[handwritten: 2.5 km-1 day precipitation data]

[handwritten: → The title a bit too futuristic.]

[handwritten: Why not use: "Dynamical downscaling of the xxx reanalysis data to generate selected locations across Europe".]

Towards the Development of a Pan-European Stochastic Precipitation Dataset

Lisa-Ann Kautz[1], Florian Ehmele[1], Patrick Ludwig[1], Hilke S. Lentink[1], Fanni D. Kelemen[1,2], Martin Kadlec[3], and Joaquim G. Pinto[1]

[1]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.
[2]now at: Institute for Atmospheric and Environmental Sciences, Goethe University Frankfurt, Frankfurt am Main, Germany.
[3]Impact Forecasting, Aon, Prague, Czech Republic.

Correspondence to: Lisa-Ann Kautz (lisa-ann.kautz@kit.edu)

[handwritten: mis-representing extreme values]

[handwritten: → I disagree with this. Stochastic models are notorious for]

Abstract.

Heavy precipitation leading to widespread river floods are one of the main natural hazards affecting Central Europe. Since extreme precipitation events associated with devastating floods have long return periods, long-term datasets are needed to adequately quantify the frequency and intensity of these events. As long-term observations of precipitation across Europe are

5 rare and not homogeneous in space nor time, they are generally not suitable to run hydrological models. In the present study, a combined approach is presented on how to generate a consistent precipitation dataset based on dynamical downscaling and post-processing statistics. Focus is given to five river catchments in Central Europe: Upper Danube, Elbe, Oder, Rhine, and Vistula. Reanalysis data are dynamically downscaled with a regional climate model and bias-corrected towards observations. Empirical quantile mapping was identified as one of the most suitable methods to correct the bias in model precipitation. For

10 most of the top ten precipitation events of large European river catchments, bias correction led to clear improvements towards the raw model data. However, results for Western European rivers (e.g., Rhine) are typically better than for Eastern European rivers (e.g., Vistula), which may also be associated with observational gaps for the latter. Two examples of severe river floods are presented in more detail: the Rhine river flood in winter 1995 and the flood in the Upper Danube and Vistula in June 2009. While the former was already well presented without bias correction, for the latter, bias correction improved underestimated

15 precipitation amounts in the Upper Danube but not in the Vistula catchment. In conclusion, this method can be applied to other extensive datasets towards the development of a Pan-European stochastic precipitation dataset.

[handwritten: not really logical or Verifiable.]

[handwritten: ① ② of]

[handwritten: investigated]

[handwritten: results cannot exceeds the boundary of observation?]

[handwritten: "compared to"]

1 Introduction

River floods are one of the most disastrous weather-related hazards in Central Europe, heading the list of the highest economic

20 losses (e.g., Alfieri et al., 2018). For example, the damage caused by the devastating flood in Germany in spring 2013 has been valued at about 12 billion Euro (Merz et al., 2014). Due to the high impact of flooding on economy, agriculture, infrastructure, transport as well as on human life, there is a high interest in quantifying the risk of flooding for Central Europe (e.g., Ward et al., 2011; Feyen et al., 2012; Jongman et al., 2014). However, such extreme events have long return periods (e.g., Pauling

[handwritten: more detail, quantitative narrative would be better. eg. well reproduced by what percent? Which physical quantity is well reproduced by what percent?]

[handwritten: not a good word choice (e.g. essential)]

[handwritten: too general — I need reference for this]

and Paeth, 2007; Hirabayashi et al., 2013) and thus are rarely presented in short-term (observational) datasets. Therefore, it
is expedient to use historical century-long datasets to estimate flooding risk (e.g., Feyen et al., 2012), which is generally re-
lated to the occurrence of heavy and/or long-lasting precipitation (e.g., Maddox et al., 1979; Hilker et al., 2009). Long-term *[handwritten: contains]*
observational records of precipitation are quite heterogeneous across Europe. Thus, they cannot be used to run hydrological
5 rainfall-runoff models, whose simulations are needed to quantify the flooding risk for Europe. On the other hand, reanalysis
provide homogeneous datasets covering long time periods with the limitation of a coarse resolution. Approaches to overcome *[handwritten: the]*
the problem of small sample sizes are either to develop stochastic precipitation models (e.g., Richardson, 1981; Ehmele and *[handwritten: the]*
Kunz, 2018) or to downscale centennial reanalysis datasets (e.g., Stucki et al., 2016). Within the Miklip project ("decadal
climate predictions"; Marotzke et al., 2016), a regional component of the decadal prediction system was developed for Europe
10 (Feldmann et al., submitted), leading to the generation of 5,800 years of regional climate model hindcast data for the last 60
years.

Dynamical downscaling of long-term reanalysis with or without subsequent bias correction was used in a variety of earlier
studies. An overview of statistical and dynamical downscaling approaches and a discussion of their advantages and disadvan-
tages are given in Maraun et al. (2010). In general, the magnitude, the temporal, and the spatial variability of a downscaled
15 variable should be represented correctly. In particular, Maraun et al. (2010) found that the performance of downscaled precipita-
tion depends on the underlying synoptic situation. Thus, precipitation associated with frontal systems can be better represented
in comparison to precipitation gained by convective systems, which may be relevant for summer flooding. As an example, *[handwritten: mostly]*
the Twentieth Century Reanalysis dataset (20CR) was dynamically downscaled with the Weather Research and Forecasting
(WRF) model to investigate gustiness in Switzerland (Stucki et al., 2016) and the Lago Maggiore flood in 1868 (Stucki et al., *[handwritten: not a good word choice ??]*
20 2018). While the timing and magnitude of the peak gusts are not well captured by the model, the simulated precipitation pat-
tern agrees well with observations. Both studies did not apply any bias correction method. A downscaling with add-on bias
correction was used in e.g., Dobler and Ahrens (2008) or Fang et al. (2015). Dobler and Ahrens (2008) compared different
downscaling approaches for precipitation in Europe and South Asia as well as different bias correction methods (namely quan-
tile mapping and local intensity scaling). They concluded that dynamical downscaling with a regional climate model (RCM) in
25 combination with a bias correction method is most suitable to simulate precipitation in Europe. Fang et al. (2015) focused more
on the comparison of different bias correction methods and found that empirical quantile mapping and power transformation
performed best for precipitation. However, they mentioned that the selection of an accurate correction method turns out to be
case sensitive.

This study is a proof of concept where we show how to generate a precipitation dataset with a resolution of 25 km, which
30 is high enough to be used as input for hydrological models (e.g., Maraun et al., 2010). We address the following research
questions: *[handwritten: the]* *[handwritten: → not for the urban floods though]*

- *Which bias correction method is most suitable to correct model precipitation?* *[handwritten: and what is its limitation?]*
- *Are historical precipitation events related to floods represented adequately in the dataset?*
- *~~What are the limitations of the approach?~~*

[handwritten: How do you define the similarity of the pattern?]

We apply a dynamical downscaling with an RCM (COSMO-CLM, Rockel et al., 2008), where reanalysis are used as initial and boundary conditions. In a second step, a bias correction towards observations is applied. The focus is laid on large European river catchments, namely Upper Danube, Elbe, Oder, Rhine, and Vistula (Fig. 1). ~~Although, temperature was also considered, as it is a necessary input variable for hydrological models to calculate discharges, we focus on precipitation in this study and~~

5 test its representation for historical flood events. The top ten flood events for each catchment are considered and two events are discussed in detail, namely the ~~winter flood in~~ 1995 ~~which affected the~~ Rhine, as well ~~as the summer flood in~~ 2009, ~~which inundated parts of the~~ Danube and Vistula ~~catchments.~~

This paper is structured as follows: The datasets which are used in this study are introduced in Sect. 2. Methods are presented in Sect. 3, including the dynamical downscaling approach as well as the tested bias correction methods. The bias correction

10 methods are evaluated in Sect 4 and their added value for historical extreme precipitation events is discussed in Sect. 5. Results for the case studies are given in Sect. 6. The general findings are discussed in Sect. 7. Finally, the conclusions are summarized in the Sect. 8.

2 Datasets

15 ### 2.1 Observational data

As reference data for the bias correction ~~as well as for a comparison to downscaled model data~~ E-OBS data is considered, which is a European land-only daily high resolution gridded dataset for precipitation, surface temperature and sea level pressure (Haylock et al., 2008; Van den Besselaar et al., 2011). In this study, E-OBS v17 is used, which ~~shows some improvements towards older versions~~ due to updates and the inclusion of new observational stations (e.g., ~~for~~ Poland). A detailed comparison

20 between older versions and v17 can be found at the official homepage[1]. In this study, we consider daily precipitation sums on a $0.22°$ resolution grid (about 25 km) for the time period from 1950 to 2015.

For a quick quality control of E-OBS we additionally use the HYRAS dataset (German: **HY**drologische **RAS**terdatensätze; hydrological raster datasets) provided by the German Weather Service (Deutscher Wetterdienst, DWD). HYRAS is a high-resolved gridded observational dataset of daily rainfall totals based on several thousand ground-based climate stations (Rauthe

25 et al., 2013). Using a specific algorithm, the more or less unevenly distributed observations are interpolated to regular grids of $1 \times 1\,\text{km}^2$ and $5 \times 5\,\text{km}^2$. During this interpolation, elevation, exposition and climatology of a specific grid point are considered. HYRAS covers the state area of Germany and surrounding catchments of the rivers Rhine, Danube, and partly Elbe and Oder in the neighboring countries. HYRAS is available for the period 1951–2006, only. Note that HYRAS data are inhomogeneous due to the changing number, location and instrumentation of the used observations over the years. Furthermore, there is a

30 certain bias in precipitation totals especially over complex terrain, where the number of observations is limited (Kunz, 2011; Ehmele and Kunz, 2018). In this study we use the 5 km version of HYRAS ~~interpolated to~~ the ~~E-OBS grid~~.

---

[1]http://cib.knmi.nl/mediawiki/index.php/Compare_E-OBS_v17.0_and_v16.0

The bias is not defined by the correlation coefficient. Please also provide the slope of the regression passing through the origin.

We calculated the correlation coefficients between E-OBS and HYRAS daily precipitation sums for the Rhine, for the northern part of the Elbe catchment, and for the western part of the Upper Danube catchment. For the Rhine, we found a correlation coefficient of 0.9908, for the Elbe of 0.9876, and for the Danube of 0.9790. Thus, the daily precipitation sums of E-OBS and HYRAS are highly correlated. The slightly lower correlation for the Danube can be explained by less well represented orographic effects.

2.2 Reanalysis data

It is good idea to define the reanalysis data here. What does it record? How is it produced?

Reanalysis data provide the initial and boundary conditions for the RCM. Since reanalysis are not affected by changes in the data assimilation system as it is the case for archived weather analyses from operational forecasting systems, a homogeneous record of the past atmospheric evolution can be produced (Dee et al., 2011).

Rephrase please. I don't understand this sentence

10 Two global reanalysis datasets provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) are considered: ERA-Interim (Dee et al., 2011) and ERA-20C (Poli et al., 2016). ERA-Interim is available on a horizontal resolution of approximately 80 km (T255 spectral) and covers the time period from 1 January 1979 onwards (Dee et al., 2011). Due to enhancements in the data assimilation system, ERA-Interim shows improvements towards the previous generation ERA-40, for example, in the representation of the hydrological cycle (Dee et al., 2011). As mentioned in the introduction, the use of

15 centennial datasets is expedient to investigate extreme events, since these events have long return periods. A centennial record of atmospheric data is provided by the ERA-20C dataset, covering the period from 1900 to 2010 (Poli et al., 2016). ERA-20C products are available on a horizontal resolution of approximately 125 km (T159 spectral). Further information about ERA-Interim as well as ERA-20C can be found at the official ECMWF homepage[2].

3 Methods

So, you are going from 80~125 km to 25 km? 25 km seems to be too coarse. Couldn't you go any finer scale? (e.g. 5 km)

20 With the aim of generating a homogeneous long-term dataset of daily precipitation sums at 25 km resolution, reanalysis (ERA-Interim and ERA-20C) are dynamically downscaled with an RCM. Afterwards, the bias of the model precipitation is corrected towards observations (E-OBS).

3.1 Dynamical downscaling

Downscaling describes a procedure, in which information at larger scales are used to make predictions at local scales (e.g.,

25 Fowler et al., 2007). In general, downscaling can be classified in three categories namely statistical, dynamical and statistical-dynamical downscaling (e.g., Gutmann et al., 2012; Reyers et al., 2015). In this study, we apply dynamical downscaling, meaning that an RCM is run on a sub-domain of the utilized large-scale forcing data (e.g., Trzaska and Schnarr, 2014). The added value of the high resolution RCM compared to global climate models with coarser resolution are discussed e.g., in Feser et al. (2011). One of the key benefits is the better representation of hydrological variables particularly over areas with complex

30 terrain, which is important to accurately model precipitation associated with flood events (Frei et al., 2000).

"improved"

---

[2]https://www.ecmwf.int/en/forecasts/datasets/archive-datasets/browse-reanalysis-datasets
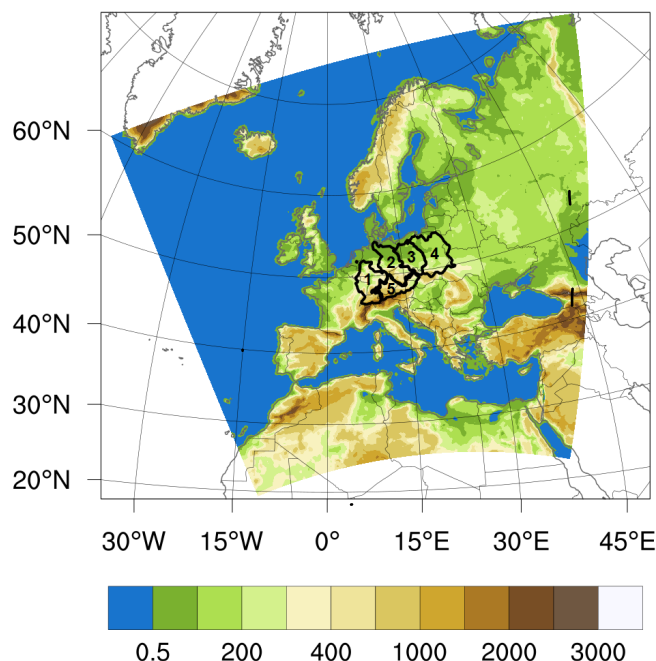
Figure 1. Surface height (in m) within the model domain. Areas marked by the black contours show the investigated river catchments, namely Rhine (1), Elbe (2), Oder (3), Vistula (4), and Upper Danube (5).

In this study, we used the non-hydrostatic COSMO model in climate mode (COSMO-CLM[3], Consortium for Small-Scale Modeling Climate Limited-area Model, thereafter CCLM; Rockel et al., 2008) in its version 5.08. CCLM is the community model of the German regional climate research community jointly further developed by the CLM-Community. In contrast to the COSMO model operationally used by the DWD, CCLM is run without data assimilation of observations and without

5   latent-heat nudging of radar data. Our CCLM simulations are performed on a 0.22° x 0.22° rotated grid with 32 unevenly distributed vertical levels (highest vertical resolution in the boundary layer) covering the EURO-CORDEX[4] domain (see Fig. 1). Physical parameterizations are used for convection (Tiedtke, 1989), subgrid-scale turbulence (Baldauf et al., 2011), longwave and shortwave radiation (Ritter and Geleyn, 1992), and microphysics (Doms et al., 2011). The initial and boundary conditions are provided by either 6-hourly ERA-Interim (1979-2015) or ERA-20C data (1900-2010). Hereafter, the abbreviation ERAI-

10   CCLM (ERA20C-CCLM) is used for the CCLM run in which ERA-Interim (ERA-20C) data were used as initial and boundary conditions. CCLM has already proved its skill to simulate extreme precipitation associated with major river floods in Europe (e.g., Kelemen et al., 2016; Huebener et al., 2016).

RCM alone? or when it is combined with the bias-correction?

performance

---

[3]www.cosmo-model.org
[4]www.euro-cordex.net

[handwritten: I am not saying that this method is wrong, but it significantly removes the model's flexibility.]

[handwritten: I think that this method bounds the downscaled values within the boundary of the observation. → Isn't the primary purpose of downscaling to detect the unseen extremes?]

3.2 Bias correction

With the dynamical downscaling approach we obtain precipitation data at the aimed resolution of 25 km. However, hydrological models are calibrated with observations, making them sensitive towards the forcing data. Thus, a bias correction of the RCM output is necessary. Moreover, it is known that RCMs produce too many wet days with low intensities (below 0.1 mm), known

5   as drizzle effect (e.g., Feldmann et al., 2008). Therefore, a previous dry-day correction had to be applied to match the number of dry days in the simulated dataset to the number of dry days in the observational dataset.

In line with Berg et al. (2012), dry days were defined as days with a daily precipitation sum below 0.1 mm. The dry-day correction was done separately for each month and included the following steps: For each grid point, the dry days were counted both in the observational data and the model run. Thereafter, for dry days in the model run, being wet days in the observations,

10   a small artificial precipitation amount was added. For wet days in the model run, being dry days in the observations, the model precipitation was reduced to an amount below 0.1 mm. This was done until the number of dry days in the model run was identical to the observed number of dry days.

The bias correction is applied to the dry-day-corrected data. Five commonly used bias correction methods were tested (e.g., Fang et al., 2015):

15   - linear scaling (LS)

- local intensity scaling (LOCI)

- power transformation (PT)

- empirical quantile mapping (EQM)

- quantile mapping with gamma distribution (GQM)

[handwritten: always starting from the beginning of the time series? I think the result of this process will differ depending on the location in the t.s. of the removal/adding of rainfall.]

20   A description of these approaches as well as references to further literature can be found in Appendix A. The needed distributions or scaling factors were calculated for each grid point by considering the eight surrounding grid points. This was done to ensure that adjacent grid points do not have independent precipitation amounts. Moreover, the distributions and scaling factors can be calculated with different calibrations, meaning e.g., the use of different time periods. Here, we compared scaling factors and distributions on a monthly, seasonal, and half-yearly basis.

25   The time period of ERA-Interim that is covered by E-OBS data (1979 to 2015) is declared as the reference period one (RPI). For ERA-20C, only the second half of the 20th century is covered by E-OBS. Therefore, a second reference period (RPII) is defined (1950 to 2010).

In the next section, the bias correction methods and calibrations are compared against each other for RPI and RPII, respectively. This comparison is done to find a bias correction method which is suitable for our approach. We do not aim to rank these

30   methods in an absolute sense. Studies which deal with a more general comparison of bias correction methods (nevertheless with focus on specific regions) are done, for instance, by Lafon et al. (2013).

Performance

a bit verbose. Simplify if possible.

4 Evaluation of bias correction methods in a climatological perspective

In this section, bias correction and calibration methods are evaluated with the aim of identifying a method which leads to improvements in the representation of precipitation in all investigated river catchments. First, we compare time series by calculating statistical variables, then we show the horizontal distribution of the mean annual precipitation during the time period from 1979 to 2015.

5   The skill of the different bias correction methods is quantified by the root mean square error (RMSE), Pearson's correlation coefficient (R) as well as the skill score (S) of Taylor (2001). The calculation of the skill score can be found in Appendix B. S and R values close to 1 and low values of the RMSE indicate a high resemblance between observations and the (corrected) model output. To calculate the statistical measures, time series (RPI for ERAI-CCLM and RPII for ERA20C-CCLM) of

10   catchment means were considered. For the Upper Danube, the results are illustrated in Table 1. For all three statistical measures, the monthly calibration leads to the best results in comparison with the seasonal and the half-yearly calibration. All bias correction methods reduce the RMSE compared to the raw model runs. The lowest RMSE can be found for PT with a monthly calibration. The highest values of S and R result for EQM with a monthly calibration. However, S and R of PT and GQM differ only at the third decimal place, which means that these methods lead to similar improvements as EQM. These statistical

15   patterns are also illustrated as Taylor diagram in Fig. 2. It can be seen that the centered root mean square difference (cRMSD, formula in Appendix B), R, and S are quite similar for the raw model and all bias-corrected ERAI-CCLM runs (Fig. 2 a). This is also true for ERA20C-CCLM (Fig. 2 b). For the Elbe catchment, a high correlation and skill is also found with EQM, while lowest RMSE values result from GQM (Table S1 in the *supplemental material*). For the Oder and the Vistula catchments, PT shows slightly better results in skill and correction than EQM, however, EQM has the lowest RMSE (Table S2 and S4).

20   The result for the Vistula and the Oder catchments are less meaningful compared to the other river catchments. This can be particularly attributed to the lower density of stations in Poland and surrounding areas than in other countries leading to a strong smoothing of peaks in the interpolation step and thus to an underestimation of precipitation amounts (shown for temperature by Kyselỳ and Plavcová, 2010). For the Rhine catchment, the correlation is higher for EQM, while the skill is higher for PT (Table S3). Moreover, lowest RMSE values are found for GQM.

25   In conclusion, EQM as well as PT with a monthly calibration lead to similar improvements towards the uncorrected model runs. We decided to use EQM, since previous studies have also found that the most comprehensive correction was achieved with EQM compared to distribution-based QM (as e.g. GQM), both for linear as well as non-linear methods (e.g., Lafon et al., 2013). In addition, PT bases on the results from LOCI, meaning that more computational effort is necessary for this bias correction method.

30   In the following, we compare the horizontal distribution of precipitation in Europe from the model output with and without bias correction. The mean annual precipitation from 1979 to 2015 for E-OBS, and the differences to uncorrected ERAI-CCLM and to ERAI-CCLM corrected with monthly EQM is shown in Fig. 3. In ERAI-CCLM (without bias correction), there are areas with overestimated precipitation and areas with underestimated precipitation (Fig. 3 b). Especially in the Alps, precipitation is overestimated, while, for example, Poland shows a weaker overestimation compared to observations. An underestimation of

Did you calculate R², RMSE, & S against EOBS? or HYDRAS?

I suggest to show the results of all test sites using different markers.

Not much differences between methods. How can you justify one method's superiority over others?

Don't you think that 0.5 of corr. coef is too low?
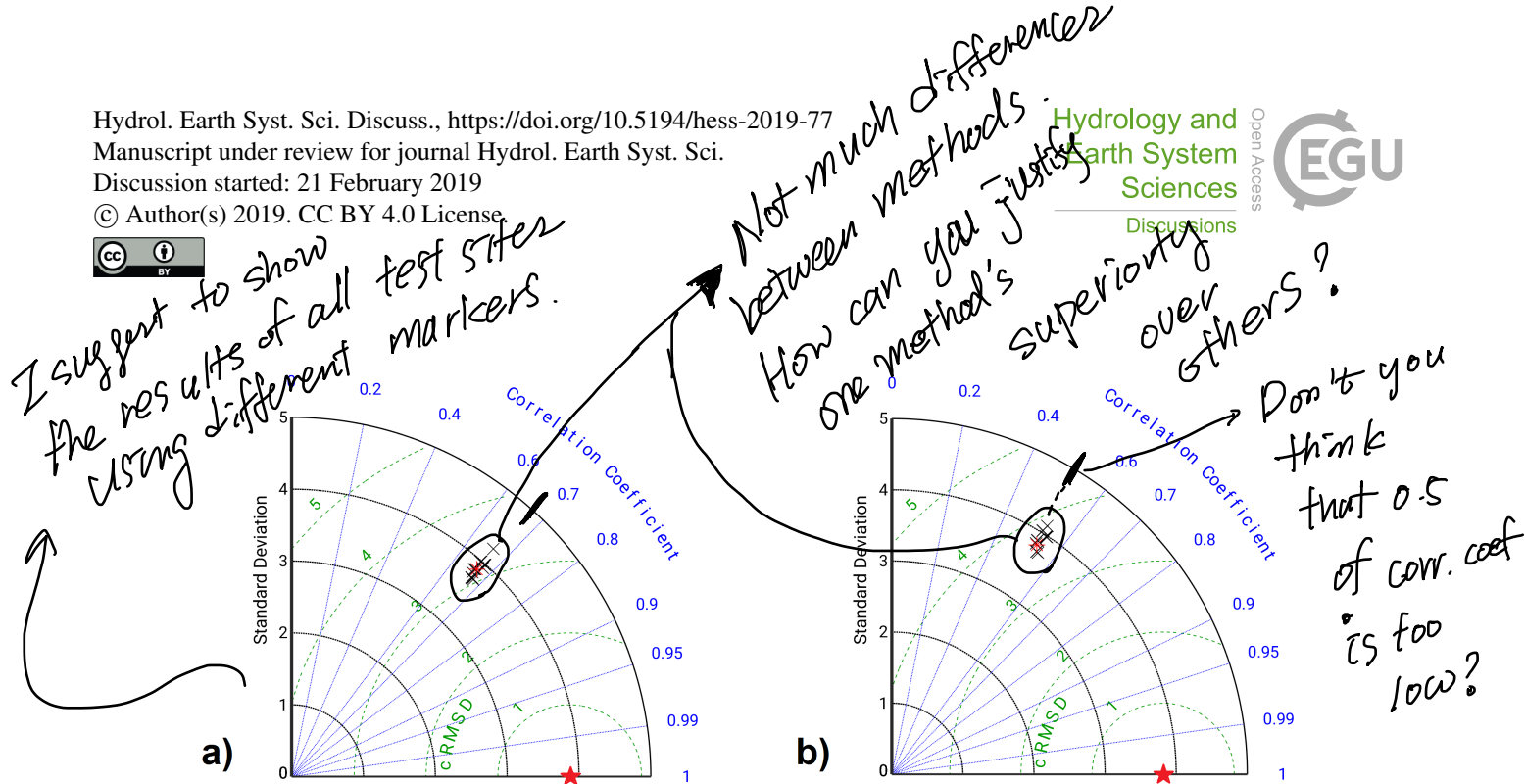
Figure 2. Diagram according Taylor (2001) for displaying statistical variables. Radial black lines mark values of the standard deviation, blue lines show values of the correlation coefficient and green lines show values for the centered RMSD. Results for the Upper Danube are shown. The red star shows E-OBS. In a) the red cross shows ERAI-CCLM and the black crosses show bias-corrected ERAI-CCLM runs. In b) the red cross shows ERA20C-CCLM and the black crosses show bias-corrected ERA20C-CCLM runs.

precipitation can be seen, for example, over Ireland or northern Italy. With bias correction, most areas in Europe are associated with an overestimation of precipitation, except Poland (Fig. 3 c). Thus, the overestimation of mean annual precipitation in the Alps is still present in the bias-corrected run but with reduced magnitude. Therefore, the bias correction (monthly EQM) leads to some improvements towards the raw model run. *To me, not enough.*

5   In contrast to ERA-Interim, ERA-20C covers also the first half of the century, where E-OBS data is not available. Thus, the bias correction for the first half of the century in ERA20C-CCLM is done with the scaling factors and distributions derived from the precipitation data of the second half of the century. To justify this procedure, the intensity spectra for the five river catchments for the first and the second half of the century are compared. For the Upper Danube, the intensity spectrum is shown in Fig. 4. The occurrence of precipitation events with amounts over 50 mm is overestimated in the uncorrected ERA20C-CCLM

10   run compared to E-OBS. However, the intensity spectrum of ERA20C-CCLM for the time period from 1900 to 1949 is similar to the time period from 1950 to 2010. Thus, it seems to be appropriate to correct the bias in precipitation for the first half of the century with the scaling factors and distributions from the second half of the century.

     We conclude from the climatological perspective that bias correction leads to improvements towards the raw model outputs, especially with EQM on a monthly basis. In the next two sections, the added value of bias correction for specific extreme

15   precipitation events is investigated.

5   Added value of bias correction for extreme precipitation events

In this section, we discuss the added value of the bias correction for the top ten events of the five investigated river catchments. The top ten events were identified by calculating the catchment-averaged 7-day running mean of daily precipitation sums

Table 1. Root mean square error (RMSE), Pearson's correlation coefficients (R) and Taylor's skill scores (S) for the raw model outputs plus for the model outputs with different bias corrections just for the Upper Danube catchment: linear scaling (LS), local intensity scaling (LOCI), power transformation (PT), empirical quantile mapping (EQM), and quantile mapping with gamma distribution (GQM). Results of ERAI-CCLM based on RPI and of ERA20C-CCLM on RPII. Best values of each measure are print in bold.

	ERAI-CCLM			ERA20C-CCLM		
	RMSE	R	S	RMSE	R	S
raw	3.353	0.666	0.946	3.844	0.534	0.950
$LS_{monthly}$	3.169	0.663	0.953	3.725	0.531	0.953
$LS_{seasonal}$	3.168	0.663	0.952	3.731	0.529	0.952
$LS_{half-yearly}$	3.157	0.661	0.951	3.713	0.528	0.952
$LOCI_{monthly}$	3.231	0.663	0.951	3.811	0.530	0.950
$LOCI_{seasonal}$	3.232	0.662	0.951	3.817	0.528	0.949
$LOCI_{half-yearly}$	3.219	0.660	0.951	3.800	0.526	0.949
$PT_{monthly}$	**3.066**	0.676	0.958	**3.599**	0.546	0.962
$PT_{seasonal}$	3.072	0.674	0.958	3.601	0.545	0.962
$PT_{half-yearly}$	3.069	0.675	0.958	3.602	0.545	0.961
$EQM_{monthly}$	3.172	**0.678**	**0.960**	3.715	**0.551**	**0.964**
$EQM_{seasonal}$	3.185	0.675	0.959	3.725	0.548	0.962
$EQM_{half-yearly}$	3.195	0.673	0.958	3.739	0.545	0.960
$GQM_{monthly}$	3.072	0.674	0.958	3.603	0.544	0.961
$GQM_{seasonal}$	3.082	0.672	0.956	3.612	0.542	0.960
$GQM_{half-yearly}$	3.090	0.670	0.955	3.617	0.540	0.958

So, RMSE is around 3.2 mm/day. Don't you think this degree of residual is enough to cause significant errors in the hydrologic modeling?

→ How does this results compare with those of other studies?

from E-OBS between 1979 and 2010. The resulting events are listed in Table 2. Then, we accumulated the precipitation sums over 14 days (9 days before and 4 days after the day of the peak value) for the 50 heavy precipitation events. We define "well captured" ("very well captured") if the model precipitation is within plus/minus 20% (10%) above/below the observed precipitation. To call the bias correction helpful, we expect an improvement of at least −33% of the initial deviation. Events

5   which are not well captured in the ERAI-CCLM and for which bias correction does not have an added value are pinrted in italic (Table 2). Events which are already well presented in the model run (indicating the accuracy of the model), so that bias correction hardly lead to further improvements are printed in standard letters. Events printed in bold letters benefit from the bias correction. The very well captured events are underlined.

     There is a clear east-west gradient in the accuracy of the bias-corrected ERAI-CCLM run. Eight of the ten events are well

10   captured for the Rhine and the Upper Danube, respectively. For the Rhine, the two bad cases are related to timing errors (cf. Fig. S1). For the Danube, the underestimation in the two not-well-captured events is related to spatial shifts and/or errors in the representation of dynamical processes (cf. Fig. S2). For the Elbe, four cases are not well captured by the model, even with
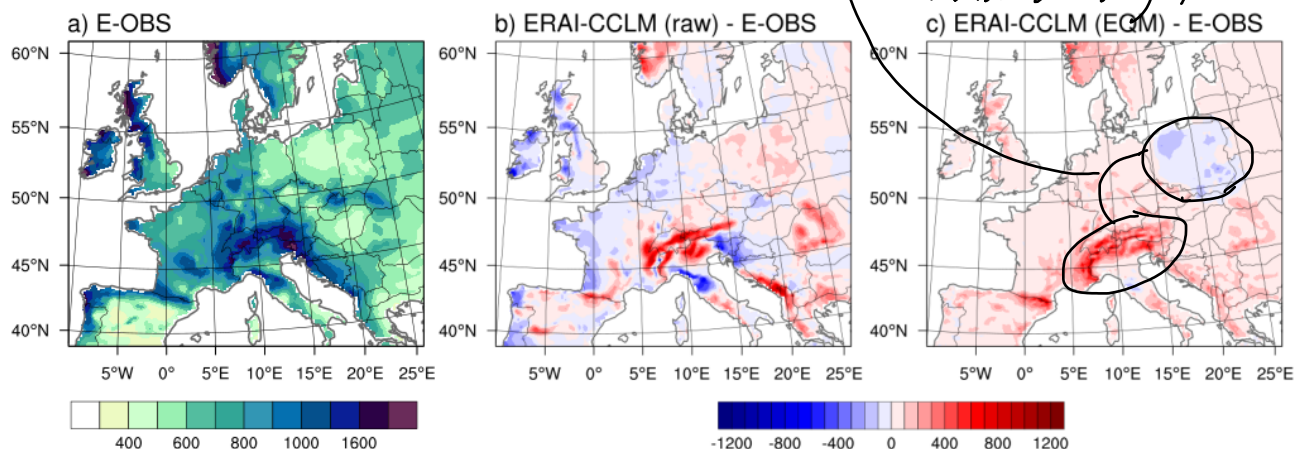
Figure 3. Mean annual precipitation (in mm a^{-1}) during the time period from 1979 to 2015 for a) E-OBS, as well as corresponding difference plots for b) ERAI-CCLM, and c) ERAI-CCLM bias-corrected with monthly EQM.
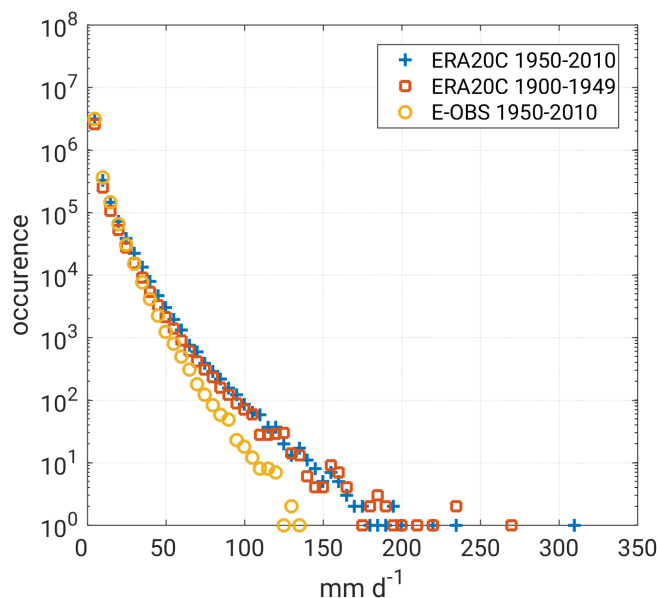
Figure 4. Precipitation intensity spectrum including all grid points within the Upper Danube catchment. Yellow circles illustrate E-OBS for the time period from 1950 to 2010. Blue crosses (orange squares) illustrate uncorrected ERA20C-CCLM for the time period from 1950 to 2010 (1900 to 1949).

bias correction, which is related to spatial shifts and the incorrect representation of dynamical processes (cf. Fig. S2). This is also true for the Oder (cf. Fig. S3). Here, one bad case (5 August 2006) results from a timing error (not shown). For the Vistula, precipitation peaks are not well presented in most of the cases, again as a consequence of spatial shifts and errors in the representation of dynamical processes (cf. Fig. S3). The fact that the Vistula floods are not well captured, can have

Table 2. Top 10 events of the five investigated river catchments for the time period from 1979 to 2010. The events are identified as the ten highest catchment-averaged 7-day running means of precipitation from E-OBS. The indicated dates (YYYY-MM-DD) address the timing of the peak value. Events where bias correction (EQM monthly) do not have an added value and the ERAI-CCLM shows large differences towards observations are printed in italic (-). Events which are already captured well in ERAI-CCLM, so that bias correction cannot lead to a further improvement, are printed standard (0). Events where bias correction has an added value are in bold print (+). The underlined events are very well captured.

Top 10	Rhine		Danube		Elbe		Oder		Vistula	
1	**1991-12-19**	(+)	*2002-08-09*	(-)	*2002-08-09*	(-)	*1997-07-06*	(-)	*2010-09-03*	(-)
2	**1998-10-26**	(+)	*1979-06-16*	(-)	**1983-08-04**	(+)	**1981-07-21**	(+)	*2006-08-09*	(-)
3	**1986-10-22**	(+)	**1981-07-18**	(+)	1986-12-30	(0)	**1996-07-09**	(+)	*1992-09-04*	(-)
4	1995-01-25	(0)	**2009-06-22**	(+)	*1997-07-18*	(-)	*1995-08-31*	(-)	*2009-06-22*	(-)
5	1990-02-13	(0)	**1985-08-04**	(+)	**1981-07-21**	(+)	*2001-07-18*	(-)	*1997-07-06*	(-)
6	1993-12-22	(0)	**1995-08-29**	(+)	**1998-10-29**	(+)	**2010-07-20**	(+)	2007-07-07	(0)
7	*1983-05-23*	(-)	**2002-03-21**	(+)	*2010-09-26*	(-)	2002-08-13	(0)	*1995-09-01*	(-)
8	2004-01-11	(0)	**2006-08-04**	(+)	*1995-08-29*	(-)	*2006-08-05*	(-)	2001-07-23	(0)
9	*1984-09-07*	(-)	**2003-10-06**	(+)	**1996-07-08**	(+)	1986-06-02	(0)	**2010-05-15**	(+)
10	1983-11-27	(0)	**1991-07-30**	(+)	2001-09-10	(0)	**1980-07-06**	(+)	*1980-10-11*	(-)

two reasons. First, observational precipitation peaks may be smoothed dramatically in the interpolation process due to a low density of stations in Poland and adjacent areas (Kyselỳ and Plavcová, 2010). Second, there are gaps in the long-term records for stations within the Vistula catchment. We counted the missing data in the E-OBS dataset and found up to 70% of missing data at certain locations within the Oder and the Vistula catchments. Thus, we assume that for the Eastern European river catchments it is not the method that fails, but the representation of precipitation in the observational dataset.

To summarize, 38% of the top ten events for all five river catchments are not well represented by the model and bias correction does not have an added value. This agrees with findings from Bennett et al. (2014) and others, who also found that even bias-corrected RCM simulations tend to underestimate the magnitude of (multi-day) heavy precipitation events. In 22% (14%) of the events, precipitation is already (very) well represented in ERAI-CCLM, even without bias correction, showing that the model is able to simulate extreme precipitation. In 40% of the cases, bias correction leads to improvements in the representation of precipitation. 18% of the events are even very well captured after bias correction. Similar results can be found for ERA20C-CCLM (cf. Table S5).

6 Case studies

In this section, we present the results of two heavy precipitation events (associated with severe flooding) for three Central European river catchments, namely the Rhine, the Danube, and the Vistula (Fig. 1). The considered modeled daily precipitation

sums were corrected with EQM based on monthly distributions, as this method turned out to be most promising (see Sect. 4). We show a case where precipitation is already well presented by the model, so that bias correction cannot lead to further improvements (Rhine 1995). In addition, we show a second case (June 2009) where bias correction has a clear added value for precipitation in the Upper Danube river catchment, while at the same time precipitation is still strongly underestimated even 5 with bias correction for the Vistula river catchment.

6.1 Rhine: Flood of January 1995

In early January, snow- and rainfall as well as meltwater coming from the Alps led to a saturation of the soil in the Rhine catchment (Chbab, 1995). These preconditions favored the flooding in the end of January in the catchment area, which was mainly caused by record-breaking rainfall events within the last ten days of January (Fink et al., 1996). Four fatalities and an 10 economic loss of 500 million German Marks were reported for Germany (Fink et al., 1996).

In winter 1994/1995, a peak in the 7-day running mean of the daily precipitation sum averaged over the river catchment is visible on 25 January (Fig. 5 a). In E-OBS, the peak reaches nearly 13 mm, which is also exceeded by both downscaled reanalysis. Overall, the timing and intensity of the heavy precipitation events are quite similar in all datasets. Therefore, there is only a small added value of the bias correction in this case. This can also be seen in the accumulated catchment-averaged 15 precipitation (Fig. 5 b). In early February, the precipitation sum is slightly overestimated by the (bias-corrected) CCLM runs.

The high agreement between E-OBS and the bias-corrected CCLM runs can also be found in the horizontal distribution of the 7-day precipitation sum (including 22 to 28 January, Fig. 6). Outside the Rhine catchment, ERAI-CCLM and ERA20C-CCLM slightly overestimate precipitation over France (Fig. 6 b-c). Within the catchment, ERAI-CCLM shows less precipitation in the Alps (Fig. 6 b), compared to E-OBS and ERA20C-CCLM. Observed maxima in precipitation around 50° N and between 5° E 20 and 10° E are captured by both CCLM runs.

In summary, the Rhine winter flood in 1995 is well represented by the model, since precipitation intensity, timing, and the horizontal rainfall distribution agree with observations. Due to this close resemblance between model outputs and observations, the effect of the bias correction is of minor importance here.

6.2 Danube and Vistula: Floods of June 2009

25 In the last eight days of June 2009, a spatially extended stationary cut-off system located over Italy led to heavy precipitation events associated with flooding in the Danube and Vistula catchments (Godina and Müller, 2009). For example, in Mostviertel, Austria, 340 mm were reported between 22 and 29 June (Godina and Müller, 2009).

In the 7-day running mean of precipitation between May and July, there is a peak reaching 13 mm on 22 June in the Danube catchment and 11 mm in the Vistula catchment (Fig. 7 a and c). For the Danube, ERAI-CCLM produces more realistic pre- 30 cipitation amounts than ERA20C-CCLM. On 22 June itself, the bias correction leads to a clear improvement compared to the modeled precipitation (Fig. 7 a). The peak is just slightly overestimated in bias-corrected ERAI-CCLM. Before and after 22 June, the bias correction leads to the overestimation in the precipitation intensity compared to observations. In the accumulated precipitation sums, there is an overestimation in the first half and at the end of June with bias correction. However, there is
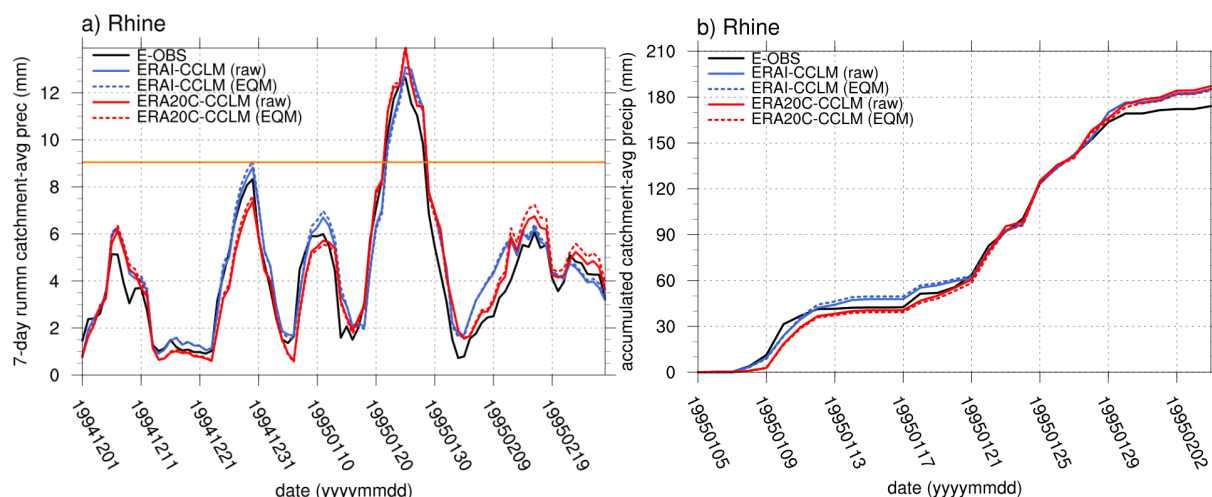
Figure 5. Time series of a) the 7-day running mean and b) the accumulated sum of the daily precipitation sum (in mm), both averaged over the Rhine catchment. The orange line marks the 99th percentile of the daily precipitation sum from E-OBS based on the time period between 1979 and 2010. Note that different time periods are presented in a) and b) (x-axis).
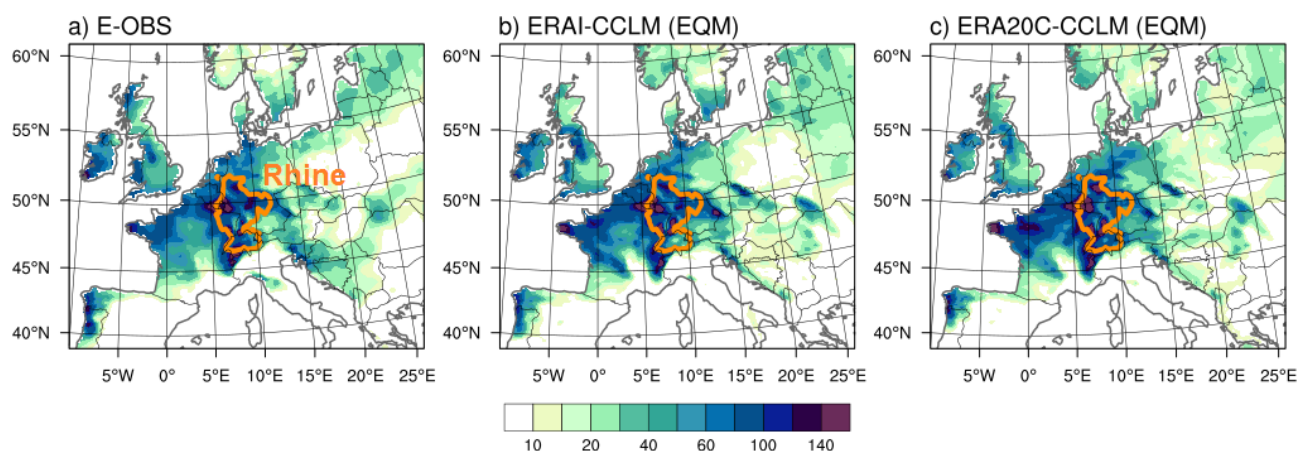
Figure 6. 7-day precipitation sum (in mm) for 22 to 28 January 1995 in a) E-OBS, b) ERAI-CCLM, and c) ERA20C-CCLM. Orange lines mark the relevant river catchment(s).

a clear benefit compared to the raw model run output (Fig. 7 b). For the Vistula, the maximum in precipitation is not well captured by the model. Furthermore, bias correction cannot improve this strong underestimation, as the raw and the bias-corrected precipitation sums lie close together (Fig. 7 c). Nevertheless, ERA20C-CCLM does a better performance than ERAI-CCLM. In the accumulated precipitation sums, the underrepresentation of precipitation in the model runs towards observations is obvious

5   for the Vistula catchment, even with bias correction (Fig. 7 d). Once again, it can be seen that the underestimation is more dramatic in ERAI-CCLM than in ERA20C-CCLM.
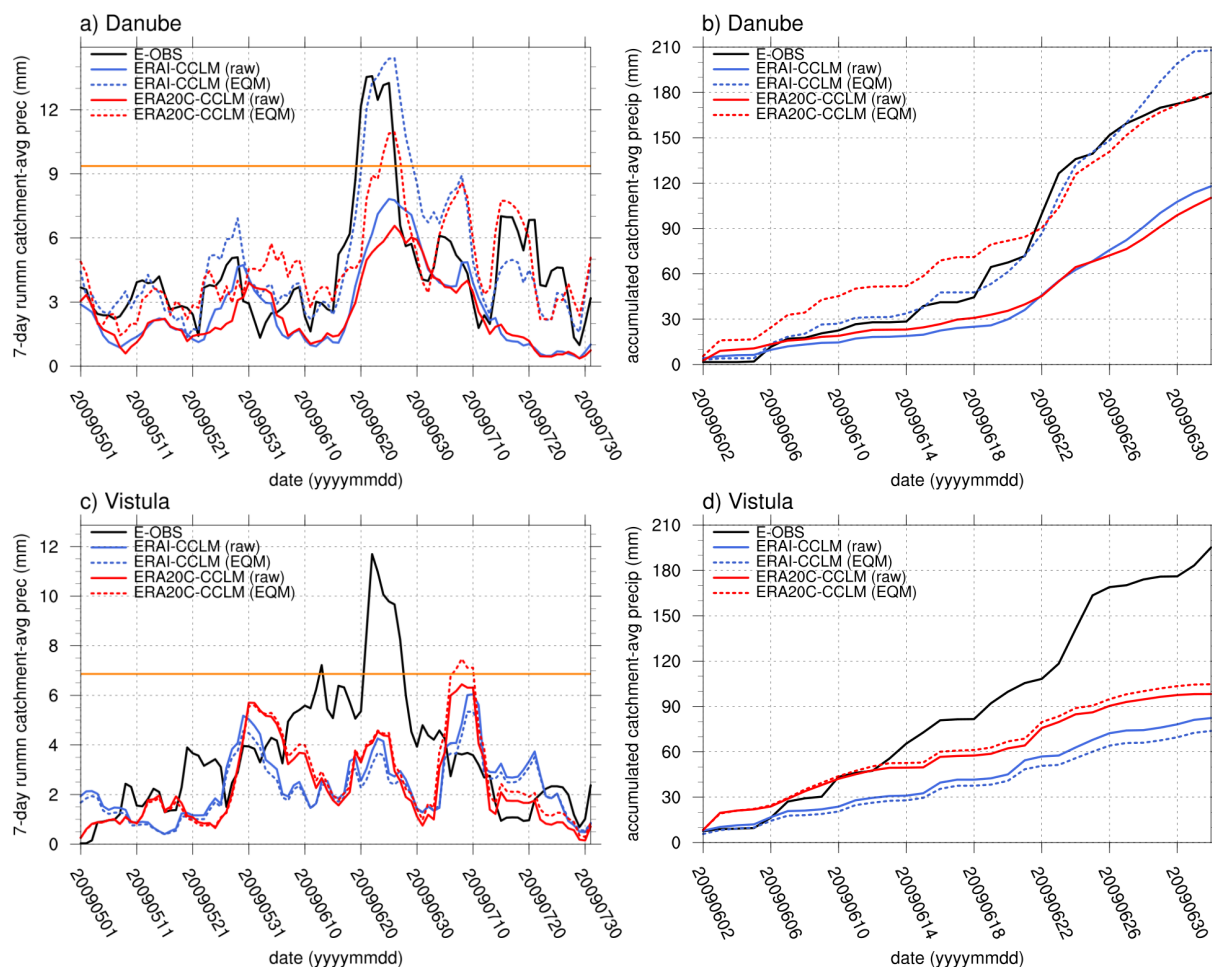
Figure 7. Similar to Fig. 5, but a) and b) show time series for the Danube and c) and d) for the Vistula.

Between 19 and 25 June, most precipitation is observed in the Danube catchment and south of it (Fig. 8 a). In the Vistula catchment, an area of heavy precipitation is only localized around 24° E and 53° N. We speculate that precipitation in surrounding areas of this peak within the Vistula catchment are underestimated and the result from observational gaps. In ERAI-CCLM, precipitation is well captured within the Danube catchment, while the horizontal distribution of precipitation within the Vistula

5 catchment differs strongly from E-OBS (Fig. 8 b). Moreover, precipitation is overestimated along coastal areas of Croatia and Italy. This is also true for ERA20C-CCLM (Fig. 8 c). However, rainfall within the Danube catchment is slightly underestimated.

In conclusion, the model runs underestimate precipitation within the Danube and the Vistula catchments. The bias correction increases the amount of precipitation for the Danube, but an underestimation still remains in the Vistula catchment.
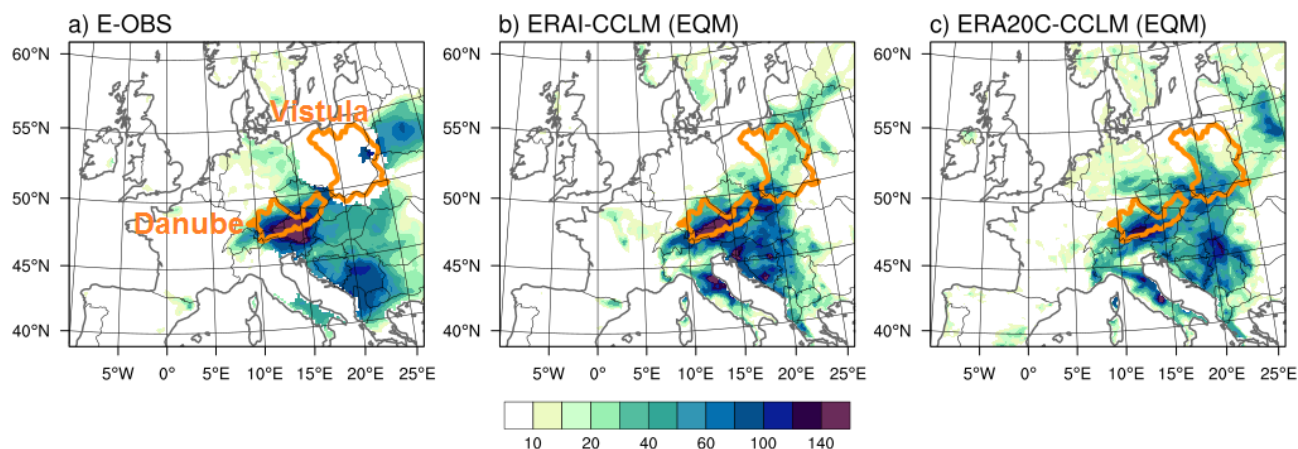
Figure 8. Similar to Fig. 6, but for 19 to 25 June 2009.

7 Discussion

Precipitation from reanalysis was dynamically downscaled with an RCM (CCLM) to generate a precipitation dataset which is suitable as input for hydrological rainfall-runoff models. We applied a statistical bias correction method (monthly EQM), in which E-OBS was used as reference data.

5   ERA-Interim and ERA-20C reanalysis were used as initial and boundary conditions for the RCM and both have advantages and disadvantages. ERA-Interim is known to produce unphysical changes in the global mean precipitation (Dee et al., 2011). Additionally, the accuracy of reanalysis data depends on the availability and quality of observations (Dee et al., 2011). ERA-20C tends to overestimate the wet-day frequency and to underestimate extreme precipitation values (Kim et al., 2018). Therefore, the choice of reanalysis may be a limiting factor in this approach.

10   Five commonly used bias correction methods have been tested, where EQM comes forth as most suitable to correct the bias in daily precipitation sums, also found in several previous studies before (e.g., Themeßl et al., 2011; Bennett et al., 2014). In addition, three calibration methods (monthly, seasonal, and half-yearly) were compared. Bias correction with monthly scaling factors and distributions lead to better results than the seasonal or the half-yearly calibrations. Haerter et al. (2011) have already shown that the improvements from bias correction depend on the considered timescales of the bias correction function. They
15   concluded that this dependency results from a different statistical behavior on different timescales. Although, we decided to use EQM, the other methods (LIN, LOCI, PT, and GQM) led also to improvements compared to the raw model runs, which agrees with findings from Teutschbein and Seibert (2012). Therefore, statistical measures, such as RMSE, R, or S, varied only slightly between the methods (see Fig. 2). Other approaches for bias correction can be found in e.g., Piani and Haerter (2012) or Moghim and Bras (2017). In Piani and Haerter (2012), copula coupling is presented, which means that precipitation and
20   temperature are corrected by fitting the two-dimensional distribution of the model run to that of observations. The representation of spring foods could benefit from this approach, as temperature is more important here in respect of snow melt. Moghim

and Bras (2017) used artificial neural networks (ANN) and showed that its output is suitable as input for hydrological models. Although ANNs are a powerful tool, a large effort is necessary to optimize an ANN for a specific application.

To show the benefit of bias correction, we investigated the top ten precipitation events of each river catchment for the period 1979–2010. As we used only precipitation sums to identify extreme events, the top ten cases do not include events in late
5   winter or early spring that are mainly caused by a strong snow melt. In most of the analyzed cases, bias correction led to clear improvements towards the raw model output. In some cases, the CCLM simulations were already close to observations, so that bias correction did not lead to further improvements, indicating the capability of the model to simulate extreme precipitation. In addition, we found a clear gradient from east to west. The underestimation of extreme precipitation amounts in the Vistula catchment could not be improved by the bias correction in most of the cases. We think that this underrepresentation of heavy
10   precipitation may be associated with observational gaps and a low density of stations within the Vistula catchment and surrounding areas (Haylock et al., 2008; Kyselỳ and Plavcová, 2010). Two case studies support this view.

The bias correction has an added value, as over- and underestimated precipitation amounts can be corrected in such a way that they are closer to observations. However, bias correction cannot improve precipitation events in respect of timing, or fix errors in the underlying dynamical processes (Haerter et al., 2011; Maraun, 2016). Dynamics may be improved by re-run the
15   model on a higher resolution (e.g., 2.8 km), where convection is resolved and orography is better depicted (e.g., Mass et al., 2002; Lean et al., 2008). Regarding convection, especially the simulations of summer floods may benefit from a higher model resolution, since embedded convection can enhance precipitation in between large-scale stratiform areas (e.g., Kirshbaum and Durran, 2004). Another source of uncertainty is the choice of reference data (here E-OBS). Although, E-OBS showed clear improvements during the last decade (e.g., for Poland), it inevitably has uncertainties. These uncertainties are caused by errors
20   and inhomogeneties in the station data as well as by (rounding) errors from the interpolation process (Hofstra et al., 2009). However, the suitability of the E-OBS dataset as reference data was shown by the high correlation coefficients between E-OBS and HYRAS daily precipitation sums (for three of the five considered river catchments, namely Elbe, Upper Danube and Rhine).

In this study, only precipitation was discussed, however, we calculated also datasets for the daily 2 m temperature, as this
25   variable is also needed as input for rainfall-runoff models. Due to its relation to snow melt, temperature plays an important role for late winter and early spring floods (e.g., Marks et al., 1998; Parajka et al., 2010). Regarding bias correction, additive linear scaling, variance scaling and quantile mapping were tested. First results indicate that the quantile mapping with a Gaussian distribution is most promising.

30   **8   Summary and Conclusions**

This study is a proof of concept on how to create a precipitation dataset which can be used as input to run hydrological models. We applied a dynamical downscaling approach with the regional climate model COSMO-CLM. ERA-Interim (∼80 km resolution) and ERA-20C (∼125 km resolution) were used as initial and boundary conditions. Then, the modeled precipitation

(∼25 km resolution) was dry-day and bias-corrected towards the observational dataset E-OBS.

The following conclusions can be made:

- *Which bias correction method is most suitable to correct model precipitation?*

  Empirical quantile mapping (EQM) based on monthly distributions turned out as one of the most suitable methods to correct model bias. However, others methods (as e.g., power transformation) showed also a clear improvement towards the raw model data. Statistical measures (e.g., the skill score of Taylor, 2001) varied only slightly between the methods.

- *Are historical precipitation events related to floods represented adequately in the dataset?*

  A comparison of all top ten flood events between 1970 and 2010 for the five river catchments showed that bias correction had an added value in most of the cases. However, for the Vistula catchment, bias correction did not lead to general improvements, which may result from errors or missing data in the underlying observational dataset E-OBS. The Rhine flood in winter 1995 was represented well in the (bias-corrected) datasets (ERAI-CCLM as well as ERA20C-CCLM). For the summer floods in 2009, the representation of precipitation was clearly improved with bias correction in the Upper Danube catchment, while the amount of precipitation was strongly underestimated in the Vistula catchment, even after bias correction.

- *What are the limitations of the approach?*

  Bias correction can improve the intensity of model precipitation towards observations, however, it is not possible to correct model precipitation in respect of timing or underlying dynamical processes. In addition, the results depend on the chosen reference data, here E-OBS. Moreover, the results are sensitive to the choice of reanalysis (here ERA-Interim and ERA-20C) used as boundary and initial conditions for the RCM.

In this pilot study, we introduced a method to generate centennial precipitation datasets which are suitable to be used as input for hydrological rainfall-runoff models. With these models, discharges can be calculated to develop probabilistic catastrophe models, with the overall goal to quantify the flooding risk for Central Europe from large ensemble datasets (e.g., Feldmann et al., submitted). Another possible pathway for the future is to investigate the decadal variability and trends of precipitation in centennial datasets such as ERA-20C and NCEP-20CR.

Appendix A: Bias correction methods

Details of the following bias correction methods can be found in, for example, Berg et al. (2012), Teutschbein and Seibert (2012), Lafon et al. (2013), Fang et al. (2015), Dobor et al. (2015), or Switanek et al. (2017).

A1 Linear scaling

Linear scaling (LS) is the simplest form of bias correction. For precipitation, a multiplicative approach is typically chosen:

$$P_{\mathrm{corr},m,d} = P_{\mathrm{raw},m,d} \cdot \frac{\mu(P_{\mathrm{obs},m})}{\mu(P_{\mathrm{raw},m})} \tag{A1}$$

$P_{\mathrm{corr},m,d}$ (m addressing the month and d the specific day) is the bias corrected precipitation, $P_{\mathrm{raw},m,d}$ the uncorrected model precipitation and $\mu(.)$ denotes the long-term monthly means of observed ($P_{\mathrm{obs},m}$) and modeled precipitation ($P_{\mathrm{raw},m}$).

A2 Local intensity scaling

Another approach is local intensity scaling (LOCI), where wet-day frequencies and intensities are corrected. Precipitation is
5   corrected as follows:

$$
\begin{aligned}
P_{\mathrm{corr},m,d} &= 0 & P_{\mathrm{raw},m,d} &< P_{\mathrm{thres},m} \\
P_{\mathrm{corr},m,d} &= P_{\mathrm{raw},m,d} \cdot s_m & &\text{otherwise}
\end{aligned}
\tag{A2}
$$

$P_{\mathrm{thres},m}$ is a wet-day threshold for the mth month from the model precipitation to ensure that the wet-day frequency of the observation is matched. s_m is a scaling factor to ensure that the mean of the corrected precipitation is equal to that of the observed precipitation. It can be calculated with:

$$
s_m = \frac{\mu(P_{\mathrm{obs},m,d}(P_{\mathrm{obs},m,d} > 0))}{\mu(P_{\mathrm{raw},m,d}(P_{\mathrm{raw},m,d} > P_{\mathrm{thres},m}))}
\tag{A3}
$$

In contrast to the applied dry-day correction, which aims for reducing model artifacts, LOCI is used to correct model offset compared to observations.

A3 Power transformation

Power transformation (PT) is a nonlinear method that can be used to correct the bias in time series of precipitation. Precipitation
15   is corrected as follows:

$$
P_{\mathrm{corr},m,d} = s_m \cdot P_{\mathrm{LOCI},m,d}^{b_m}
\tag{A4}
$$

$P_{\mathrm{LOCI},m,d}^{b_m}$ is the corrected precipitation from LOCI. b_m is estimated by minimizing the following function:

$$
f(b_m) = \frac{\sigma(P_{\mathrm{obs},m})}{\mu(P_{\mathrm{obs},n})} - \frac{\sigma(P_{\mathrm{LOCI},m}^{b_m})}{\mu(P_{LOCI,m}^{b_m})}
\tag{A5}
$$

where $\sigma(.)$ is the standard deviation. The scaling factor s_m is calculated by dividing the mean of observed precipitation by the
20   mean of $P_{\mathrm{LOCI},m,d}^{b_m}$:

$$
s_m = \frac{\mu(P_{\mathrm{obs},m})}{\mu(P_{\mathrm{LOCI},m}^{b_m})}
\tag{A6}
$$

A4 Quantile mapping

Quantile mapping (QM) is a distribution-driven bias correction method. The corrected precipitation can be determined as follows:

25
$$
P_{\mathrm{corr},m,d} = ecdf_{\mathrm{obs},m}^{-1}(ecdf_{\mathrm{raw},m}(P_{\mathrm{raw},m,d}))
\tag{A7}
$$

where $ecdf$ is the empirical cumulative distribution function (cdf) and $ecdf^{-1}$ its inverse. Instead of an empirical estimated cdf, quantile mapping can be applied using a given density function just like a gamma distribution for precipitation. In this case the corrected precipitation is calculated with:

$$P_{\mathrm{corr},m,d} = F_{\mathrm{obs},m}^{-1}(F_{\mathrm{raw},m}(P_{\mathrm{raw},m,d})) \tag{A8}$$

5   where F is the cdf of the gamma distribution.

Appendix B: Statistical quantities

B1   Skill score

The skill score of Taylor (2001) is defined as follows:

$$S = \frac{4 \cdot (1+R)}{\left(\sigma + \frac{1}{\sigma}\right)^2 (1+R_0)} \tag{B1}$$

10   where σ is defined as the standard deviation of the model data divided through the standard deviation of the observations. R is Pearson's correlation coefficient and R_0 the maximum attainable correlation coefficient. R_0 is defined as maximum of all coefficients from the different datasets (here, the model runs with different bias correction) plus 10%, following Ehmele and Kunz (2018). S ranges from 0 to 1, where values close to 1 indicate a high resemblance between the model and the observational data.

15   **B2   Centered root mean square difference**

The centered root mean square difference (cRMSD) can be calculated as:

$$cRMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [(x_i - \overline{x}) \cdot (y_i - \overline{y})]^2} \tag{B2}$$

where N is the number of data points, and x_i and y_i specific data points of two datasets. The overline denotes an average of all data points.

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

Acknowledgements.   We thank Aon Central and Eastern Europe a.s. for funding. We thank the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the ERA-Interim and ERA-20C data. We acknowledge the 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). In addition, we thank the German Climate Computing Centre (DKRZ, Hamburg) for computing and storage resources. JPG thanks the AXA Research fund
25   for support.

References

Alfieri, L., Dottori, F., Betts, R., Salamon, P., and Feyen, L.: Multi-model projections of river flood risk in Europe under global warming, Climate, 6, 6, https://doi.org/10.3390/cli6010006, 2018.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities, Mon. Wea. Rev., 139, 3887–3905, https://doi.org/10.1175/MWR-D-10-05013.1, 2011.

Bennett, J. C., Grose, M. R., Corney, S. P., White, C. J., Holz, G. K., Katzfey, J. J., Post, D. A., and Bindoff, N. L.: Performance of an empirical bias-correction of a high-resolution climate dataset, Int. J. Climatol., 34, 2189–2204, https://doi.org/10.1002/joc.3830, 2014.

Berg, P., Feldmann, H., and Panitz, H.-J.: Bias correction of high resolution regional climate model data, J. Hydrology, 448, 80–92, https://doi.org/10.1016/j.jhydrol.2012.04.026, 2012.

Chbab, E.: How extreme were the 1995 flood waves on the rivers Rhine and Meuse?, Phys. Chemistry Earth, 20, 455–458, https://doi.org/10.1016/S0079-1946(96)00005-5, 1995.

Dee, D. P., Uppala, S. M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quart. J. Roy. Meteor. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dobler, A. and Ahrens, B.: Precipitation by a regional climate model and bias correction in Europe and South Asia, Meteor. Z., 17, 499–509, https://doi.org/10.1127/0941-2948/2008/0306, 2008.

Dobor, L., Barcza, Z., Hlásny, T., Havasi, Á., Horváth, F., Ittzés, P., and Bartholy, J.: Bridging the gap between climate models and impact studies: the FORESEE Database, Geosci. data j., 2, 1–11, https://doi.org/10.1002/gdj3.22, 2015.

Doms, G., Förstner, J., Heise, E., Herzog, H.-J., Raschendorfer, M., Schrodin, R., Reinhardt, T., and Vogel, G.: A description of the nonhydrostatic regional model LM. Part II: Physical parameterization, Tech. Rep., Consortium for Small-Scale Modelling, p. 161pp, http://www.cosmo-model.org, 2011.

Ehmele, F. and Kunz, M.: Flood-Related Extreme Precipitation in Southwestern Germany: Development of a Two-Dimensional Stochastic Precipitation Model, Hydrol. Earth Syst. Sci., https://doi.org/10.5194/hess-2018-147, 2018.

Fang, G., Yang, J., Chen, Y., and Zammit, C.: Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in China, Hydrology Earth Syst. Sci., 19, 2547–2559, https://doi.org/10.5194/hess-19-2547-2015, 2015.

Feldmann, H., Frueh, B., Schaedler, G., Panitz, H.-J., Keuler, K., Jacob, D., and Lorenz, P.: Evaluation of the precipitation for South-western Germany from high resolution simulations with regional climate models, Meteor. Z., 17, 455–465, https://doi.org/10.1127/0941-2948/2008/0295, 2008.

Feldmann, H., Pinto, J. G., Laube, N., Uhlig, M., Moemken, J., Pasternack, A., Früh, B., Pohlmann, H., and Kottmeier, C.: Characterization of Skill and Added Value of the MiKlip Regional Decadal Prediction System for Temperature over Europe, Tellus A, submitted.

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. Amer. Meteor. Soc., 92, 1181–1192, https://doi.org/10.1175/2011BAMS3061.1, 2011.

Feyen, L., Dankers, R., Bódis, K., Salamon, P., and Barredo, J. I.: Fluvial flood risk in Europe in present and future climates, Climatic Change, 112, 47–62, https://doi.org/10.1007/s10584-011-0339-7, 2012.

Fink, A., Ulbrich, U., and Engel, H.: Aspects of the January 1995 flood in Germany, Weather, 51, 34–39, https://doi.org/10.1002/j.1477-8696.1996.tb06182.x, 1996.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, Int. J. Climatol., 27, 1547–1578, https://doi.org/10.1002/joc.1556, 2007.

Frei, C., Davies, H. C., Gurtz, J., and Schär, C.: Climate dynamics and extreme precipitation and flood events in Central Europe, Integrated Assessment, 1, 281–300, https://doi.org/10.1023/A:1018983226334, 2000.

Godina, R. and Müller, G.: Das Hochwasser in Österreich vom 22. bis 30. Juni 2009, Beschreibung der hydrologischen Situation Abteilung VII/3–Wasserhaushalt (HZB), Bundesministerium für Land-und Forstwirtschaft, Umwelt und Wasserwirtschaft, Vienna, 2009.

Gutmann, E. D., Rasmussen, R. M., Liu, C., Ikeda, K., Gochis, D. J., Clark, M. P., Dudhia, J., and Thompson, G.: A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain, J. Climate, 25, 262–281, https://doi.org/10.1175/2011JCLI4109.1, 2012.

Haerter, J., Hagemann, S., Moseley, C., and Piani, C.: Climate model bias correction and the role of timescales, Hydrology Earth Syst. Sci. 15, 1065–1073, https://doi.org/10.5194/hess-15-1065-2011, 2011.

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

Hilker, N., Badoux, A., and Hegg, C.: The Swiss flood and landslide damage database 1972-2007, Nat. Hazards Earth Syst. Sci., 9, 913–925, https://doi.org/10.5194/nhess-9-913-2009, 2009, 2009.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanae, S.: Global flood risk under climate change, Nature Climate Change, 3, 816, https://doi.org/10.1038/nclimate1911, 2013.

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.: Atmospheres, 114, https://doi.org/10.1029/2009JD011799, 2009.

Huebener, H., Stoffels, B., and Brahmer, G.: Assessment and projection of high and low precipitation extremes simulated by COSMO-CLM4.8 for four small river catchments in Hesse, Germany, Meteor. Z., pp. 563–575, https://doi.org/10.1127/metz/2016/0708, 2016.

Jongman, B., Hochrainer-Stigler, S., Feyen, L., Aerts, J. C., Mechler, R., Botzen, W. W., Bouwer, L. M., Pflug, G., Rojas, R., and Ward, P. J.: Increasing stress on disaster-risk finance due to large floods, Nature Climate Change, 4, 264, https://doi.org/10.1038/nclimate2124, 2014.

Kelemen, F. D., Ludwig, P., Reyers, M., Ulbrich, S., and Pinto, J. G.: Evaluation of moisture sources for the Central European summer flood of May/June 2013 based on regional climate model simulations, Tellus, 68, 29 288, https://doi.org/10.3402/tellusa.v68.29288, 2016.

Kim, D.-I., Kwon, H.-H., and Han, D.: Exploring the Long-Term Reanalysis of Precipitation and the Contribution of Bias Correction to the Reduction of Uncertainty over South Korea: A Composite Gamma-Pareto Distribution Approach to the Bias Correction, Hydrol. Earth Syst. Sci., https://doi.org/10.5194/hess-2018-36, 2018.

Kirshbaum, D. J. and Durran, D. R.: Factors governing cellular convection in orographic precipitation, J. Atmos. Sci., 61, 682–698, https://doi.org/10.1175/1520-0469(2004)061<0682:FGCCIO>2.0.CO;2, 2004.

Kunz, M.: Characteristics of large-scale orographic precipitation in a linear perspective, J. Hydrometeor., 12, 27–44, https://doi.org/10.1175/2010JHM1231.1, 2011.

Kyselỳ, J. and Plavcová, E.: A critical remark on the applicability of E-OBS European gridded temperature data set for validating control climate simulations, J. Geophys. Res., 115, https://doi.org/10.1029/2010JD014123, 2010.

Lafon, T., Dadson, S., Buys, G., and Prudhomme, C.: Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods, Int. J. Climatol., 33, 1367–1381, https://doi.org/10.1002/joc.3518, 2013.

Lean, H. W., Clark, P. A., Dixon, M., Roberts, N. M., Fitch, A., Forbes, R., and Halliwell, C.: Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom, Mon. Wea. Rev., 136, 3408–3424, https://doi.org/10.1175/2008MWR2332.1, 2008.

Maddox, R. A., Chappell, C. F., and Hoxit, L. R.: Synoptic and meso-α scale aspects of flash flood events, Bull. Amer. Meteor. Soc., pp. 115–123, 1979.

Maraun, D.: Bias correcting climate change simulations-a critical review, Current Climate Change Reports, 2, 211–220, https://doi.org/10.1007/s40641-016-0050-x, 2016.

Maraun, D., Wetterhall, F., Ireson, A., Chandler, R., Kendon, E., Widmann, M., Brienen, S., Rust, H., Sauter, T., Themeßl, M., et al.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, https://doi.org/10.1029/2009RG000314, 2010.

Marks, D., Kimball, J., Tingey, D., and Link, T.: The sensitivity of snowmelt processes to climate conditions and forest cover during rain-on-snow: A case study of the 1996 Pacific Northwest flood, Hydrological Processes, 12, 1569–1587, https://doi.org/10.1002/(SICI)1099-1085(199808/09)12:10/11<1569::AID-HYP682>3.0.CO;2-L, 1998.

Marotzke, J., Müller, W. A., Vamborg, F. S., Becker, P., Cubasch, U., Feldmann, H., Kaspar, F., Kottmeier, C., Marini, C., Polkova, I., et al.: MiKlip: a national research project on decadal climate prediction, Bull. Amer. Meteor. Soc., 97, 2379–2394, https://doi.org/10.1175/BAMS-D-15-00184.1, 2016.

Mass, C. F., Ovens, D., Westrick, K., and Colle, B. A.: Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific Northwest, Bull. Amer. Meteor. Soc., 83, 407–430, https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2, 2002.

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, https://doi.org/10.1051/lhb/2014001, 2014.

Moghim, S. and Bras, R. L.: Bias correction of climate modeled temperature and precipitation using artificial neural networks, J. Hydrometeor., 18, 1867–1884, https://doi.org/10.1175/JHM-D-16-0247.1, 2017.

Parajka, J., Kohnová, S., Bálint, G., Barbuc, M., Borga, M., Claps, P., Cheval, S., Dumitrescu, A., Gaume, E., Hlavčová, K., et al.: Seasonal characteristics of flood regimes across the Alpine–Carpathian range, J. hydrology, 394, 78–89, https://doi.org/10.1016/j.jhydrol.2010.05.015, 2010.

Pauling, A. and Paeth, H.: On the variability of return periods of European winter precipitation extremes over the last three centuries, Climate of the Past, 3, 65–76, https://doi.org/10.5194/cp-3-65-2007, 2007.

Piani, C. and Haerter, J.: Two dimensional bias correction of temperature and precipitation copulas in climate models, Geophys. Res. Lett., 39, https://doi.org/10.1029/2012GL053839, 2012.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G., Peubey, C., Thépaut, J.-N., et al.: 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.

Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., and Gratzki, A.: A Central European precipitation climatology–Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS), Meteor. Z., 22, 235–256, https://doi.org/10.1127/0941-2948/2013/0436, 2013.

Reyers, M., Pinto, J. G., and Moemken, J.: Statistical–dynamical downscaling for wind energy potentials: evaluation and applications to decadal hindcasts and climate change projections, Int. J. Climatol., 35, 229–244, https://doi.org/10.1002/joc.3975, 2015.

Richardson, C. W.: Stochastic simulation of daily precipitation, temperature, and solar radiation, Water resources res., 17, 182–190, https://doi.org/10.1029/WR017i001p00182, 1981.

Ritter, B. and Geleyn, J.-F.: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations, Mon. Wea. Rev., 120, 303–325, https://doi.org/10.1175/1520-0493(1992)120<0303:ACRSFN>2.0.CO;2, 1992.

5 Rockel, B., Will, A., and Hense, A.: The regional climate model COSMO-CLM (CCLM), Meteor. Z., 17, 347–348, https://doi.org/10.1127/0941-2948/2008/0309, 2008.

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, 68, 31 820, https://doi.org/10.3402/tellusa.v68.31820, 2016.

10 Stucki, P., Bandhauer, M., Heikkilä, U., Rössler, O., Zappa, M., Pfister, L., Salvisberg, M., Froidevaux, P., Martius, O., Panziera, L., et al.: Reconstruction and simulation of an extreme flood event in the Lago Maggiore catchment in 1868, Nat. hazards earth syst. sci. discussions, 2018, 1–28, https://doi.org/10.5194/nhess-2018-134, 2018.

Switanek, M. B., Troch, P. A., Castro, C. L., Leuprecht, A., Chang, H.-I., Mukherjee, R., and Demaria, E.: Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes, Hydrology Earth Syst. Sci., https://doi.org/10.5194/hess-21-2649-2017, 2017.

15 Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, https://doi.org/10.1029/2000JD900719, 2001.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, J. Hydrology, 456, 12–29, https://doi.org/10.1016/j.jhydrol.2012.05.052, 2012.

20 Themeßl, M. J., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, Int. J. Climatol., 31, 1530–1544, https://doi.org/10.1002/joc.2168, 2011.

Tiedtke, M.: A comprehensive mass flux scheme for cumulus parameterization in large-scale models, Mon. Wea. Rev., 117, 1779–1800, https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2, 1989.

Trzaska, S. and Schnarr, E.: A review of downscaling methods for climate change projections, United States Agency for International Development by Tetra Tech ARD, pp. 1–42, 2014.

25 Van den Besselaar, E., Haylock, M., Van der Schrier, G., and Klein Tank, A.: A European daily high-resolution observational gridded data set of sea level pressure, J. Geophys. Res., 116, https://doi.org/10.1029/2010JD015468, 2011.

Ward, P. J., Moel, H. d., Aerts, J., et al.: How are flood risk estimates affected by the choice of return-periods?, Nat. Hazards Earth Syst. Sci., 11, 3181–3195, https://doi.org/10.5194/nhess-11-3181-2011, 2011.