Fig. 9c).
All cases in which the ensemble reveals ambiguous tendencies are neglected (gray areas).

To a high degree the single members show the same behavior, especially for the longer time series where positive trends are dominant. On a decadal time scale (diagonal line in Fig. 9), some oscillations appear with phases of increasing and decreasing precipitation. This signal might be smoothed as it is not expected that the decadal simulations of data blocks 2 and 4 cover the natural variability at this time scale in detail. Furthermore, these simulations are not expected to be in phase with the long lasting simulations of data blocks 1 and 3. The trends on this time scale reach rates of up to 0.1 mm a⁻¹ or 1 mm per decade,

lasting simulations of data blocks 1 and 3. The trends on this time scale reach rates of up to 0.1 mm a⁻¹ or 1 mm per decade, respectively. The overall trend is weaker with rate of 0–0.02 mm a⁻¹ or 0–2 mm per century, respectively. Positive trends are more often significant than the negative, while only a small part of the ensemble shows significant trends. Similar results can be found for the Alpine region AL (Fig. S12). The trends on the decadal time scale reach higher rates but the oscillation is less pronounced than in ME. Again, most of the positive trends are significant, while just a few members with negative trends are significant.

For the 99.9% percentile of ME(Fig. S13), large parts of LAERTES-EU show positive trends -(Fig. S13). On the decadal time scale a clear sequence of positive and negative trends is visible. Both the increases and decreases are more pronounced than for the 99% percentile but only a few members are significant. In the Alpine region (Fig. S14) For AL, even more parts of the ensemble have the same tendency of heavy precipitation and a higher number of members have a significant trend -(Fig. S14).

These trends exceed rates of decisively more than ± 0.1 mm a⁻¹. In contrast to the results above, the 99.9 % percentile for AL 495 seems to have a multidecadal oscillation, while the overall trend of the complete time series is negative.

5.2.3 Future projections predictions

With respect for the upcoming decade (TP2; 2018-2028), LAERTES-EU predicts an continuation of the current trend with an increase especially for the 99.9 % percentile (Fig. 7, and S6-S8; reddish area). In comparison to the last decade (2007-2017), the RCM mean of the 99 % percentile increases of about 0.6 % for ME and about 2.1 % for AL. The 99.9 % percentile increases

500

about 2.0% for ME and 3.0% for AL.

Further to this absolute change, the number of days exceeding the climatological 99% percentile shows an increase of 4.9% for ME and 8.4 % for AL, and 6.7 % (ME) and 22.4 % (AL) in case of the 99.9 % compared to the mean of 2007–2017. This also manifests in the relative anomaly (Fig. 8, and S10-S11; green bars).

505

Relative deviation of (a) the R95pTOT index and (b) the R99pTOT index of the LAERTES-EU mean compared to the elimatology (climTP) for ME. Red bars indicate negative (dry) anomalies, blue bars positive (wet) anomalies. The predictions (TP2) are given in green. The black line indicates a linear regression.

Nevertheless, a more detailed trend analysis illustrated in Fig. 9 and also Fig. S12-14 reveals that LAERTES-EU shows no clear tendency for the 99% for during TP2. Just in a few cases, more than 60% of the members have a similar mainly 510 positive trend signal, which however, however, is not significant. In case of the 99.9% percentile, 60–70% of the members show a strong positive trend of more than 0.1 mm a^{-1} with 20–40 % of them being significant. Although the tendency for TP2 is ambiguous and less significant, it shows continuity to the present decadeand so we conclude that a positive trend is likely.

5.3 **Climate change indices**

- The results described in the previous section sections also manifest in the considered ETCCDI climate change indices (Table 5). R95pTOT shows a positive trend for ME (Fig. 10a) with a relative change of about 18% and a strong negative trend of 515 approximately -15% for AL (Fig. S15). Remarkably, there is a high positive deviation in the early first half of the 20th century compared to the climTP amount for AL which might be artificial due to the mentioned problems of the dry-day correction. R99pTOT shows a positive change for ME (Fig. 10b) and a slightly negative trend for AL (Fig. S16). The overemphasis overestimation for AL in the early century is less pronounced for this index. Considering only the TP1b, the tendencies are the same in all cases. The positive trends for ME are less pronounced, while the negative trends for AL are stronger. The estimated 520
- trends are highly significant except for the R99pTOT of AL for the whole time series.

Compared to the present decade, the projections show a continuation of the positive trend for ME with an increase of 2 % for R95pTOT and 5% for R99pTOT. In contrast, both indices show a positive trend for AL with an increase of 7% for R95pTOT and 8% for R99pTOT, which is a complete reversion of the overall trend.

525 Compared to the present decade, the predictions show a continuation of the positive trend for ME with an increase of 2% for R95pTOT and 5% for R99pTOT. In contrast, both indices show a positive trend for AL with an increase of 7% for R95pTOT and 8% for R99pTOT, which is a complete reversion of the overall trend.



Figure 10. Relative deviation of (a) the R95pTOT index and (b) the R99pTOT index of the LAERTES-EU ensemble mean of daily spatial mean precipitation (wet days and land only) compared to the climatology (climTP; 1961–1990; Table 5) for Mid-Europe (ME). Red bars indicate negative (dry) anomalies, blue bars positive (wet) anomalies. The predictions (TP2; 2018–2028) are given in green. The black line indicates a linear regression.

Table 5. Climatological mean 1961–1990 (climTP; 1961–1990) of ETCCDI quantities for Mid-Europe (ME) and the Alps (AL), linear regression (LR) and relative change (RC) compared to climTP for different investigation periods (TP), and related significance (p–value; $\alpha = 0.05$). Both indices are based on wet days only of daily spatial mean precipitation (land only).

area	ETCCDI	climTP	TP	LR	RC	\mathbf{p}_{lpha}
		(mm)		(mm)	(%)	
ME	R95pTOT	157.5	1+2	28.4	18	1.0
			1b	20.1	13	1.0
	R99pTOT	43.8	1+2	15.6	36	1.0
			1b	12.2	28	1.0
AL	R95pTOT	306.7	1+2	-46.3	-15	1.0
			1b	-54.3	-18	1.0
	R99pTOT	88.5	1+2	-4.5	-5	0.8953
			1b	-10.8	-12	0.9891

6 Summary and Conclusions

We have presented a novel combined the novel ensemble LAERTES-EU of combining various regional climate model simulations to better estimated heavy done with COSMO-CLM to analyze long-term variability and trends of flood related intensive areal precipitation across central Europe. The whole RCM ensemble was divided into four data blocks depending on forcing data, assimilation schemes, or the initialization of the driving global model MPI-ESM. The setup of the COSMO model remained the same for all simulations. In total, the presented LAERTES-EU consists of over 1100 simulation runs with approximately 12.500 simulated years on a 25 km horizontal resolution.

- The focus of investigation was laid on the PRUDENCE regions Mid–Europe (ME) and the Alps (AL). Regarding heavy precipitationintensive areal precipitation, we concentrated on high percentiles, namely 99% and 99.9%. It was not expectable that the simulations are of spatially averaged daily precipitation amounts. Note that it was not expected that LAERTES-EU was able to reproduce historical precipitation events on a daily base in detail, but have a better more accurate performance regarding long-term variations, and statistical distributions on a larger scale perspective. Furthermore, the given resolution restricts the
- 540 consideration of convective processes, so we analyzed time series of spatial mean precipitationconcentrated on larger scale phenomena.

With respect to our <u>initial</u> research questions, the following main conclusions can be drawn <u>and summed up</u> out of the presented results, which will be discussed more detailed afterwards:

- (1) Extreme precipitation is well represented in LAERTES-EU is capable of representing the range of extreme areal precipitation
- 545

550

similar to the used observational data sets and also fits into the range of previous studies (e.g. Früh et al., 2010). The four data blocks are consistent and have similar precipitation distributions(IPCs), which are within the uncertainty of the observations. The ensemble range also covers the observed temporal evolution.

- (2) The added-value benefits of the large ensemble size manifests in a strong increase of the signal-to-noise ratio beyond the typically used ensemble sizes and in high statistical significances of estimated trends for the ensemble mean. Furthermore, the distribution of precipitation totals is represented in a more concise way taking the limitations of the considered observations into account.
- (3) Long-term trends reveal spatial differences in sign and strengthand between the members. These tendencies are partly significant. Despite a quite large ensemble spread, the ensemble mean shows more explicit results. Distinct oscillations can also be found on shorter time scales (e.-g. decades).
- (4) The projections predictions for the upcoming decade show a continuation of past tendencies with increasing heavy precipitation in terms of both intensity and occurrence frequency for ME without any discontinuity - Howeverto the previous time period. On the other hand, LAERTES-EU shows no clear signal and less significance for the projections for AL:

Regarding the validation (1), grid point based intensity-probability-curves (IPCs)and, areal mean precipitation distributions (internal variability σ_{Γ} and linear error in probability space \overline{L}), and Q-Q distributions have been analyzed. In all cases, the IPCs of the simulations show an overestimation of precipitation in order of 10–20% compared to E–OBSof about one third. Haylock et al. (2008) found out that E–OBS has can have a certain negative bias of up to 30% compared to raw single ground based punctual observations. Taking this into account, the IPCs are almost coincident. NeverthelessFurthermore, the IPCs of LAERTES-EU show only small deviation compared to the high resolution HYRAS data set - Distinct differences

- 565 mainly appear in the Alpine Mountains, which can be explained by less spatial coverage of observations. Furthermore, the IPCs and (aggregated to the model grid). The IPCs and also the Q-Q distributions of all four data blocks are coincident which was a prerequisite for the combination to one large ensemble. The Q-Q distributions of spatially aggregated mean precipitation reveal less differences between modeled and observed precipitation compared to LAERTES-EU and E-OBSand, but an underestimation of simulated rainfall compared to HYRAS by about 10%. The linear error in probability space \overline{L} shows
- 570 a good agreement of LAERTES-EU with observations in terms of the distribution of daily areal mean precipitation totals. For different aggregation intervals from daily values up to 10-year running means, the internal variability (standard deviation σ_{Γ}) of LAERTES-EU matches to a high degree with that of both observations. Note that both quantities \overline{L} and σ_{Γ} do not indicate whether the simulated absolute precipitation values coincide with the observations, but rather show the agreement of statistical properties.
- 575 Regarding (2), LAERTES-EU reveals a clear added value due to the large sample size. Estimates of long return periods are more robust compared to smaller ensembles which is of importance, for instance, for risk and insurance applications. Furthermore, trends at least in the ensemble mean are highly significant. The IPCs also show a clear added value benefit of RCM data compared to coarser global models the coarser global model (MPI–ESM) or the 20CR global reanalysis. Regarding extremes, LAERTES-EU includes a broader range of precipitation totals with even higher values, which are not covered by
- 580 observations due to their limited temporal availability. Although the presented results reveal a broad range of realizations within LAERTES-EU, the statistics of the ensemble mean clearly benefit from the large ensemble size with a better signal-to-noise ratio.

Besides a proper representation of precipitation, long-term trends and temporal variations were of special interest. Regarding (3), the presented results show a good-reasonable agreement of LAERTES-EU concerning the temporal evolution of the con-

- 585 sidered percentiles of spatially aggregated daily precipitation totals for the different investigation areas. The ensemble mean is within the range of the observations and the spread (min-to-maxspread (minimum to maximum)) covers the observed variability except a few peaks. The ensemble mean shows a small positive bias compared to both observational data sets. Throughout the complete time period TP1 (1900–2017), positive and significant trends can be found for ME in both percentiles and (99 % and 99.9 %) and also in the number of days exceeding the climatological mean (1961–1990). For AL, there is no clear trend
- 590 signal in the ensemble mean but an increase in the maximum values. In contrast, the number of days exceeding the elimatology is decreasing. The positive trends for ME with relative changes about 7–8% are coincident with the theoretical 6–7% per Kelvin temperature change (CC rate) as Moberg et al. (2006) found an increase of approximately 1 K during the 20th century for Europe. The negative trends for AL, however, do not fit in this theoretical estimate. The maximum simulated percentile values increase with a super–CC rate up to a factor 4.

595 climatological mean percentiles is decreasing in this area. Comparing the trends of TP1 to the shorter TP1b (1951–2006), the tendencies are the same but less pronounced in TP1b. On a decadal time scale, some oscillations can be found with periods of increasing precipitation and such with decreasing values. Similar results as for time series of percentiles can be found using climate change indices (ETCCDI).

Regarding (4), the projections for the upcoming decade until 2028 predictions for the next decade 2018–2028 (TP2) reveal ongoing tendencies of heavy precipitation indices. A special case is the Alpine region AL where the slightly negative trends in the past (TP1) turn to positive onceones. Both the continuity for ME and the reversion for AL appear in all time series, namely the number of days or ETCCDI variablesand all of threshold exceedance, ETCCDI variables, and investigated percentiles. While there is are a clear signal and high significance building for the ensemble meanfirst, the trends are ambiguous and less significant when the ensemble members were considered separately. However, we conclude that this tendencies are likely as

605 it is a continuation of the results of the present decade. Similar results for parts of LAERTES-EU were found by Reyers et al. (2019).

Precipitation remains a challenging task for both reanalyses and climate model simulations of the past and the future with partly contrasting results shown by several previous studies. Furthermore, long-term comprehensive observations are not available which makes a validation difficult due to the high spatial variability of precipitation. This also affects analyses of

- 610 trends or climate variability. What is known is a theoretical increase of the water vapor capacity according to the Clausius–Clapeyron (CC) equation of about 6–7% per degree of temperature increase (e.g. Trenberth et al., 2003; Berg et al., 2009), which assumes a near constant relative humidity. The CC rate is generally thought to be a proxy for future precipitation projections (Westra et al., 2013). A recent discussion about the validity of the CC rate as an estimate for future projections of heavy precipitation can be found in Zhang et al. (2017). They pointed out that beside the thermodynamic responses, changes in heavy precipitation may be also
- 615 influenced by dynamical effects. Furthermore, Pfahl et al. (2017) and Kröner et al. (2017) showed that precipitation trends can be regionally influenced by contributions from both lapse-rate and circulation effects.

The ensemble mean of LAERTES-EU shows an increase of about 1.9 °C for ME and 2.3 °C for AL for the yearly mean 2 m-temperature of spatial means during the 20th century (TP1; 1900–2017). Including the predictions (TP2), the increase is about 2.4 °C for ME and 2.8 °C for AL. For instance, Simmons et al. (2017) found an increase over European land masses of

- 620 approximately 2 °C in the mean compared to pre-industrial conditions. Moberg et al. (2006) found an increase of about 1 °C for temperature extremes. Thus, LAERTES-EU is within the range of observed changes. The increase in temperature over the entire time period is equivalent to a CC scaling of about 15–20 %. The extracted changes of the high precipitation percentiles for ME make up to 50 % compared to the theoretical CC value. However, the negative tendencies for AL do not fit into this theoretical estimate.
- The presented LAERTES-EU data set can be used for various applications -fields. In particular, the simulations are used as input for hydrological modeling and further applications such as flood risk assessments. The presented ensemble in this case can be used as a stochastic weather generator treating the single simulations independently. This leads to the production of a quasi-stochastic hydrological discharge data set. Due to the large ensemble size, estimates of high return periods become more robust. However, it has to be mentioned that the composition of the four data blocks to one ensemble restricts the temporal

- 630 homogeneity. Nevertheless, the agreement with intensity distributions, observations, and statistics is very high. Moreover, the validation showed a positive bias of the ensemble mean which, together with the overestimation of low intensities, requires a bias correction to avoid unrealistic discharges. This application as well as the bias correction of LAERTES-EU will be addressed in a consecutive study.
- In this studywe have been, we have focused on all-year variances, oscillations, or trends. Future investigations will can address a seasonal differentiated analysis of trends and oscillations as well as a more detailed investigation of the spatial distribution of these findings - In particular, the simulations can be used as input for hydrological modeling and further applications such as flood risk assessments. The presented ensemble in this case acts as a stochastic weather generator treating the single simulations independently. Estimates of high return periods become more robust.

Furthermore, analyses of possible mechanisms behind observed oscillations are in preparationand potential mechanisms
640 behind the observed variability. Previous studies indicated that there is a strong relation between precipitation in Europe and the North Atlantic Oscillation (NAO), especially during wintertime (e.g., Hurrell, 1995; Rîmbu et al., 2002; Haylock and Goodess, 2004; Nissen et al., 2010; Pinto and Raible, 2012). Moreover, Casanueva et al. (2014) found a connection between extreme precipitation and the Atlantic Multidecadal Oscillation (AMO) during the whole year. The investigations of Bloomfield et al. (2018) revealed long-term changes in mean sea level pressure in the North Atlantic region and related

645 storminess over Europe, which might be an artifact of a rising number of available and assimilated observations in the last decades.

Data availability. The E–OBS data (Haylock et al., 2008) is online available after registration at https://www.ecad.eu/download/ensembles/ ensembles.php. The 20CR data (Compo et al., 2011) can be found on https://www.esrl.noaa.gov/psd/data/20thC_Rean/. HYRAS (Rauthe et al., 2013) can be requested at the German Weather Service (DWD). The RCM data (MiKlip data) will be made available via the CERA database (http://cera-www.dkrz.de/; last access: July 2019) of the German Climate Computing Center (DKRZ).

650

Appendix A: Statistical Quantities

The linear error in probability space L uses the difference of probabilities ΔC defined as:

$$\Delta C_{\rm r}(x) = ecdf_{mod,r}(x) - ecdf_{obs}(x) , \qquad (A1)$$

where $ecdf_{mod,r}$ is the empirical cumulative density function of the model run r, and $ecdf_{obs}$ that of the observation up to precipitation intensity x. The linear error in probability space L_r for a model run r is then defined as (Déqué, 2012; Wahl et al., 2017):

$$L_{\rm r} = \frac{1}{n} \cdot \sum_{x=1}^{n} |\Delta C_{\rm r}(x)| \,. \tag{A2}$$

L_r describes the mean value of ΔC_r over the entire range of precipitation intensities x grouped into n classes. Using absolute values avoids a compensation of positive and negative values. The better both distributions coincide, the lower the value of L_r.
 The ensemble mean of L_r is given by:

$$\overline{L} = \frac{1}{M} \sum_{r=1}^{M} L_r , \qquad (A3)$$

with M being the total number of simulation runs.

...

670

The model performance on different frequency intervals is further validated using the standard deviation of a gamma distribution σ_{Γ} (Wilks, 2006), which is given by:

$$665 \quad \sigma_{\Gamma}^2 = \alpha \beta^2 \,. \tag{A4}$$

In this formulation, α is the shape parameter of the gamma distribution, and β its scale parameter.

The quantile-quantile analysis uses the Pearson correlation coefficient (Wilks, 2006) is given by given by:

$$R = \frac{\sum_{k=1}^{N} \left\{ \left[y_k - \overline{y_k} \right] \cdot \left[x_k - \overline{x_k} \right] \right\}}{\sqrt{\sum_{k=1}^{N} \left[x_k - \overline{x_k} \right]^2} \cdot \sqrt{\sum_{k=1}^{N} \left[y_k - \overline{y_k} \right]^2}} ,$$
(A5)

with the data series x and y of length N. The range of R is $R \in [-1; +1]$ with a perfect anti-correlation at R = -1 and a perfect correlation at R = +1.

The Gumbel distribution signal-to-noise ration S2N in this study is defined as:

$$S2N = \frac{RV_{\rm T,Gumbel}}{CI_{90,\rm T}},\tag{A6}$$

with the return level RV of the Gumbel distribution at return period T divided by its 90 % confidence interval at T (Früh et al., 2010). Small values of S2N indicate a more uncertain estimate, high values a more robust one. The Gumbel distribution (Wilks, 2006) 675 is an extreme value type-I distribution and often used for return period estimation. Its cumulative density function (cdf) is given by:

$$F(x) = \exp\left(-\exp\left(-\frac{x-\beta}{\alpha}\right)\right) , \qquad (A7)$$

with the free parameters $\beta = \sigma \sqrt{6} \cdot \pi^{-1}$ and $\alpha = \overline{x} - \gamma \beta$, where σ is the standard deviation of the sample x assuming a normal distribution, and $\gamma = 0.57721$ Euler's constant. For x, usually a series of yearly maximum values is used. The relationship between the cdf and the return period T is given by (Wilks, 2006):

$$T = \frac{1}{1 - F(x)} .$$
 (A8)

The signal-to-noise ration S2N in this case is defined as-

$$S2N = \frac{RV_{\mathrm{T,Gumbel}}}{CI_{90,\mathrm{T}}} ,$$

with the return level *RV* of the Gumbel distribution at return period *T* divided by the 90 % confidence interval at *T* (Früh et al., 2010).
685 Small values of *S*2*N* indicate a more uncertain estimate, high values a more robust one.

Appendix B: ETCCDI quantities

Two out of the 27 indices introduced and recommended by the Expert Team on Climate Change Detection and Indices⁴ (ETCCDI; Karl et al., 1999; Peterson, 2005) are used in this study. R95pTOT describes the annual total precipitation sum of all values above the climatological 95% percentile of wet days (RR > 1 mm) during the reference period 1961–1990. The R95pTOT of the year k is defined as:

690

680

$$R95pTOT_{k} = \sum_{w=1}^{W} RR_{wk} \quad \forall RR_{wk} > RR_{p95} , \qquad (B1)$$

where RR_{wk} is the daily precipitation amount on a wet day during year k, RR_{p95} is the climatological 95% percentile, and W the total number of wet days in year k. Analogously, the R99pTOT is defined replacing the 95% with the 99% percentile.

$$R99pTOT_{k} = \sum_{w=1}^{W} RR_{wk} \quad \forall RR_{wk} > RR_{p99} .$$
(B2)

695 Appendix C: Trends and Significance

A Mann–Kendall Test (Mann, 1945; Kendall, 1955) is performed for the detection of trends and its related significance. To account for possible oscillations within long time series, we first split the complete time series into sub-series with a minimum

⁴http://etccdi.pacificclimate.org/

length of 10 years and up to over 100 years (trend matrix). The Mann-Kendall Test uses a standardized test statistic S_{τ} following a standard Gaussian distribution (SGD). S_{τ} is given by:

700
$$S_{\tau} = \begin{cases} \frac{\tau - 1}{\sqrt{\sigma_{\tau}^2}} &, \tau > 0 ,\\ 0 &, \tau = 0 ,\\ \frac{\tau + 1}{\sqrt{\sigma_{\tau}^2}} &, \tau < 0 . \end{cases}$$
(C1)

Here, τ is known as the Kendall's τ and σ_{τ}^2 is the variance of the standard Gaussian distribution (SGD). A detected trend is significant if S_{τ} lies within the upper and lower quantile z of the SGD at a given significance level α with $S_{\tau} \in [z_{\frac{\alpha}{2}}; z_{1-\frac{\alpha}{2}}]$ $S_{\tau} \in [z_{\frac{\alpha}{2}}\sigma_{\tau}; z_{1-\frac{\alpha}{2}}\sigma_{\tau}]$, respectively (Yue et al., 2002).

Yue et al. (2002) pointed out some weaknesses of the Mann–Kendall test in case of inherent autocorrelation. To avoid a distortion of the statistic by autocorrelation, Yue et al. (2002) presented the Trend–Free Pre–Whitening (TFPW) method. The first step is the estimation of a linear trend between two time steps t = i and t = j using the Theil-Sen Approach (TSA; Theil, 1950; Sen, 1968). The slope *b* of this linear regression is given by:

$$b = median\left(\frac{x_j - x_i}{j - i}\right), \forall i < j.$$
(C2)

In a second step, the original time series x is detrended by subtracting b at each time step t:

710
$$x'_t = x_t - b \cdot t$$
. (C3)

Afterwards, the lag-1 autocorrelation coefficient r_1 is removed from the trend-free series x':

$$x_t'' = x_t' - r_1 \cdot x_{t-1}' , \tag{C4}$$

where r_1 is given by:

$$r_{1} = \frac{\frac{1}{N-1} \cdot \sum_{i=1}^{N-1} \left(x_{i}' - \overline{x'} \right) \cdot \left(x_{i+1}' - \overline{x'} \right)}{\frac{1}{N} \cdot \sum_{i=1}^{N} \left(x_{i}' - \overline{x'} \right)^{2}} .$$
 (C5)

715 The modified TFPW time series x^* result results by re-adding the TSA-slope b:

$$x_t^* = x_t^{\prime\prime} + b \cdot t . \tag{C6}$$

This modified time series conserves the trend, but is free of autocorrelation. The Mann–Kendall Test is performed on the TFPW time series x*. According to Yue et al. (2002), TFPW has to be considered in cases with non-zero TSA-slope and significant lag-1 autocorrelation. The significance of a trend or autocorrelation is tested on the 90 % (α = 0.1), 95 % (α = 0.05), and 99 %
(α = 0.01) significance level.

Author contributions. FE, LAK, HF, and JGP designed the study. HF performed (parts of) the RCM simulations. LAK applied the dry-day correction. FE did the analysis and plots, and wrote the initial draft. All Authors contributed with discussions and revisions.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors thank the National Centers for Environmental Prediction (NCEP) for providing the 20CR data. We ac knowledge the E-OBS E-OBS data set from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu). We also thank the German Weather Service (DWD) for providing HYRAS. In addition, we thank the German Climate Computing Center (DKRZ, Hamburg) for computing and storage resources. We thank the BMBF for funding the MiKlip project (contract number 01LP1129A/D) and AON for funding the extreme weather events project. We acknowledge open access publishing fund of Karlsruhe Institute of Technology (KIT). Joaquim G. Pinto JGP thanks the AXA Research Fund for support. We thank the

reviewers for their valuable comments that helped to improve this study, and the handling editor for guidance throughout the entire process.

References

- Barbero, R., Fowler, H. J., Lenderink, G., and Blenkinsop, S.: Is the intensification of precipitation extremes with global warming better detected at hourly than daily resolutions?, Geophys. Res. Lett., 44, 974–983, https://doi.org/10.1002/2016GL071917, 2017.
- Berg, P., Haerter, J. O., Thejll, P., Piani, C., Hagemann, S., and Christensen, J. H.: Seasonal characteristics of the relationship between daily
 precipitation intensity and surface temperature, J. Geophys. Res. Atm., 114, 2009.
 - Berg, P., Feldmann, H., and Panitz, H.-J.: Bias correction of high resolution regional climate model data, J. Hydrol., 448-449, 80 92, https://doi.org/10.1016/j.jhydrol.2012.04.026, 2012.
 - Berg, P., Moseley, C., and Haerter, J. O.: Strong increase in convective precipitation in response to higher temperatures, Nat. Geosci., 6, 181, 2013.
- 740 Bloomfield, H. C., Shaffrey, L. C., Hodges, K. I., and Vidale, P. L.: A critical assessment of the long-term changes in the wintertime surface Arctic Oscillation and Northern Hemisphere storminess in the ERA20C reanalysis, Environ. Res. Lett., 13, 094004, https://doi.org/10.1088/1748-9326/aad5c5, 2018.
- Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Müller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., and Eade, R.: The decadal climate prediction project (DCPP)
 contribution to CMIP6, Geosci. Model Dev., 9, 3751–3777, https://doi.org/10.5194/gmd-9-3751-2016, 2016.
 - Brönnimann, S.: Weather extremes in an ensemble of historical reanalyses, in: Historical Weather Extremes in Reanalyses, pp. 7–22, Geographica Bernensia, g92 edn., https://doi.org/10.4480/GB2017.G92.01, 2017.
- Brönnimann, S., Romppainen-Martius, O., Franke, J., Stickler, A. N., and Auchmann, R.: Historical weather extremes in the "Twentieth Century Reanalysis, in: Weather extremes during the past 140 years, pp. 7–17, Geographica Bernensia, g89 edn., https://doi.org/10.4480/GB2013.G89.01, 2013.
- Cahill, N., Rahmstorf, S., and Parnell, A. C.: Change points of global temperature, Environ. Res. Lett., 10, 084002, https://doi.org/10.1088/1748-9326/10/8/084002, 2015.
 - Casanueva, A., Rodríguez-Puebla, C., Frías, M. D., and González-Reviriego, N.: Variability of extreme precipitation over Europe and its relationships with teleconnection patterns, Hydrol. Earth Syst. Sci., 18, 709–725, https://doi.org/10.5194/hess-18-709-2014, 2014.
- 755 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, https://doi.org/10.1007/s10584-006-9210-7, 2007.
- Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The twentieth century reanalysis project, Q. J. R. Meteorol. Soc., 137, 1–28, https://doi.org/10.1002/qj.776, 2011.
- Déqué, M.: Deterministic Forecasts of Continuous Variables, chap. 5, pp. 77–94, John Wiley & Sons Ltd, https://doi.org/10.1002/9781119960003.ch5, 2012.
- Dittus, A. J., Karoly, D. J., Lewis, S. C., Alexander, L. V., and Donat, M. G.: A Multiregion Model Evaluation and Attribution Study of Historical Changes in the Area Affected by Temperature and Precipitation Extremes, J. Climate, 29, 8285–8299, https://doi.org/10.1175/JCLI D-16-0164.1, 2016.
- Donat, M. G., Alexander, L. V., Herold, N., and Dittus, A. J.: Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations, J. Geophys. Res. Atm., 121, 11,174–11,189, https://doi.org/10.1002/2016JD025480, 2016.

Easterling, D. R., Evans, J. L., Groisman, P. Y., Karl, T. R., Kunkel, K. E., and Ambenje, P.: Observed Variability and Trends in Extreme Climate Events: A Brief Review, Bull. Am. Meteorol. Soc., 81, 417-426, https://doi.org/10.1175/1520-0477(2000)081<0417:OVATIE>2.3.CO;2, 2000.

- 770
 - Ehmele, F. and Kunz, M.: Flood-related extreme precipitation in southwestern Germany: development of a two-dimensional stochastic precipitation model, Hydrol. Earth Syst. Sci., 23, 1083-1102, https://doi.org/10.5194/hess-23-1083-2019, 2019.
 - Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, 2016.
- Feldmann, H., Früh, B., Schädler, G., Panitz, H.-J., Keuler, K., Jacob, D., and Lorenz, P.: Evaluation of the precipitation for South-775 western Germany from high resolution simulations with regional climate models, Meteorol. Z., 17, 455-465, https://doi.org/10.1127/0941-2948/2008/0295, 2008.
 - Feldmann, H., Schädler, G., Panitz, H.-J., and Kottmeier, C.: Near future changes of extreme precipitation over complex terrain in Central Europe derived from high resolution RCM ensemble simulations, Int. J. Climatol., 33, 1964–1977, https://doi.org/10.1002/joc.3564, 2013.
- Feldmann, H., Pinto, J. G., Laube, N., Uhlig, M., Moemken, J., Pasternack, A., Früh, B., Pohlmann, H., and Kottmeier, C.: Skill and 780 Added Value of the MiKlip Regional Decadal Prediction System for Temperature over Europe, Tellus A: Dynamic Meteorology and Oceanography, 0, 0, 2019.
 - Feser, F., Rockel, B., von Storch, H., Winterfeldt, J., and Zahn, M.: Regional Climate Models Add Value to Global Model Data: A Review and Selected Examples, Bull. Am. Meteorol. Soc., 92, 1181–1192, https://doi.org/10.1175/2011BAMS3061.1, 2011.
- 785 Field, C. B., Barros, V., Stocker, T. F., Qin, D., Dokken, D. J., Ebi, K. L., Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Allen, S. K., Tignor, M., and Midgley, P. M. e.: 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, https://doi.org/10.1017/CBO9781139177245, 582 pp., 2012.
- Fischer, E. M. and Knutti, R.: Observed heavy precipitation increase confirms theory and early models, Nat. Clim. Change, 6, 986, 790 https://doi.org/10.1038/NCLIMATE3110, 2016.
 - Folland, C. K., Boucher, O., Colman, A., and Parker, D. E.: Causes of irregularities in trends of global mean surface temperature since the late 19th century, Science Advances, 4, https://doi.org/10.1126/sciadv.aao5297, 2018.
 - Früh, B., Feldmann, H., Panitz, H.-J., Schädler, G., Jacob, D., Lorenz, P., and Keuler, K.: Determination of precipitation return values in complex terrain and their evaluation, J. Climate, 23, 2257-2274, 2010.
- 795 Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, J. Hydrol., 228, 113-129, https://doi.org/10.1016/S0022-1694(00)00144-X, 2000.
 - Hagemann, S., Loew, A., and Andersson, A.: Combined evaluation of MPI-ESM land surface water and energy fluxes, J. Adv. Model. Earth Syst., 5, 259-286, https://doi.org/10.1029/2012MS000173, 2013.

Haylock, M. R. and Goodess, C. M.: Interannual variability of European extreme winter rainfall and links with mean large-scale circulation,

- 800 Int. J. Climatol., 24, 759-776, https://doi.org/10.1002/joc.1033, 2004.
 - Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., 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. Atm., 113, https://doi.org/10.1029/2008JD010201, 2008.
 - Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanae, S.: Global flood risk under climate change, Nat. Clim. Change, 3, 816, https://doi.org/10.1038/nclimate1911, 2013.

- 805 Hofstra, N., Haylock, M., New, M., and Jones, P. D.: Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature, J. Geophys. Res. Atm., 114, 2009.
 - Hurrell, J. W.: Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation, Science, 269, 676–679, https://doi.org/10.1126/science.269.5224.676, 1995.
- Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Geor-
- 810 gopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Reg. Environ. Change, 14, 563–578, https://doi.org/10.1007/s10113-013-0499-2, 2014.
- 815 Journel, A. G. and Huijbregts, C. J.: Mining Geostatistics, Academic press, London, 600 pp., 1978.
- Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and von Storch, J.-S.: Characteristics of the ocean simulations in MPIOM, the ocean component of the MPI-earth system model, J. Adv. Model. Earth Syst., 5, 146–172, https://doi.org/10.1002/jame.20023, 2013.
- Kanamitsu, M., Alpert, J. C., Campana, K. A., Caplan, P. M., Deaven, D. G., Iredell, M., Katz, B., Pan, H.-L., Sela, J., and White,
 G. H.: Recent changes implemented into the global forecast system at NMC, Weather Forcast., 6, 425–435, https://doi.org/10.1175/1520-0434(1991)006%3C0425:RCIITG%3E2.0.CO;2, 1991.
 - Karl, T. R., Nicholls, N., and Ghazi, A.: Clivar/GCOS/WMO workshop on indices and indicators for climate extremes workshop summary, in: Weather Clim. Extrem., pp. 3–7, Springer, 1999.

Kendall, M. G.: Rank correlation methods, Charles Griffin, London, UK, 196 pp., 1955.

- 825 Kienzler, S., Pech, I., Kreibich, H., Müller, M., and Thieken, A. H.: After the extreme flood in 2002: changes in preparedness, response and recovery of flood-affected residents in Germany between 2005 and 2011, Nat. Hazards Earth Syst. Sci., 15, 505–526, https://doi.org/10.5194/nhess-15-505-2015, 2015.
 - Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate modeling on European scales:
- a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geosci. Model Dev., 7, 1297–1333, https://doi.org/10.5194/gmd-7-1297-2014, 2014.
 - Kröner, N., Kotlarski, S., Fischer, E., Lüthi, D., Zubler, E., and Schär, C.: Separating climate change signals into thermodynamic, lapse-rate and circulation effects: theory and application to the European summer climate, Clim. Dynam., 48, 3425–3440, https://doi.org/10.1007/s00382-016-3276-3, 2017.
- 835 Lenderink, G., Mok, H. Y., Lee, T. C., and Van Oldenborgh, G. J.: Scaling and trends of hourly precipitation extremes in two different climate zones–Hong Kong and the Netherlands, Hydrol. Earth Syst. Sci., 15, 3033–3041, 2011.

Mann, H. B.: Nonparametric Tests Against Trend, Econometrica, 13, 245-259, https://doi.org/10.2307/1907187, 1945.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themešl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, 1–34, https://doi.org/10.1029/2009RG000314, 2010.

- Marotzke, J., Müller, W. A., Vamborg, F. S. E., Becker, P., Cubasch, U., Feldmann, H., Kaspar, F., Kottmeier, C., Marini, C., Polkova, I., Prömmel, K., Rust, H. W., Stammer, D., Ulbrich, U., Kadow, C., Köhl, A., Kröger, J., Kruschke, T., Pinto, J. G., Pohlmann, H., Reyers, M., Schröder, M., Sienz, F., Timmreck, C., and Ziese, M.: MiKlip: a national research project on decadal climate prediction, Bull. Am. Meteorol. Soc., 97, 2379–2394, https://doi.org/10.1175/BAMS-D-15-00184.1, 2016.
- Martel, J.-L., Mailhot, A., and Brissette, F.: Global and Regional Projected Changes in 100-yr Subdaily, Daily, and Multiday Precipitation Extremes Estimated from Three Large Ensembles of Climate Simulations, J. Climate, 33, 1089–1103, https://doi.org/10.1175/JCLI-D-18-0764.1, 2020.

Merz, B., Elmer, F., Kunz, M., Mühr, B., Schröter, K., and Uhlemann-Elmer, S.: The extreme flood in June 2013 in Germany, La Houille Blanche, pp. 5–10, 2014.

- Mieruch, S., Feldmann, H., Schädler, G., Lenz, C.-J., Kothe, S., and Kottmeier, C.: The regional MiKlip decadal forecast ensemble for Europe: the added value of downscaling., Geosci. Model Dev., 7, https://doi.org/10.5194/gmd-7-2983-2014, 2014.
 - Moberg, A. and Jones, P. D.: Trends in indices for extremes in daily temperature and precipitation in central and western Europe, 1901–99, Int. J. Climatol., 25, 1149–1171, https://doi.org/10.1002/joc.1163, 2005.
- Moberg, A., Jones, P. D., Lister, D., Walther, A., Brunet, M., Jacobeit, J., Alexander, L. V., Della-Marta, P. M., Luterbacher, J., Yiou, P., Chen, D., Klein Tank, A. M. G., Saladié, O., Sigró, J., Aguilar, E., Alexandersson, H., Almarza, C., Auer, I., Barriendos, M., Begert, M., Bergström, H., Böhm, R., Butler, C. J., Caesar, J., Drebs, A., Founda, D., Gerstengarbe, F.-W., Micela, G., Maugeri, M., Österle, H., Pandzic, K., Petrakis, M., Srnec, L., Tolasz, R., Tuomenvirta, H., Werner, P. C., Linderholm, H., Philipp, A., Wanner, H., and Xoplaki, E.: Indices for daily temperature and precipitation extremes in Europe analyzed for the period 1901–2000, J. Geophys. Res. Atm., 111,
- 860 https://doi.org/10.1029/2006JD007103, 2006.

845

- Moorthi, S., Pan, H.-L., and Caplan, P.: Changes to the 2001 NCEP operational MRF/AVN global analysis/forecast system, in: Technical Procedures Bulletin 484, US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Office of Meteorology, Program and Plans Division, 2001.
- Mudelsee, M., Börngen, M., Tetzlaff, G., and Grünewald, U.: No upward trends in the occurrence of extreme floods in central Europe, Nature, 425, 166, https://doi.org/10.1038/nature01928, 2003.
- Mudelsee, M., Börngen, M., Tetzlaff, G., and Grünewald, U.: Extreme floods in central Europe over the past 500 years: Role of cyclone pathway "Zugstrasse Vb", J. Geophys. Res. Atm., 109, https://doi.org/10.1029/2004JD005034, 2004.
 - Müller, W. A., Baehr, J., Haak, H., Jungclaus, J. H., Kröger, J., Matei, D., Notz, D., Pohlmann, H., von Storch, J. S., and Marotzke, J.: Forecast skill of multi-year seasonal means in the decadal prediction system of the Max Planck Institute for Meteorology, Geophys. Res.
- 870 Lett., 39, https://doi.org/10.1029/2012GL053326, 2012.
 - Müller, W. A., Pohlmann, H., Sienz, F., and Smith, D.: Decadal climate predictions for the period 1901–2010 with a coupled climate model, Geophys. Res. Lett., 41, 2100–2107, https://doi.org/10.1002/2014GL059259, 2014.
 - Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., Bunzel, F., Esch, M., Ghosh, R., Haak, H., Ilyina, T., Kleine, T., Kornblueh, L., Li, H., Modali, K., Pohlmann, H., Roeckner, E., Stemmler, I., Tian, F., and Marotzke, J.: A Higher-
- 875 resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1. 2-HR), J. Adv. Model. Earth Syst., 10, 1383–1413, https://doi.org/10.1029/2017MS001217, 2018.

MunichRe: NatCatSERVICE, natcatservice.munichre.com/, accessed: 23 Aug 2018, 2017.

Nissen, K. M., Leckebusch, G. C., Pinto, J. G., Renggli, D., Ulbrich, S., and Ulbrich, U.: Cyclones causing wind storms in the Mediterranean: characteristics, trends and links to large-scale patterns, Nat. Hazards Earth Syst. Sci., 10, 1379–1391, https://doi.org/10.5194/nhess-10-1379-2010, 2010.

880

Nissen, K. M., Ulbrich, U., and Leckebusch, G. C.: Vb cyclones and associated rainfall extremes over Central Europe under present day and climate change conditions, Meteorol. Z., 22, 649–660, https://doi.org/10.1127/0941-2948/2013/0514, 2013.

O'Gorman, P. A.: Precipitation extremes under climate change, Current climate change reports, 1, 49–59, 2015.

- Pal, J. S., Giorgi, F., and Bi, X.: Consistency of recent European summer precipitation trends and extremes with future regional climate
 projections, Geophys. Res. Lett., 31, https://doi.org/10.1029/2004GL019836, 2004.
- Pauling, A. and Paeth, H.: On the variability of return periods of European winter precipitation extremes over the last three centuries, Clim. Past., 3, 65–76, https://doi.org/10.5194/cp-3-65-2007, 2007.
 - Peleg, N., Fatichi, S., Paschalis, A., Molnar, P., and Burlando, P.: An advanced stochastic weather generator for simulating 2-D high-resolution climate variables, J. Adv. Model. Earth Sy., 9, 1595–1627, https://doi.org/10.1002/2016MS000854, 2017.
- 890 Peterson, T. C.: Climate change indices, WMO Bull., 54, 83–86, 2005.
 - Pfahl, S., O'Gorman, P. A., and Fischer, E. M.: Understanding the regional pattern of projected future changes in extreme precipitation, Nat. Clim. Change, 7, 423, https://doi.org/10.1038/nclimate3287, 2017.
 - Pinto, J. G. and Raible, C. C.: Past and recent changes in the North Atlantic oscillation, WIRES Clim Change, 3, 79–90, https://doi.org/10.1002/wcc.150, 2012.
- Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G. H., Peubey, C., Thépaut, J.-N., Trémolet, Y., Hólm, E. V., Bonavita, M., Isaksen, L., and Fisher, M.: ERA-20C: An atmospheric reanalysis of the twentieth century, J. Climate, 29, 4083–4097, https://doi.org/10.1175/JCLI-D-15-0556.1, 2016.
 - Potts, J. M., Folland, C. K., Jolliffe, I. T., and Sexton, D.: Revised LEPS scores for assessing climate model simulations and long-range forecasts, J. Climate, 9, 34–53, https://doi.org/10.1175/1520-0442(1996)009<0034:RSFACM>2.0.CO;2, 1996.
- 900 Primo, C., Kelemen, F. D., Feldmann, H., Akhtar, N., and Ahrens, B.: A regional atmosphere–ocean climate system model (CCLMv5. 0clm7-NEMOv3. 3-NEMOv3. 6) over Europe including three marginal seas: on its stability and performance, Geosci. Model Dev., 12, 5077–5095, https://doi.org/10.5194/gmd-12-5077-2019, 2019.
- 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), Meteorol. Z., 22, 235–256, https://doi.org/10.1127/0941-2948/2013/0436, 2013.
- Reyers, M., Feldmann, H., Mieruch, S., Pinto, J. G., Uhlig, M., Ahrens, B., Früh, B., Modali, K., Laube, N., Moemken, J., Müller, W., Schädler, G., and Kottmeier, C.: Development and prospects of the regional MiKlip decadal prediction system over Europe: predictive skill, added value of regionalization, and ensemble size dependency, Earth Syst. Dynam., 10, 171–187, https://doi.org/10.5194/esd-10-171-2019, 2019.
- 910 Rhodes, R. I., Shaffrey, L. C., and Gray, S. L.: Can reanalyses represent extreme precipitation over England and Wales?, Q. J. R. Meteorol. Soc., 141, 1114–1120, https://doi.org/10.1002/qj.2418, 2015.
 - Rîmbu, N., Boroneanţ, C., Buţă, C., and Dima, M.: Decadal variability of the Danube river flow in the lower basin and its relation with the North Atlantic Oscillation, International Journal of Climatology, 22, 1169–1179, https://doi.org/10.1002/joc.788, 2002.
- Rockel, B., Will, A., and Hense, A.: The regional climate model COSMO-CLM (CCLM), Meteorol. Z., 17, 347-348,

⁹¹⁵ https://doi.org/10.1127/0941-2948/2008/0309, 2008.

- Schewe, J., Gosling, S. N., Reyer, C., Zhao, F., Ciais, P., Elliott, J., Francois, L., Huber, V., Lotze, H. K., Seneviratne, S. I., van Vliet, M. T. H., Vautard, R., Wada, Y., Breuer, L., Büchner, M., Carozza, D. A., Chang, J., Coll, M., Deryng, D., de Wit, A., Eddy, T. D., Folberth, C., Frieler, K., Friend, A. D., Gerten, D., Gudmundsson, L., Hanasaki, N., Ito, A., Khabarov, N., Kim, H., Lawrence, P., Morfopoulos, C., Müller, C., Müller Schmied, H., Orth, R., Ostberg, S., Pokhrel, Y., Pugh, T. A. M., Sakurai, G., Satoh, Y., Schmid, E., Stacke, T.,
- 920 Steenbeek, J., Steinkamp, J., Tang, Q., Tian, H., Tittensor, D. P., Volkholz, J., Wang, X., and Warszawski, L.: State-of-the-art global models underestimate impacts from climate extremes, Nature communications, 10, 1005, https://doi.org/10.1038/s41467-019-08745-6, 2019.
 - Schröter, K., Kunz, M., Elmer, F., Mühr, B., and Merz, B.: What made the June 2013 flood in Germany an exceptional event? A hydrometeorological evaluation, Hydrol. Earth Syst. Sci., 19, 309–327, https://doi.org/10.5194/hess-19-309-2015, 2015.
- 925 Sen, P. K.: Estimates of the Regression Coefficient Based on Kendall's Tau, J. Am. Stat. Assoc., 63, 1379–1389, https://doi.org/10.1080/01621459.1968.10480934, 1968.
 - Simmons, A. J., Berrisford, P., Dee, D. P., Hersbach, H., Hirahara, S., and Thepaut, J.-N.: A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets, Q. J. R. Meteorol. Soc., 143, 101–119, https://doi.org/10.1002/qj.2949, 2017.
- 930 Singer, M. B., Michaelides, K., and Hobley, D. E. J.: STORM 1.0: a simple, flexible, and parsimonious stochastic rainfall generator for simulating climate and climate change, Geosci. Model Dev., 11, 3713–3726, 2018.
 - Stephens, G. L. and Ellis, T. D.: Controls of global-mean precipitation increases in global warming GCM experiments, J. Climate, 21, 6141–6155, 2008.
 - Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast,
- 935 I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T., and Roeckner, E.: Atmospheric component of the MPI-M earth system model: ECHAM6, J. Adv. Model Earth Syst., 5, 422–446, https://doi.org/10.1002/jame.20015, 2013.
 - Stucki, P., Rickli, R., Brönnimann, S., Martius, O., Wanner, H., Grebner, D., and Luterbacher, J.: Weather patterns and hydro-climatological precursors of extreme floods in Switzerland since 1868, Meteorol. Z., 21, 531–550, https://doi.org/10.1127/0941-2948/2012/368, 2012.
- Stucki, P., Dierer, S., Welker, C., Gómez-Navarro, J. J., Raible, C. C., Martius, O., and Brönnimann, S.: Evaluation of downscaled wind
 speeds and parameterised gusts for recent and historical windstorms in Switzerland, Tellus A: Dynamic Meteorology and Oceanography, 68, 31 820, https://doi.org/10.3402/tellusa.v68.31820, 2016.
 - Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bull. Am. Meteorol. Soc., 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.
- Theil, H.: A rank-invariant method of linear and polynomial regression analysis, Proceedings of the Royal Netherlands Academy of Sciences, 53, Part I: 386–392, Part II:521–525, Part III: 1397–1412, https://doi.org/10.1007/978-94-011-2546-8_20, 1950.
- Toreti, A., Xoplaki, E., Maraun, D., Kuglitsch, F.-G., Wanner, H., and Luterbacher, J.: Characterisation of extreme winter precipitation in Mediterranean coastal sites and associated anomalous atmospheric circulation patterns, Nat. Hazards Earth Syst. Sci., 10, 1037–1050, https://doi.org/10.5194/nhess-10-1037-2010, 2010.
- Torma, C., Giorgi, F., and Coppola, E.: Added value of regional climate modeling over areas characterized by complex terrain–Precipitation over the Alps, J. Geophys. Res. Atm., 120, 3957–3972, https://doi.org/10.1002/2014JD022781, 2015.
- Trenberth, K. E., Dai, A., Rasmussen, R. M., and Parsons, D. B.: The changing character of precipitation, Bull. Am. Meteorol. Soc., 84, 1205–1218, 2003.

- Uhlemann, S., Thieken, A. H., and Merz, B.: A consistent set of trans-basin floods in Germany between 1952-2002, Hydrol. Earth Syst. Sci., 14, 1277, https://doi.org/10.5194/hess-14-1277-2010, 2010.
- 955 Ulbrich, U., Brücher, T., Fink, A. H., Leckebusch, G. C., Krüger, A., and Pinto, J. G.: The central European floods of August 2002: Part 1 Rainfall periods and flood development, Weather, 58, 371–377, https://doi.org/10.1256/wea.61.03A, 2003a.
 - Ulbrich, U., Brücher, T., Fink, A. H., Leckebusch, G. C., Krüger, A., and Pinto, J. G.: The central European floods of August 2002: Part 2 –Synoptic causes and considerations with respect to climatic change, Weather, 58, 434–442, https://doi.org/10.1256/wea.61.03B, 2003b.
- van den Besselaar, E. J. M., Haylock, M. R., van der Schrier, G., and Klein Tank, A. M. G.: A European daily high-resolution observational
 gridded data set of sea level pressure, J. Geophys. Res. Atm., 116, https://doi.org/10.1029/2010JD015468, 2011.
 - van der Wiel, K., Wanders, N., Selten, F. M., and Bierkens, M. F. P.: Added Value of Large Ensemble Simulations for Assessing Extreme River Discharge in a 2 °C Warmer World, Geophys. Res. Lett., 46, 2093–2102, https://doi.org/10.1029/2019GL081967, 2019.
- Wahl, S., Bollmeyer, C., Crewell, S., Figura, C., Friederichs, P., Hense, A., Keller, J. D., and Ohlwein, C.: A novel convective-scale regional reanalyses COSMO-REA2: Improving the representation of precipitation, Meteorol. Z., 26, 345–361, https://doi.org/10.1127/metz/2017/0824, 2017.
 - Ward, M. N. and Folland, C. K.: Prediction of seasonal rainfall in the north nordeste of Brazil using eigenvectors of sea-surface temperature, Int. J. Climatol., 11, 711–743, https://doi.org/10.1002/joc.3370110703, 1991.
 - Westra, S., Alexander, L. V., and Zwiers, F. W.: Global increasing trends in annual maximum daily precipitation, J. Climate, 26, 3904–3918, 2013.
- 970 Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, vol. 91 of *International Geophysics Series*, Academie Press, San Diego, California, USA, 2nd edn., 2006.
 - Yue, S., Pilon, P., Phinney, B., and Cavadias, G.: The influence of autocorrelation on the ability to detect trend in hydrological series, Hydrol. Process., 16, 1807–1829, https://doi.org/10.1002/hyp.1095, 2002.
- Zhang, X., Zwiers, F.and Li, G., Wan, H., and Cannon, A.: Complexity in estimating past and future extreme short-duration rainfall, Nat.
 Geosci., 10, https://doi.org/10.1038/NGEO2911, 2017.
- Zolina, O., Simmer, C., Kapala, A., Bachner, S., Gulev, S., and Maechel, H.: Seasonally dependent changes of precipitation extremes over Germany since 1950 from a very dense observational network, J. Geophys. Res. Atm., 113, https://doi.org/10.1029/2007JD008393, 2008. Zwiers, F. W., Alexander, L. V., Hegerl, G. C., Knutson, T. R., Kossin, J. P., Naveau, P., Nicholls, N., Schär, C., Seneviratne, S. I., and Zhang,

X.: Climate extremes: challenges in estimating and understanding recent changes in the frequency and intensity of extreme climate and

980

weather events, in: Climate Science for Serving Society, pp. 339-389, Springer, 2013.