Event selection for dynamical downscaling: a neural network approach for physically-constrained precipitation events

J. J. Gómez-Navarro[1] · C. C. Raible[2] · J. A. García-Valero[3] · M. Messmer[2,4] · J. P. Montávez[1] · O. Martius[5]

Abstract

This study presents a new dynamical downscaling strategy for extreme events. It is based on a combination of statistical downscaling of coarsely resolved global model simulations and dynamical downscaling of specific extreme events constrained by the statistical downscaling part. The method is applied to precipitation extremes over the upper Aare catchment, an area in Switzerland which is characterized by complex terrain. The statistical downscaling part consists of an Artificial Neural Network (ANN) framework trained in a reference period. Thereby, dynamically downscaled precipitation over the target area serve as predictands and large-scale variables, received from the global model simulation, as predictors. Applying the ANN to long term global simulations produces a precipitation series that acts as a surrogate of the dynamically downscaled precipitation for a longer climate period, and therefore are used in the selection of events. These events are then dynamically downscaled with a regional climate model to 2 km. The results show that this strategy is suitable to constraint extreme precipitation events, although some limitations remain, e.g., the method has lower efficiency in identifying extreme events in summer and the sensitivity of extreme events to climate change is underestimated.

1 Introduction

Extreme precipitation is a necessary precursor for flooding, which can cause high economic and human losses in densely populated areas. Extremely rare events are characterized by long return periods (Salvadori et al. 2011), and are used for risk assessments of critical infrastructure that requires special protection, such as nuclear power plants or dams (Requena et al. 2013). Moreover, extreme events may be affected by climate change, as pointed out by growing evidence that relates climate change with an intensification in the frequency and severity of extreme episodes (Seneviratne et al. 2012). However, an important challenge in the characterisation of the risks associated with these events is that they are, by definition, extremely rare. Given the relatively short instrumental records of rainfall, the characterisation of extremes whose return period exceeds centuries is affected by large uncertainties.

A prominent way to tackle this problem is through climate modelling. Large ensembles of simulations carried out with comprehensive Earth System models (ESM) (e.g. CMIP6, Eyring et al. 2016) provide a valuable source of information about the evolution of the hydrological cycle for the future. However, climate models only contain a simplified representation of precipitation processes, and one of

✉ J. J. Gómez-Navarro
  jjgomeznavarro@um.es

  C. C. Raible
  raible@climate.unibe.ch

  J. A. García-Valero
  garciavalero@aemet.es

  M. Messmer
  messmer@climate.unibe.ch

  J. P. Montávez
  montavez@um.es

  O. Martius
  olivia.romppainen@giub.unibe.ch

[1] Department of Physics, University of Murcia, Murcia, Spain

[2] Climate and Environmental Physics, Physics Institute and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

[3] Agencia Estatal de Meteorología (AEMET), Murcia, Spain

[4] School of Earth Sciences, The University of Melbourne, Melbourne, Victoria, Australia

[5] Department of Geography and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

the prominent drawbacks of state-of-the-art ESMs is their coarse spatial resolution, that limits their applicability in impact studies at local scales (e.g., Messmer et al. 2017; García-Valdecasas Ojeda et al. 2017; Felder et al. 2018). For this reason, the output of ESMs is brought to a higher spatial resolution using either statistical or dynamical downscaling methods (Maraun et al. 2010, 2015). For dynamical downscaling, Regional Climate Models (RCM) are run over a limited spatial domain at a higher spatial resolution (e.g. Torma et al. 2015; Fantini et al. 2016; Giorgi et al. 2016; Gómez-Navarro et al. 2018, among many others). However, the computational cost of such regional simulations is still considerably high. This has motivated the development of hybrid approaches that take advantage of statistical relationships to extend the results drawn from short RCM simulations over longer time periods. Martinez et al. (2012) developed a statistical–dynamical downscaling procedure that relies on Empirical Orthogonal Function analysis to generate large-scale atmospheric patterns, which are then dynamically downscaled. This allowed the researchers to construct regional time series, and was successfully used to generate realistic regional series of wind with 6-hour resolution. Li et al. (2011) used a limited number of existing RCM simulations over North America to fit a linear regression model between the RCM output and the driving ESM fields. This statistical model was then used, together with a large set of ESM simulations, to produce a probabilistic projection of high-resolution temperature change in North America, which even allowed the researchers to quantify the different sources of uncertainty.

Within the RCM community, large on-going initiatives such as The Coordinated Regional Climate Downscaling Experiment (CORDEX) have been formed to coordinate the computational effort, therefore facilitating and maximising the exchange of information derived from these costly simulations. Nowadays, the resolution in most RCM simulations is about 10 km (e.g., https://guidelines.euro-cordex.net). This resolution is sufficient to demonstrate the added value of RCM compared to ESM simulations, especially regarding precipitation processes (Torma et al. 2015; Fantini et al. 2016; Bowden et al. 2016). However, there exists added value of going beyond 10 km, entering the scale of convection permitting simulations (Ban et al. 2014; Giorgi et al. 2016; Zittis et al. 2017; Gómez-Navarro et al. 2018), in particular in areas of complex topography and during extreme precipitation events (Giorgi et al. 2016; Chan et al. 2017). Still, the high computational cost is the bottleneck that has limited the number of simulations currently available of this nature, and it may become an unavoidable limitation precluding the RCM community from taking full advantage of the new CMIP6 ensemble.

To overcome this limitation, some researchers have proposed to make a previous selection of dates to be simulated to avoid the computational cost of running transient climate simulations. Meredith et al. (2018) presented a classification algorithm based on geopotential height as a mean to select dates with an elevated potential for extreme precipitation in a narrow river catchment. This approach enables a clever selection of events to be dynamically downscaled that discards situations of fewer interest, and so high-resolution RCM simulations can be selectively performed saving important computational resources. Felder et al. (2018) aimed at simulating worst-case events using a range of computational models across spatial scales, from an ESM to a damage and loss model reaching the scale of individual buildings. To keep the physical consistency among models, the research team selected events within the ESM, using the extreme precipitation (averaged over Switzerland) as criterion and dynamically downscaled these events to 2 km. The analysis of the downscaled events showed that this criterion leads to unsatisfactory results, and suggested that any refinement of the approach shall include more variables from the ESM used as predictors. Chan et al. (2017) selected three large-scale predictors of extreme precipitation: Mean Sea Level Pressure, 850-hPa relative vorticity and static stability. They used regression analysis to identify large-scale precursors of extreme precipitation events in convection-permitting climate simulations, and found that indeed these three variables have skill in predicting precipitation extremes in simulations both at 12 and 1.5 km spatial resolution.

In general, good predictor variables should include the main processes contributing to heavy precipitation on a scale that is captured by the ESM. Heavy and extreme precipitation requires a steady supply of moisture and a lifting mechanism that brings the moist air to saturation (e.g., Doswell et al. 1996). In Switzerland moisture is transported towards the Alps from the south prior to and during regional-scale heavy precipitation events on the Alpine south side (Martius et al. 2006; Winschall et al. 2012). During regional-scale heavy precipitation on the Alpine north side, sustained and intensive moisture transport against the orography can occur from the east during Vb weather situations, from the north and from the west (Piaget et al. 2015; Giannakaki and Martius 2016; Froidevaux and Martius 2016; Messmer et al. 2017). The moist air masses reaching Switzerland are lifted within warm convey belts (Pfahl et al. 2014), along the orography (Giannakaki and Martius 2016), or in areas of flow convergence (Giannakaki and Martius 2016).

Here, we propose a comprehensive and flexible framework that blends statistical and dynamical downscaling and, similarly as the one presented by Meredith et al. (2018), it provides a suitable identification of candidates to be local extreme precipitation events in long ESM simulations. As tested for this method, we use it to forecast daily extreme precipitation in a region of complex orography, i.e. the catchment of the Aare river upstream of Bern (Switzerland).

2 Data and methods

2.1 Community Earth System Model (CESM)

The Community Earth System Model (CESM, 1.0.1 release; Hurrell et al. 2013) was developed at the National Center for Atmospheric Research. This ESM has been run with a horizontal resolution of about 1° (about 110 km in the equator) in all physical model components, i.e. atmosphere, ocean, land and sea ice (Gent et al. 2011). Further, the carbon cycle was explicitly simulated. The reader interested in the full details of this particular model configuration is referred to Lehner et al. (2015) for a comprehensive description.

We use data from two CESM simulations: (1) a 400-year simulation with perpetual AD 850 conditions, hereafter referred as CESM-control simulation, (2) a seamless 850–2099 AD simulation driven by reconstructions of external forcings for the historical period 850–2005 and RCP8.5 forcing for the future period 2006–2099 (Lehner et al. 2015). In this study, we use this data but consider it as split in two periods: 1850–2005, hereafter referred as CESM-historical simulation, and 2006–2099, hereafter CESM-future simulation.

2.2 Weather Research and Forecasting Model (WRF)

The dynamical downscaling of the CESM simulations is performed with the Weather Research and Forecasting model (WRF), version 3.5 (Skamarock et al. 2008). We use a setup with 4 nested domains reaching a spatial resolution of 2 km in its innermost domain spanning the Alpine region entirely (see Fig. 1). This high resolution allows us to explicitly simulate convective processes, which is of foremost importance in extreme event phenomena, precisely those that this study tackles (Ban et al. 2014; Giorgi et al. 2016). A comprehensive description of the details of these simulations, as well as an evaluation of the model performance in the particular configuration employed in this study is presented in Gómez-Navarro et al. (2018).

The simulations performed with WRF include (1) a transient present-climate simulation that continuously spans the period 1979–2005 and is driven by the CESM-historical simulation, hereafter referred as WRF-reference period; (2) a transient simulation over the period 2080–2099 nested to CESM-future simulation, hereafter WRF-future; (3) a number of single-day case studies selected from the CESM-control, CESM-historical and CESM-future simulations. The criteria for the selection of dates is described below, and in each case a short spinup of 12 h is used.
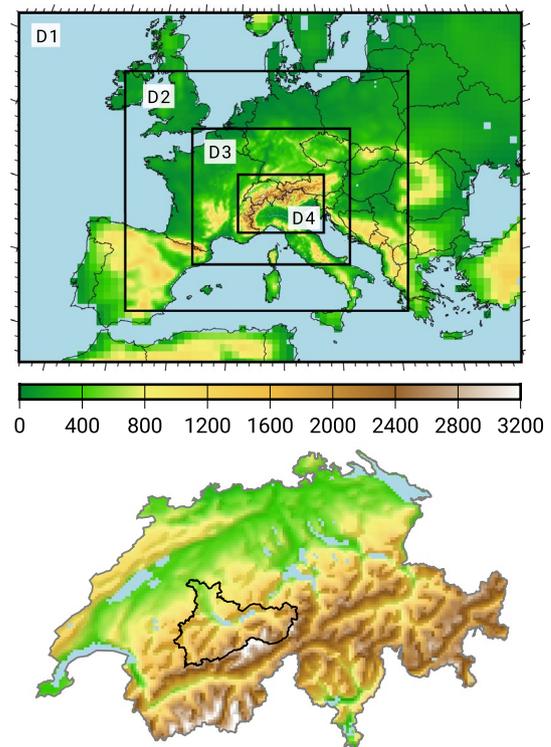

Fig. 1 Top: configuration of the four nested domains used to downscale CESM with WRF. Bottom: detail of the actual orography implemented in the 2-km resolution simulation over Switzerland. The black contour outlines the target of the study, the catchment of the Aare river upstream of Bern

2.3 Artificial neural networks as statistical downscaling tool

An ANN is a mathematical model that acts as a function, relating certain n-dimensional input vectors to m-dimensional output vectors (Schalkoff 1997). This model is not new in meteorological applications. Dawson and Wilby (1998) proposed a novel rainfall-runoff model based on ANNs, and used it to forecast the river flow in two different UK catchments with a skill comparable to operational systems. Lee et al. (1998) used an ANN to build a model to forecast precipitation in Switzerland, a region characterised by complex orography. ANNs have also found early applications as downscaling technique in a very similar manner as we aim here (Zorita and Storch 1999). The reader is referred to the former references for a more comprehensive explanation of the algorithm, as we just briefly outline its most important aspects in the following.

An ANN is composed of various layers, each of which contains so-called neurons, that can be regarded as computation units. A network contains at least an input layer (with n neurons, the dimension of the input vector) and an output layer (with m neurons, the dimension of the output vector).

In between, there can be a number of so-called hidden layers, each of which including a variable number of neurons. Each neuron is connected to all neurons in the following layers through connections that are characterised by a weight. Given an input, the network calculates a unique output that results from a relatively simple and therefore computationally inexpensive calculation that involves the input vector and all the weights, which act as the parameters of the model. Therefore, the calibration of the network consists of finding the set of weights that optimise the output of the network for a given metric. We use supervised learning, in which the training of the network consists of using a number of input–output couples from the reference period, i.e. the predictors and the predictand, to find the set of weights that minimise the difference between the transient RCM and the output of the ANN, averaged for the whole pool of samples that conform the reference period. The search of an optimal solution is computationally moderately demanding, and is based on the backward propagation algorithm, that can be viewed as an application of the chain rule in differentiation (Schalkoff 1997).

The performance of the ANN approach to select events is based on a number of skill metrics: correlation, Hit Rate and Symmetric Extremal Dependence Index (SEDI), comprehensively described in the Appendix. Note that the truth we aim at reproduce with the help of an ANNs is not the actual precipitation, but the one produced by the CESM-WRF system. Therefore, "observation" refers hereafter to the daily precipitation simulated by WRF averaged over the Aare catchment upstream of Bern, whereas "prediction" refers to the output of an ANN once it is trained during the reference period to mimic the CESM-RCM relationships over such region.

Finally, we use the ANN from the R package neuralnet (https://CRAN.R-project.org/package=neuralnet). The geometry of the network and the number of variables used as input predictors for the ANN are not part of the calibration, but have to be determined beforehand, according to a number of comprehensive tests described in Sect. 3.2.

2.4 Strategy for the selection of events

This work proposes a strategy for the selection of dates candidate to extreme event that consists of the following steps (Fig. 2):

1. The coarse fields from an ESM are dynamically downscaled with an RCM. This is a computationally demanding step, so this reference period is inadmissibly short for most climate applications.
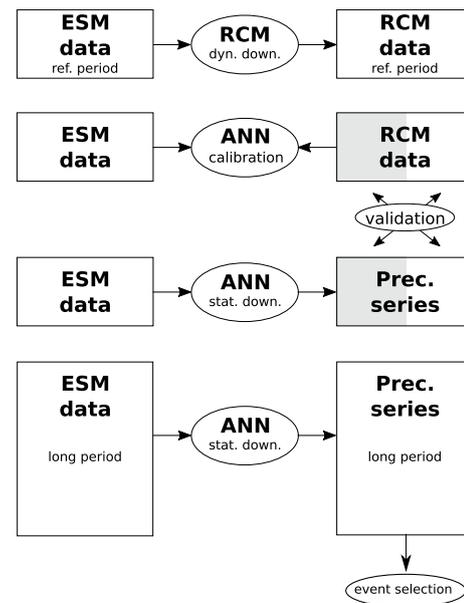2. An ANN is trained to learn the relationship between the large-scale variables in the ESM and the daily pre-

Fig. 2 Scheme of the different steps to obtain a tool for the selection of dates candidates to extreme events. First, a computationally extensive high-resolution simulation is carried out over the domain of interest during a reference period. Next, the dynamically downscaled dataset, together with the driving ESM, are used to calibrate an ANN. Only half of the reference period is used for the calibration, as the other half is reserved for validation of the ANN against the RCM data (this is symbolised with the white/grey shading). The role of the calibration and validation periods is exchanged to ensure a correct validation. Finally, the calibrated ANN is applied to obtain a statistically downscaled series over a longer period that serves for the selection of events

cipitation simulated by the RCM in an orographycally complex region.
3. The calibration and validation periods are swapped to carefully assess the performance of the ANN with independent data during the reference period.
4. Finally, the calibrated ANN can be used to statistically downscale a longer ESM simulation. This way, the ANN tries to emulate the series we would obtain running the RCM for the longer period in case it would be computationally feasible. This series shall be used in the selection of dates candidates to extreme events.

To demonstrate the feasibility of this approach, we implement it in the following sections. We use the WRF-reference simulation comprehensively described in Gómez-Navarro et al. (2018). The arguments for the selection of predictors are presented in Sect. 3.1, followed by a range of tests that allow to determine the optimal ANN geometry in Sect. 3.2. A comprehensive calibration using various statistics associated to the forecast of extreme events is presented in Sect. 3.3. Finally, in Sect. 4 we apply the obtained ANN during a perturbed climate period the ANN was not calibrated

for. This allows us to assess the limitations derived from the stationary hypothesis implicit in many statistical downscaling exercises.

3 Event selection based on ANN during the reference period

3.1 Large scale variables—potential predictors of extreme events

Similarly to Chan et al. (2017), we first gain insight on how local extreme events in precipitation over the Aare catchment (bottom panel in Fig. 1) are connected to the large scale atmospheric dynamics in the WRF-CESM coupled system. For this, we use a composite analysis applied to different variables using the 1-day extreme precipitation events, defined here as those days exceeding the 95th percentile of daily precipitation and applied to each season separately. All days in the WRF-reference simulation are filtered out according to this criterion and then averaged. The selection of the variables is based on previous studies (Martius et al. 2006; Winschall et al. 2012; Messmer et al. 2017): sea level pressure (SLP), geopotential height at 850 and 500 hPa (Z850 and Z500), integrated water vapour and vapour transport (IWV and IWVT, respectively) and precipitation (PREC). The annual cycle is removed from each variable to obtain anomalies. Note that one limitation of this analysis is that these composites reflect the mean large-scale flow patterns. Thus, for individual events the large-scale flow and the moisture transport can substantially deviate from these composites as presented by Giannakaki and Martius (2016) who identified several relevant flow patterns associated with heavy precipitation events in Northern Switzerland.

Composites for selected variables for winter are presented in Fig. 3. Extreme daily precipitation events are related to a west-east oriented pressure anomaly dipole with high pressure centred between the Iberian Peninsula and the Bay of Biscay and low pressure over eastern Europe (Fig. 3a). This pressure dipole has a barotropic vertical structure (Fig. 3b). The height thickness between 500- and 850-hPa, a measure of temperature anomalies in this layer, indicates a dipole with warmer air over western and colder air over eastern Europe (Fig. 3c). The strongest pressure gradient is located over Germany and Switzerland and suggests strong north to northwesterly winds over this region. Consistently extreme daily precipitation events in winter are associated with a strong north to northwesterly integrated water vapour transport (Fig. 3e). As expected, CESM generates a positive precipitation anomaly over Switzerland when sampling over extreme precipitation events in the WRF simulation (Fig. 3d). The autumn patterns (not shown) resemble the winter ones while the spring ones are rotated counter

clockwise by 10° with respect to the winter pattern for all but the precipitation pattern (therefore not shown).

The corresponding composites for summer are shown in Fig. 4. Extreme daily precipitation events are associated with a surface low pressure system centred over Austria and the Czech Republic. At 500 hPa, two low pressure minima are found. A stronger one located over northern France and a weaker one over south-eastern Europe. The thickness between 500- and 850-hPa shows warm air over western Europe, whereas over the British Isles and eastern Europe a cold anomaly is present (Fig. 4c). The moisture fluxes over Switzerland are weaker than in winter, and the main source of humidity in these situations is the Mediterranean Sea, in good agreement with results of Messmer et al. (2017). A positive precipitation anomaly is found over Switzerland in the ESM (Fig. 4d), which again shows the link between the precipitation simulated by the RCM and its driving dataset, i.e. the relationship exploited by Felder et al. (2018).

In summary, we find that extreme precipitation events identified in the WRF simulation are related to large scale circulation patterns of the driving CESM simulation. The flow patterns vary depending on the season. In winter (and similarly for autumn and spring) a west east dipole pattern with low pressure at the east becomes an important predictor of extreme events. In summer a low pressure system centred over Austria and the Czech Republic indicates a so-called Vb-cyclone situation (e.g., van Bebber 1891; Stucki et al. 2012; Messmer et al. 2015, 2017). This information is used below to define meteorological indices that are exploited by the ANN.

3.2 Network geometry and predictor variables

Once the variables candidate to be used as predictors are identified, we need to determine the geometry of the ANN, which includes the number of variables considered, but also the number and size of neurons in the hidden layers. The geometry of an ANN used in this study consists of an input, an output and a single hidden layer. The output layer of the model is the predictand, and consists of a single number, i.e. the daily mean precipitation in the Aare catchment. Thus, we need a single neuron in the output layer. The input layer contains the predictors, and has as many neurons as the number of variables considered for the downscaling. For the sake of simplicity, we set one single hidden layer, whose size is variable. Its number of neurons is not determined arbitrarily, but exhaustive tests are carried out to identify in each season the optimal number of neurons of the hidden layer that yields the highest skill of the ANN (steps 1 and 2 in Fig. 5).

A number of variables obtained from the ESM are considered as input for the network. The full list is shown in Table 1, and is based on previous literature (e.g., Martius et al. 2006; Chan et al. 2017) as well as the composite

Fig. 3 Composite analysis for extreme 1-day precipitation events for winter, i.e. DJF months: **a** Sea level pressure, **b** 500-hPa geopotential height, **c** difference between 500- and 850-hPa geopotential height, **d** precipitation and **e** water vapour transport vertically integrated up to 700 hPa
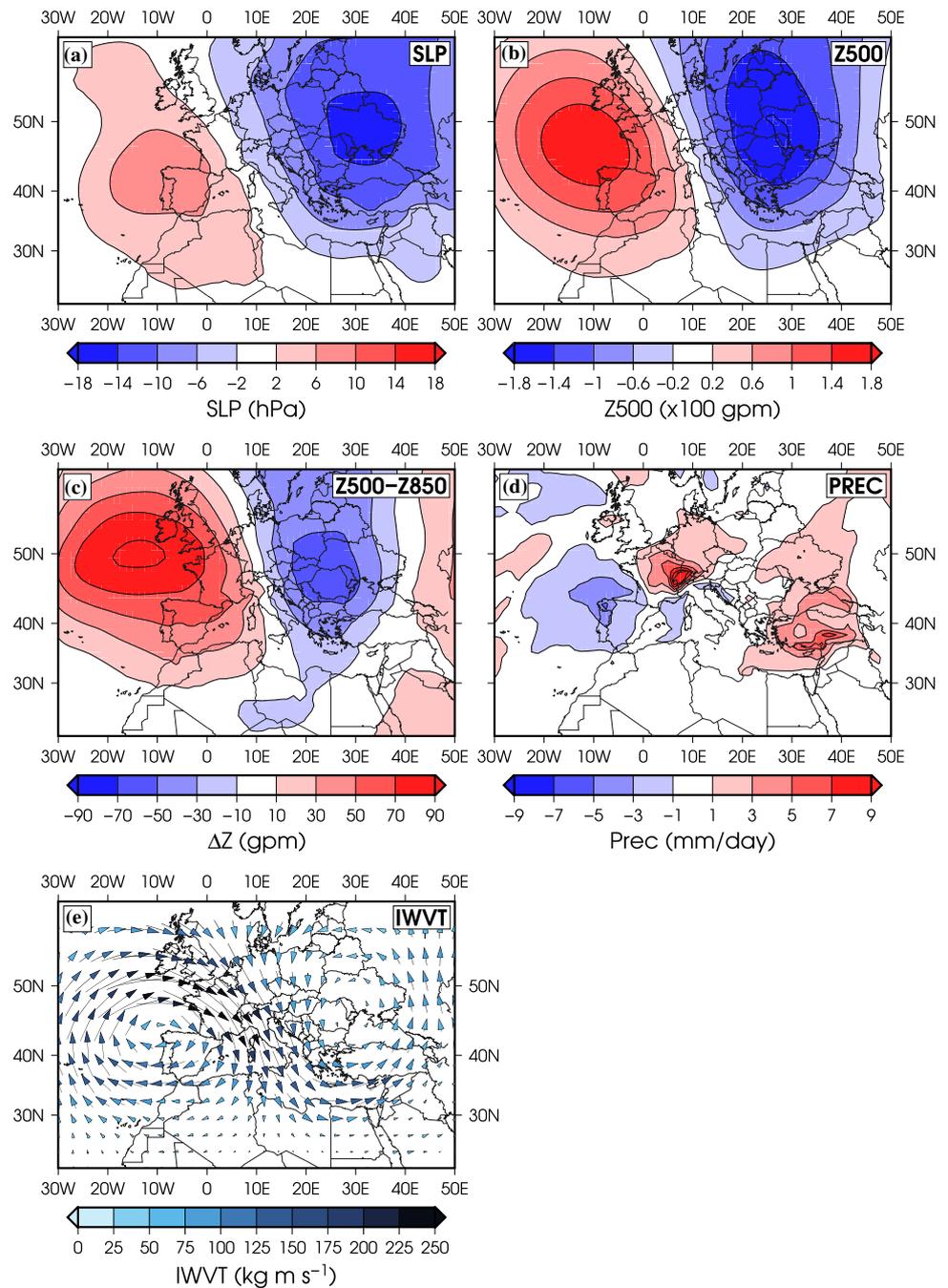
analysis described in Sect. 3.1 (note that we use different variables according to each season). Although in principle all variables are considered by the model to avoid prejudges and gain in generality, a first test allows to stablish which variables seem more closely related to the output variable, i.e., precipitation over the Aare catchment (see diagram labelled 1 in Fig. 5). To do so, each variable is used separately to build a simple ANN where it is the only input, and a single neuron exists in the hidden layer. In each case, the ANN is calibrated using half of the reference period, i.e. 1979–1992, and the fitted ANN is used to produce a

prediction for the other half, i.e. 1993–2005. This is then compared to the expected output, i.e. the dynamically downscaled precipitation, and their mutual correlation is calculated, which serves as metric to build the aforementioned ranking of the variables. In this step, other metrics than correlation were considered, such as Root Mean Square Error or Mean Averaged Bias. The results indicate modest sensitivity to the choice of the metric, so correlation was finally the metric used.

In a second step (see diagram labelled 2 in Fig. 5), the ranking is used as the base to calibrate more complex ANNs

Fig. 4 As Fig. 3 but for summer, i.e. JJA months
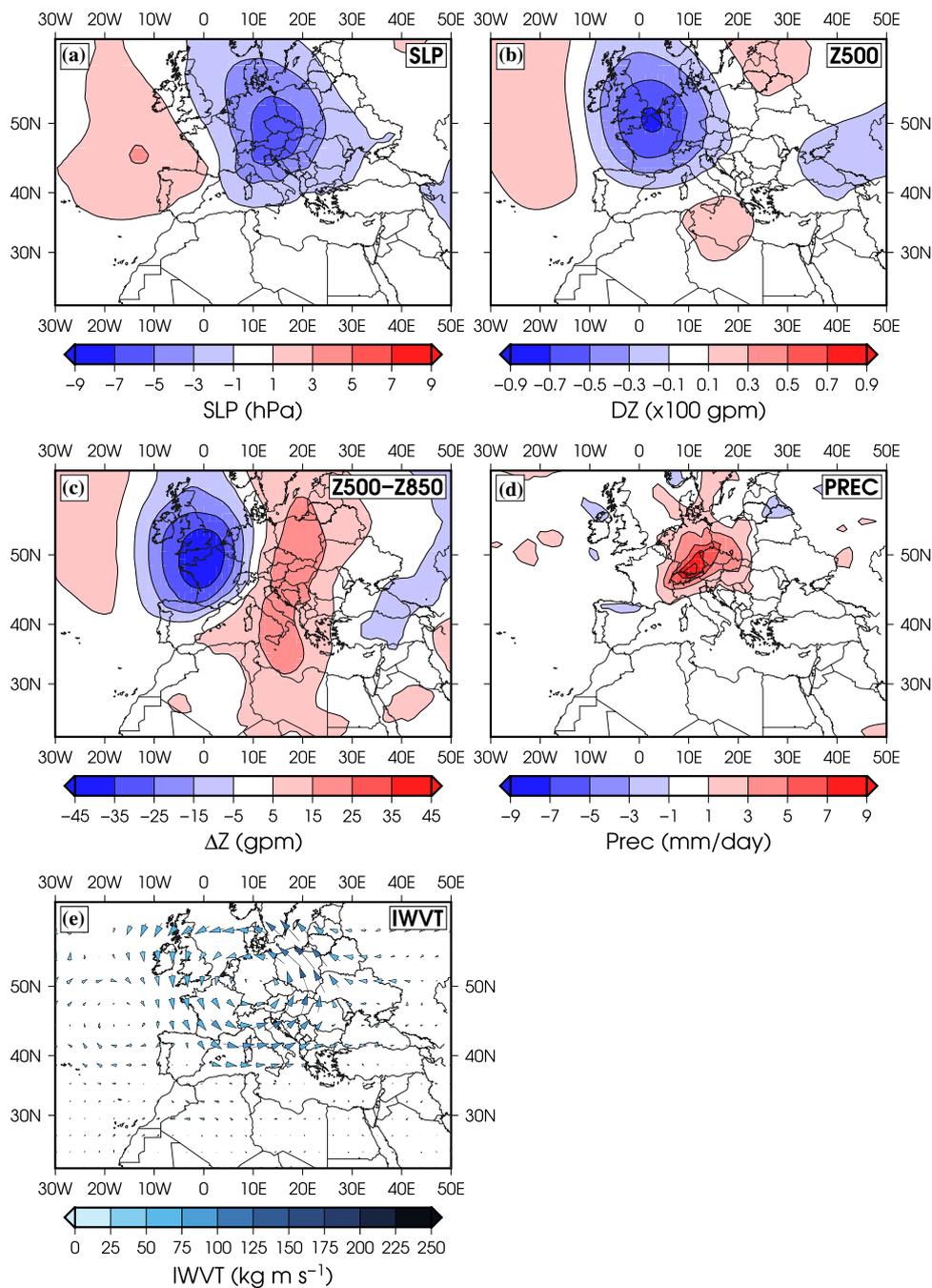
that include a growing number of variables and their interactions, and allow to determine the optimal geometry. To do so, we loop in the variables in the order defined by the previous ranking, including them in a growing set of predictors. In each step i of this loop, another loop is considered in j, the number of neurons in the hidden layer, which varies between 1 and i. Therefore, a total of $\frac{N^2+N}{2}$ ANNs, where N is the total number of variables in Table 1, are tested for each season. As before, the period 1979–1992 is used for calibration and 1993–2005 for validation. From all the combinations of number of variables and neurons, the optimal one, in the sense of maximising the correlation, is chosen as ultimate ANN geometry to perform the downscaling. Note that the role of calibration and validation periods is exchanged to complete the full reference period, allowing the cross validation of the results without circularity.

These tests result in a distinct combination of predictors and number of neurons in the hidden layer for each season separately. These configurations, together with the correlations obtained during the validation are shown in Table 2. Note that the number of variables does not grow monotonically, but reaches an optimal number for each season.
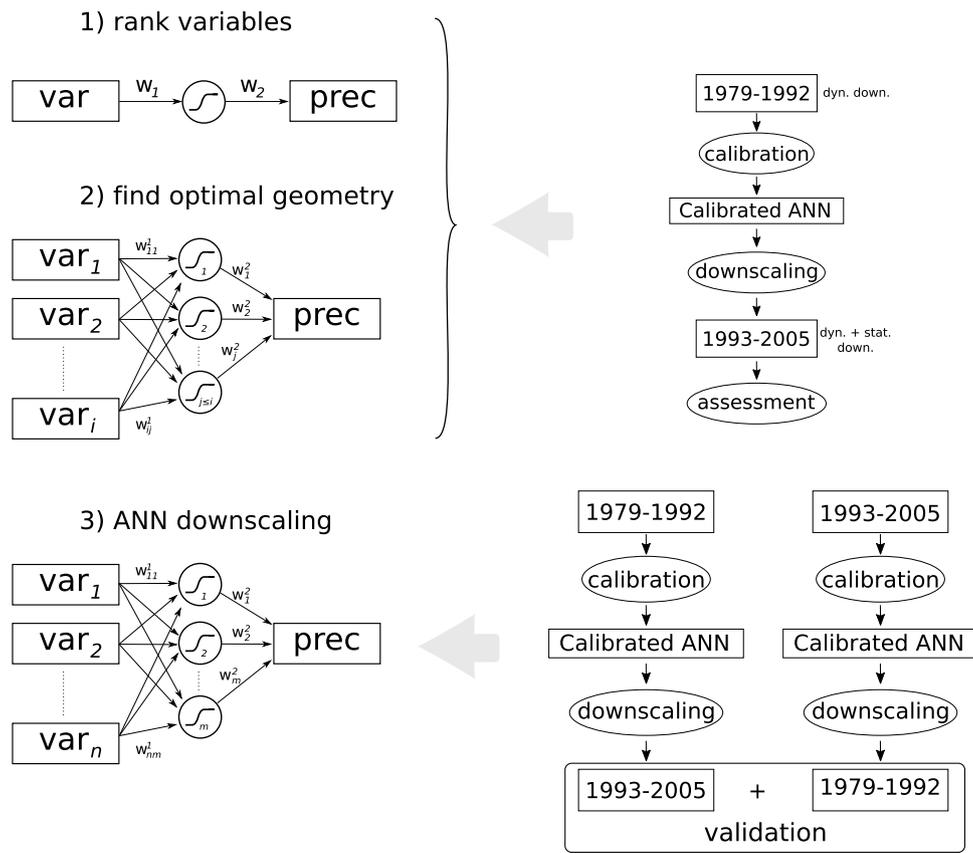
Fig. 5 Steps to find the optimal ANN geometry, which is then used to produce a statistical downscaling that can be used to validate the skill of the ANN during the WRF-reference period. (1) All variables considered are tested individually with a trivial ANN to build a ranking. Each ANN is calibrated during 1979–1992, and assessed by comparing the output of the ANN with the dynamically downscaled data set for the period 1993–2005. (2) This ranking is used to iteratively find the optimal network geometry among more realistic ANNs that allow interactions between variables. For this, all possible combinations are evaluated within a loop, which goes from 1 to N, where N is the total number of variables candidates to be included in the input

layer. In the step *i* of this iterative process, another loop is considered that evaluates ANNs with a variable number of neurons in the hidden layer between 1 and *i*. A case with *i* variables and $j \leq i$ neurons is represented in the figure. The calibration and assessment periods are defined as in the former step. (3) Once the number of variables and neurons is identified (labelled *n* and *m*, respectively), these parameters are fixed, and the WRF-reference period is statistically downscaled to validate the skill of the ANN. For this analysis a more comprehensive validation is carried out by splitting the full period in two halves and using the complementary part to calibrate and validate, respectively

Similarly, the inclusion of more neurons in the hidden layer does not necessarily improve the ANN. As expected, the precipitation of the ESM emerges as an important variable predicting precipitation in the target region, but in all cases the addition of further large-scale predictors improves the performance of the ANN. Indeed, in autumn zonal wind at 850 hPa is more skilful predicting local precipitation than the precipitation of the ESM. In agreement with the results in Sect. 3.1, the variables most relevant for DJF, MAM and SON are similar, being wind at various levels very illuminating predictors of extreme events. This is indicative of the close relationship between extreme events in this region and the moisture provided by the large-scale circulation. Summer again stands out as a distinct season, where the moisture provided by the Mediterranean Sea associated to low pressure centres in central Europe (note the presence

of IWVTXFRANCE and SLPCZECH, in the most relevant variables for JJA in Table 2), as well as instability measures (KI), emerge as important key variables to predict summer precipitation in this region.

3.3 Results of the event selection in the reference period

To assess the performance of the event selection based on ANNs, we compare it with the skill achieved by the simple approach employed by Felder et al. (2018). They used only extreme precipitation in the CESM-control simulation averaged over the Switzerland (bottom map of Fig. 1) as a criterion to identify dates which potentially deliver extreme precipitation over the Aare catchment. The selected dates correspond to the four most extreme precipitation events

Table 1 Variables directly taken or derived from the ESM used as predictors in the ANN

Season	Acronym	Levels	Description
All	PREC	Surface	Precipitation
All	SLP	Surface	Sea level pressure
All	Q	3D	Water vapour mixing ration
All	RH	3D	Relative humidity
All	T	3D	Temperature
All	U	3D	Zonal wind
All	V	3D	Meridional wind
All	Z	3D	Geopotential height
All	KI	Surface	K-Index
All	TTI	Surface	Total totals index
All	DIV	Surface, 500	Divergence
All	PV	3D	Potential vorticity
DJF, MAM, SON	SLPGRAD	Surface	SLP gradient
DJF, MAM, SON	IWVFRANCE	Atm. integrated	IWV over France
JJA	IWVTXFRANCE	Atm. integrated	Zonal component of IWV over France
JJA	SLPCZECH	Surface	SLP over the Czech Republic
JJA	Z500FRANCE	500	Geopotential over the France

Some variables are used systematically for all seasons, whereas others are used for certain seasons only, based on the results of the composite analysis described in Sect. 3.1. 3D refers to variables in several vertical levels, i.e. 1000, 850, 700 and 500 hPa. IWV refers to integrated water vapour. SLPGRAD is calculated as the difference between the SLP averaged in the regions 46°–49°N, 7°–12°W, and 45°–51°N, 28°–38°E. IWVFRANCE is defined as IWV averaged over the region 46°–49°N, 1°–7°W. IWVTXFRANCE is the zonal component of the integrated water vapour transport averaged over the region 39°–45°N, 5°–15°E. SLPCZECH is SLP averaged over the region 45°–52°N, 10°–20°E. Z500FRANCE is the geopotential height at 500 hPa averaged over the region 45°–53°N, $-2 - 8$°E

Table 2 Combination of variables and number of neurons in the hidden layer that, once calibrated in half of the reference period (1979–1992), lead to ANNs that maximise the correlation in the other half (1993–2005) for different seasons

Season	Predictors	Hidden neurons	Correlation
DJF	PREC, U700, U850, U500, V500, RH700, SLPGRAD, V700	8	0.83
MAM	PREC, U850, RH850, U700, U500, RH700, Z850, HR500, Z700, Z1000, Z500, SLP, KI, U1000, Q700, T850, VPO700, VPO850, T700, Q500, SLPGRAD, Q1000, TTI, V700, T500, RH1000, Q850, VPO500, V1000	2	0.69
JJA	PREC, IWVTXFRANCE, SLPCZECH, RH700, Z500FRANCE, Z1000, Z850, SLP, KI	2	0.69
SON	U8500, PREC, V700, SLPGRAD, RH700, V500, U700, RH500, Z700, U500, RH_850, Z850, Z500, Z1000	3	0.80

for each season in CESM-control, i.e. 16 cases in total, and were dynamically downscaled with WRF. Their results demonstrated that the precipitation obtained for these cases in winter and summer over the target region was far lower than expected, and motivated the new method presented in this study. The results by Felder et al. (2018) are extended here to the four seasons and analysed in more detail. The downscaled precipitation in all the selected cases is relatively large, exceeding all but one case the 90th percentile (see Table 3). Still, the precipitation obtained for these events is lower than the maximum precipitation during the reference period in nearly all cases. Given the length of the CESM-control simulation that was used to search for extremes, the

selection of dates aimed at providing physically consistent precipitation events with return periods of up to 400 years. However, the method failed in this regard. Only 1 event in summer leads to precipitation that can be considered as extreme in a 400-year frame, whereas for 10 out of 15 cases the estimated return period is below 10 years.

The skill of event selection based on precipitation within the ESM alone can be further assessed using the whole WRF-reference simulation. This way, we can compute various skill metrics between the precipitation predicted by the ESM over the target region and the one simulated by WRF during the complete reference period 1979–2005 (left column of Fig. 6). The precipitation within the ESM alone is

Table 3 Results of the downscaling of four single cases per season in the CESM-control simulation based solely on the precipitation simulated over Switzerland by the ESM

	Prec.	Perc.	Ret. Per.
DJF	13.1	76.4	1.0
	35.2	96.2	1.0
	85.3	99.9	10.6
	–	–	–
MAM	26.6	97.2	1.1
	28.2	97.5	1.2
	61.4	99.6	4.9
	78.1	99.8	11.3
JJA	12.0	94.2	1.0
	29.7	99.8	3.5
	43.3	99.9	23.7
	86.0	100.0	> 400
SON	13.9	91.6	1.0
	21.4	95.9	1.1
	41.9	99.2	2.6
	86.0	99.9	18.0

The table depicts, separately for each season, the precipitation accumulated over the Aare catchment in each of the four cases once dynamically downscaled with WRF (in winter only three cases could be run due to numerical instabilities). For each event, three numbers are presented: the precipitation value in mm, the percentile it represents within the PDF obtained for the WRF-reference period (1979–2005), and the return period (in years) of such precipitation estimated using the data in the same period. In the latter case, when the value exceeds 400 years, it is indicated as "> 400". The return periods are obtained by fitting the parameters of a Generalised Extreme Value distribution to the data with the aid of the extRemes package of R (Gilleland and Katz 2016)

a poor predictor of precipitation at local scale. Considering all days, correlations are around 0.4 in all seasons (note that the Hit Rate trivially converges to 1 at low percentiles by its very definition). However, when higher percentiles are considered, the skill rapidly deteriorates in all seasons. Above the 90th percentile, the correlation ranges between 0.3 (in winter) and 0.1 (in summer), and the Hit Rate is about 0.3 in all seasons. For percentiles above the 99th percentile, the Hit Rate rapidly approaches zero and becomes insignificantly different from random chance in all seasons. SEDI is stable through most of the percentiles, as it is expected for the properties that define this index (Ferro and Stephenson 2011). Again it shows that the selection of events has certain skill in moderate percentiles, but it rapidly deteriorates towards rarer events.

We focus now on the approach based on ANNs. The optimal geometry for each season is fixed as indicated in Table 2 and determined by the tests in Sect. 3.2. Next, we apply the approach represented by the diagram 3 in Fig. 5. This is, an ANN for each season is calibrated during the first half of the WRF-reference period, and used to forecast the

daily precipitation in the Aare catchment during the second half. These periods are exchanged to ensure that the full period is statistically downscaled using independent data for the calibration. The validation, based on the comparison between the WRF output and the optimal ANN for each season, is summarised in the results of the middle column of Fig. 6. Correlations are systematically higher than those for the more simple method in all seasons and percentiles. Considering all days, correlation is about 0.8 in spring and autumn, with slightly higher and lower values in winter and summer, respectively. As before, the skill decreases towards higher percentiles, but in clear contrast to the performance of the simple approach by Felder et al. (2018), in all seasons but summer the correlation is nearly constant up to the highest percentiles. The results are worse in summer, where the ANN monotonically decreases its performance towards higher percentiles. Still, in this season the ANN demonstrates valuable skill up to the 99th percentile, where correlation reaches a critical value of 0.3. This is again in contrast to the simple approach (left column of Fig. 6) showing no skill above the 90th percentile. The Hit Rate also shows the seasonal differences. The ANN's ability to select the right dates is best in winter, with probabilities of detecting an event that actually happen above 0.6 in percentiles above 99th, and worst in summer, when it drops to 0.1. But even in that case, and unlike in the simpler approach, the ANN is able to capture the 99th percentile events better than pure chance. In both, spring and autumn, there is a somewhat unexpected improvement in the metrics beyond the 90th percentile that is provoked by the ability of the ANN to capture the precise ordering of the most extreme cases (not shown), and implies that the ANN is able to adequately select between 60 and 80% of the extreme events in these seasons. SEDI remains remarkably stable, and even grows towards higher percentiles in all seasons but summer. The values are systematically higher across all percentiles than those of the simple approach, and the skill is comparable to that obtained for the 12–24 h precipitation forecasts of the ECMWF, where the seasonal cycle in the forecasting performance is also observed (North et al. 2013).

The ability of the ANN to predict the correct extreme events during the reference period is further evaluated by comparing the magnitude of the precipitation in observed and predicted extremes. This is presented through boxplots in Fig. 7. This figure shows, for each season, the distribution of precipitation in the days when it is observed to be above the 99th percentile (black), and compares it to the distribution when the days are those predicted by the ANN (blue). For comparative purposes, this figure also includes dark cyan boxes that represent the same results obtained when the precipitation in the ESM alone is used as predictor of precipitation, i.e. the simple approach used by Felder et al. (2018). First, we note that the left tail of the distribution is longer
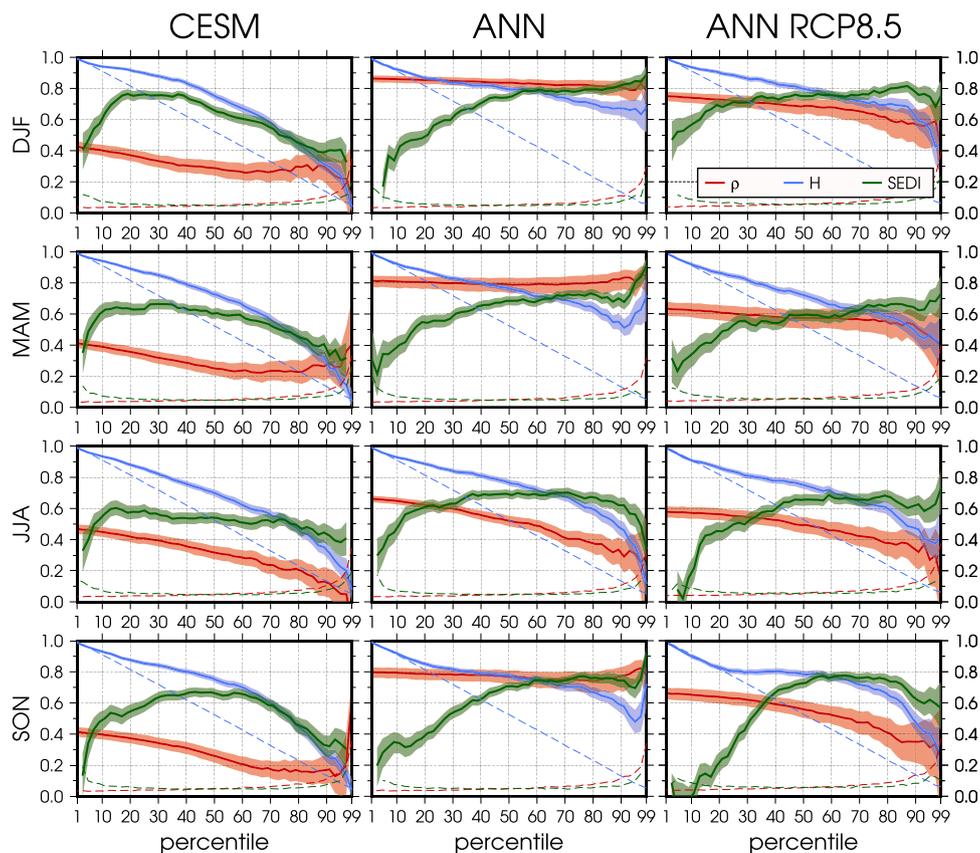
Fig. 6 Agreement between various predictors of simulated precipitation over the Aare catchment and the eventually dynamically downscaled one. Each row represents the result for a given season. The left column compares the precipitation averaged over Switzerland in the ESM data with the WRF-reference simulation during the 1979–2005 period (i.e. the approach by Felder et al. 2018). The central column shows the same, but with respect to the output of the ANNs calibrated for each season separately. The right column compares the WRF-future simulation with the output of the ANNs driven by the CESM-future in the period 2080–2099, but calibrated during the reference period (see Sect. 4). Three metrics are shown: correlation (red), Hit Rate (blue) and SEDI (green). The results are shown as a function of the percentile p used to filter out the series to keep the days where precipitation is above the given quantile. The solid lines represent the median, whereas shadows represent the 5–95 range, as obtained by bootstraping the sample with repetition. Dashed lines represent the threshold to reject the null hypothesis of skill by random chance at the 95% confidence level

in the events predicted by the ANNs (compare black and blue boxplots across seasons), and the precipitation in the cases selected by the simple approach is strongly underestimated. This is the expected behaviour, as we already know from the previous analysis that about 50% of the events are incorrectly attributed by the ANN to extreme events in this percentile. Still, in all seasons but summer the median is well captured, as well as the right tail of the distribution, i.e. the absolute most extreme events during the period. In agreement with the analysis above, in summer the predictive skill is the lowest. Only around 10% of the cases are correctly predicted (see Hit Rate in Fig. 6) and these correspond to two marked outliers. In this regard, it noteworthy that all the absolute maxima for each season during the reference period have been captured by the ANNs.

4 Detection of extremes in climate change projections

For the ANNs to be successful, and more generally for any statistical downscaling tool, the inferred predictor-predictand relationships must be stationary between the period used to calibrate and the one where the model is applied. But given that the actual climate is not completely stationary, these relations could not be either, which can limit the reliability of statistical methods for downscaling under perturbed climate conditions. We have taken the opportunity provided by the WRF-future simulation to evaluate the performance of statistical downscaling under the unfavourable circumstances of a climate severely perturbed with respect to the one used to calibrate the model.

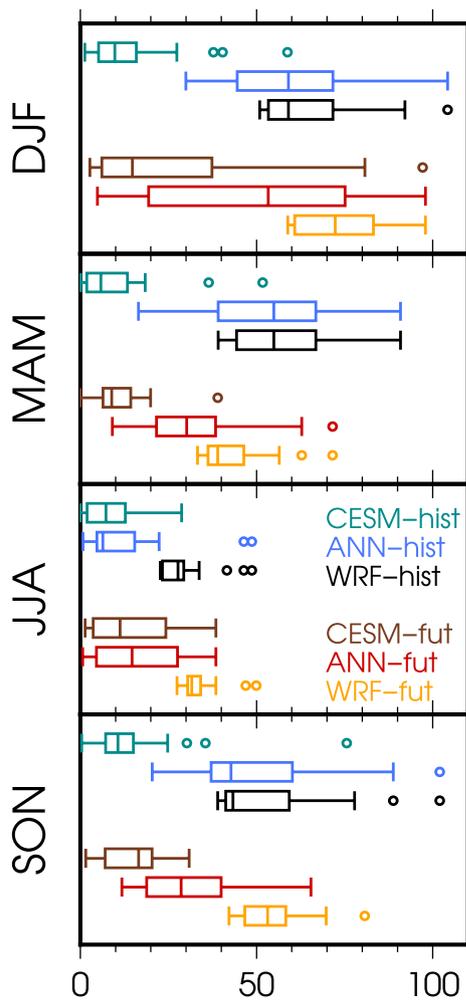In detail, the optimal ANN calibrated during the WRF-reference period has been used to downscale the

Fig. 7 Distribution of daily precipitation in the Aare catchment. Each boxplot represents different datasets and periods. Black, blue and dark cyan correspond to the WRF-reference simulation, and highlight daily precipitation above the 99th percentile during the reference period (1979–2005): black corresponds to observed extremes, whereas blue and dark cyan corresponds to the extreme precipitation days in WRF, but as predicted by the ANN and the precipitation in the ESM alone, respectively. Red, orange and brown represent the same information but for the WRF-future simulation (2080–2099), discussed in Sect. 4: orange represents the actual extreme events observed in the WRF-future simulation, whereas red and brown correspond to the cases predicted by the ANN and the precipitation in the ESM alone, respectively

CESM-future simulation during the period 2080–2099. This prediction is compared with the WRF-future simulation following the same protocol than in the Sect. 3.3. The result of this analysis is presented in the right column of Fig. 6. Compared to the reference period, the ANN generally presents lower predictive skill. In winter, correlation ranges between 0.75 and 0.6, whereas it is above 0.8 in the reference period. The Hit Rate is very similar up to the 80th percentile, but above this point the probability of predicting extreme events drops faster, reaching 0.4 for the 99th

percentile (compare this with the value about 0.7 during the reference period). This worsening of the skill is also visible in lower SEDI values in the highest percentiles. A similar behaviour is found in spring, albeit with slightly lower correlations, closely resembling the results during the reference period. The Hit Rate is very similar in both periods until the 90th percentile, where the increase of the Hit Rate observed in the reference period is absent in the future period, thus also SEDI is reduced. In summer, the ANN presents the lowest correlations, but they are remarkably similar to those during the reference period, showing that the reduction of skill of the ANN under future climate conditions is not very pronounced for summer. Indeed, above 80th percentile the Hit Rate is higher under future climate conditions, leading to values comparable to other seasons. In autumn, correlations are strongly reduced compared to the reference period, and are close to those for summer. This difference is also found in the Hit Rate above 80th percentile, which leads to the lowest Hit Rate of all seasons in the future simulation, and the strongest reduction of SEDI. This is in clear contrast to the behaviour of the reference period.

The lower skill in selecting the days with the most severe precipitation under future conditions compared to the reference period is related to the generally lower correlation and Hit Rate (see middle and right columns in Fig. 6), in particular during autumn. This reduced performance is attributed to the fact that the ANNs are trained to learn the relationships between synoptic and local-scale variables during a relatively short period, which are then implicitly assumed to be stationary as part of the statistical downscaling exercise. The reduction in skill under future climate conditions, however, suggest that at least part of this stationary is not perfectly fulfilled, so that climate change can indeed affect the mechanisms learned by the ANN and exploited during the statistical downscaling, making the calibration sensitive to the period used as reference. This has important implications in the way the results of this approach in the detection of trends shall be interpreted, as discussed below.

Figure 7 represents the distribution of extremes observed (orange) and predicted by the ANNs (red) under climate change conditions. As before, for comparison we include the results when the extreme events are searched with the precipitation in the driving ESM alone (brown). As with the reference period, the longer left tails are expected as a consequence of the non-perfect Hit Rate. Unlike in the reference period, a systematic bias stands out, i.e. a general underestimation of precipitation in the cases selected by the ANNs. Although it is far from the low skill of the simple approach, the median for the ANN-selected cases is below the 25th percentile for the WRF-future simulation. Regarding the most extreme events, in winter and spring the ANNs are able to predict the most extreme events, whereas

in summer and autumn the two most extreme events are not identified with the ANNs, respectively.

The signal of climate change on extreme events should be sought in a shift in the blue and red boxplots in Fig. 7, respectively. However, instead of a systematic shift towards more severe extremes, which could be expected according to other studies (e.g. Seneviratne et al. 2012; Rajczak et al. 2013, 2016; Messmer et al. 2017) as well as basic thermo-dynamic relationships (e.g. O'Gorman 2015), we notice a rather stationary behaviour with modest changes attributable to sampling uncertainty. Further, the analysis of the distribution of these extremes in consecutive periods shows a lack of trend towards more severe or frequent extremes (not shown). Therefore, the ANNs suggest a lack of sensitivity of extreme precipitation events to climate change in all seasons.

This contrasts with the results we can draw by comparing the WRF-reference and WRF-future runs (see the horizontal shift in the black vs. orange boxplots in Fig. 7). Using the full transient runs, we find more severe extremes in winter, summer and to a lesser extent in autumn, with a strong opposite behaviour in spring. This behaviour is concurrent with an overall increase in precipitation not only in the extremes but also on average in these seasons (not shown), which better agrees with the sensitivity of extremes to a warming climate reported for the aforementioned studies. Similar conclusions can be drawn from the application of Extreme Value Theory to these data. Figure 8 depicts the return level plots for daily precipitation in each season in the WRF-reference and WRF-future simulations. In winter and summer, the return levels are systematically higher in the WRF-future simulation than in WRF-reference. The lower climate change signal described above for autumn can be understood under the light of this analysis as a mixed

behaviour between the events with return levels below and above 5 years. In contrast to winter and autumn, the events with longer return periods are ameliorated by climate change in autumn according to these simulations, although the uncertainty in this range is large due to the modest number of events that support this conclusion. Finally, as described above respect to Fig. 7, spring stands out in the return level plot as an anomalous season, where climate change seems to reduce the occurrence of extreme events.

Based on the discussion above, the apparent lack of sensitivity to climate change identified by the ANNs has to be attributed, at least in all seasons but spring, to a limitation of statistical downscaling regarding its ability to identify extremes under perturbed climate conditions, rather than to an outcome of the CESM-WRF simulations. This is a problem than can introduce subtle and non-systematic biases that largely affect the study of trends, so it would be desirable to identify its causes. In this regard, in the application of statistical downscaling it is recommended to use predictors that clearly capture the local thermodynamic state of the atmosphere through which the climate change signals are communicated (Hewitson et al. 2014). In principle, the downscaling exercise we present based on ANNs fulfils this condition, as we introduce through the input layer variables that represent the thermodynamic state of the atmosphere, e.g. temperature, relative humidity or geopotential height. However, we do not force these thermodynamic relationships to be more strongly exploited than others not closely related to climate change. It is the calibration during the reference period what implicitly determines the relative importance of all physical mechanisms through the determination of the set of weights that define the ANNs behaviour. A possible modification of this approach could try to

Fig. 8 Return levels for daily precipitation over the Aare catchment in the WRF-reference (black) and WRF-future (orange) simulations. The analysis relies on the Peak Over Threshold (POT) approach of Extreme Value Theory. Solid lines depict the Generalised Pareto Distribution fitted to the values above the 95 percentile (dots) for each simulation. Dashed lines represent the confidence interval, obtained from the observed information matrix. This analysis is carried out using the POT package of the R language (https://cran.r-project.org/web/packages/POT/index.html)
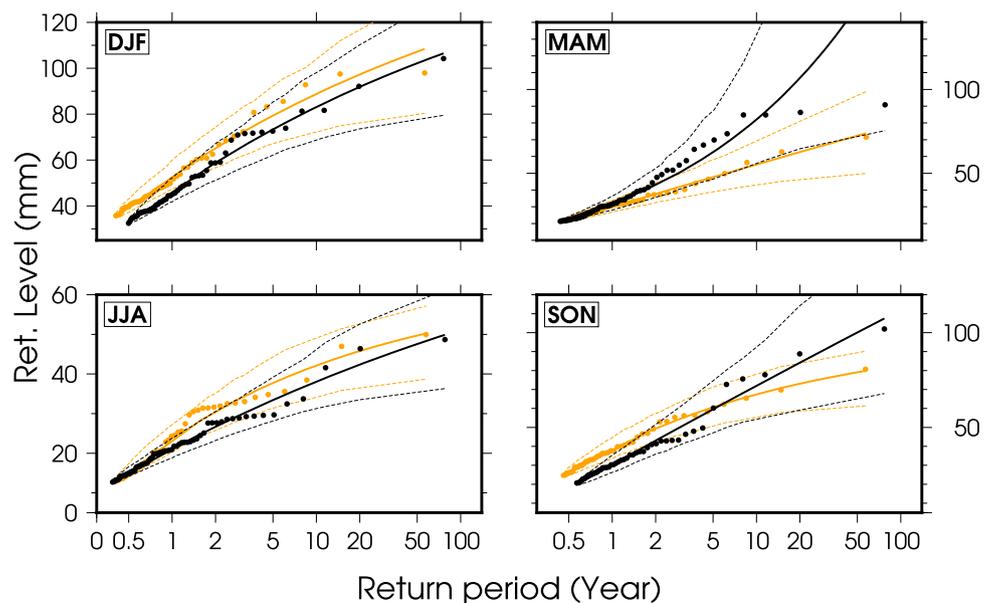
minimise the importance of certain mechanisms through the removal of some variables not related to the moisture/thermodynamic state of the atmosphere. This could improve the predictability of extreme events under perturbed climate conditions at the expense of a having sub-optimal performance during the reference period. However the identification of what variables can be removed is cumbersome and prone to the introduction of further biases. This is illustrated by the results obtained with the simple approach by Felder et al. (2018), which can be seen as an extreme example of this idea (compare dark cyan vs. brown boxes in Fig. 7). This simple approach, which completely ignores the thermodynamic state or even the large-scale dynamic of the atmosphere, seems to capture certain sensitivity to climate change, especially in winter when the large-scale precipitation is a better predictor of its local counterpart. But in spring the signal is reversed, and generally the skill of the approach is so low and barely distinguishable from random chance that this climate change signal might be simple statistical artefact.

5 Conclusions

This study proposes and evaluates the feasibility of a dynamical downscaling strategy to study extreme precipitation events at local scales from low-resolution comprehensive ESMs. It is based on the simulation of case studies, rather than on running continuous and in many cases unaffordable simulations. The main advantage is the reduced computational cost, which in turn can be used to increase the spatial resolution, thus becoming an approach especially suitable for the simulation of extreme precipitation in regions of very complex topography (Ban et al. 2014; Gómez-Navarro et al. 2018). The central challenge of this approach is the selection of the adequate dates to be downscaled, as internal variability within this type of freely evolving ESM simulations precludes the selection of known historical events.

We propose a method to select target days to downscale from the ESM simulations that blends dynamical and statistical downscaling, and is similar in its aim to the method proposed by Meredith et al. (2018). First we set up an ANN that uses large-scale ESM variables as predictors, and local downscaled precipitation as predictand. This model is trained to mimic the ESM-RCM coupling over the target region in a computationally affordable period, in this case the Aare catchment during 1979–2005. After a careful training and crossvalidation, we use the obtained ANNs, one for each season, to produce precipitation series that span an arbitrary long period within the ESM run, and that is used to search for candidate extreme events. Unlike the approach by Meredith et al. (2018), our method relies on existing RCM simulations rather than observations.

The results of the ANNs are evaluated by comparing them to the dynamically downscaled precipitation over the reference period. These results show that the ANNs are able to effectively blend information from different variables, and result in a powerful predictor of local precipitation. The ability of the statistically downscaled series to select the most extreme precipitation events at local scales is worse when higher percentiles are considered, although this effect becomes noticeable only in the highest percentiles in all seasons but summer. In summer, the method provides considerably lower skill in all percentiles, although still significantly better than a pure random selection. This is to some extent expected, as extreme precipitation events in this season are less strongly driven by the large-scale circulation, but by convective processes (e.g., Panziera et al. 2018), and therefore the information that can be provided by the driving ESM has fewer potential to explain the variability of precipitation at such local scale.

Finally, we use an existing high-resolution climate change projection to evaluate the sensitivity of the method to the non-stationarity of actual climate. The ANNs trained during the reference period have been tested under the RCP 8.5 scenario, searching for events above the 99th percentile using the CESM-future simulation as input for the statistical model, and comparing the output with WRF-future. The events selected after the application of the ANNs are overall extreme, with the majority of events above the expected percentile. Still, the performance of the ANN method is lower than during the reference period, which can be attributed to the fact that the model exploits relationships between variables learnt for different climate conditions. A number of events are erroneously identified as severe, although this is expected and agrees with the Hit Rates obtained during the reference period, and demonstrates the ability of the ANNs to predict extreme events with remarkable performance even in climate conditions very different to the ones used to calibrate the model. The analysis of the response of the severity of these events to climate change evidences no trend. This could be erroneously interpreted as a lack of sensitivity of extreme precipitation under climate change in the simulations, which would be in contradiction with other studies (e.g. Seneviratne et al. 2012; Rajczak et al. 2013). Indeed, the comparison of the transient simulations WRF-reference and WRF-future rules out this possibility, and instead hints to a limitation of the ANNs to capture the thermodynamic mechanisms responsible for this trend. As indicated by Hewitson et al. (2014), the predictor variables must represent the thermodynamic state of the atmosphere through which climate change is communicated. This condition is in principle satisfied by our approach, although we do not overweight in any way the relationships related to climate change. This suggests a possible way to improve the methodology, but such modifications are not self-evident and are

prone to severely reduce the skill of the ANNs. Therefore, our results suggest that statistical methods which rely on the assumption of stationarity of the statistical relationships between predictand and predictors for reference and change climate states, may be unsuitable for correctly identifying trends.

Appendix: Skill metrics

Correlation

We use Pearson correlation. This metric evaluates the co-variability of two series disregarding possible systematic biases, therefore being especially suitable for the evaluation of the ANN to predict the right timing of extreme events. We repeated the calculation with Spearman correlation and the results are similar (not shown).

As we are especially interested in the performance towards the most extreme events, correlations are successively calculated after the daily series are filtered out to retain only the values of precipitation above a given quantile q that corresponds to percentiles p between 1 and 99. In detail, all days in which precipitation in the dynamically downscaled series above q are selected, and the correlation coefficient between the latter and the series for the ANN within this subset of dates is calculated. This process of successive recalculation of the statistics filtering out the data towards higher percentiles is repeated for all skill metrics described here. Note that as we move towards higher percentiles, the length of the series becomes shorter, which leads to larger uncertainty in the estimation of the skill metrics. This uncertainty is estimated by jointly bootstraping the series with repetition (shadings in Fig. 6 represent the confidence interval at $\alpha = 0.1$, while solid lines represent the median). Further, the value that rejects the null hypothesis of no skill at $\alpha = 0.05$ is obtained by independently bootstrapping both series with repetition (dashed curves in Fig. 6).

Hit rate *F*

In the evaluation of the skill of predicting rare events, it is common to use contingency tables (Skamarock 2000; Ferro and Stephenson 2011). Thereby, each event can fall in one out of four categories: either it is correctly predicted (hit), incorrectly predicted when it did not happen (false alarm), incorrectly non predicted with it actually happened (missed event) or it can be correctly rejected (most common situation). It is customary to name the number of the events within these disjoint sets as a, b, c and d, respectively. Given this notation, the Hit Rate is defined as (e.g. Skamarock 2000):

$$H = \frac{a}{a + c} = \hat{p}(f|o), \tag{1}$$

which can be interpreted as the probability of predicting a situation (event f, where f stands for "predicted") given that it actually happened (event o, where o stands for "observed"). In a similar fashion, we can define the false alarm rate F as:

$$F = \frac{b}{b + d} = \hat{p}(f|\bar{o}), \tag{2}$$

representing the probability of incorrectly having predicted a situation that did not happen.

A detail to be determined is how to define whether an event happened or not in either the observations or the predicted dataset. For instance, if a given threshold of precipitation is fixed for both datasets, it might be that the total number of events above such threshold differs between the two datasets, leading to a systematic bias, defined as:

$$B = \frac{a + b}{a + c}. \tag{3}$$

Values of B other than 1 indicate a systematic bias between the observations and the predicted dataset. However, this bias is meaningless to us, as we are not interested in the given values of precipitation provided directly by the ANN, but in their ranking of most extreme values, which will be ultimately used to select the events to be downscaled dynamically. Therefore, we carry out a form of hedging to the data that consists of working with quantiles. This is, for a given a percentile p, we obtain the corresponding quantiles separately for the statistical and dynamical downscaling series (as they are in general different if the ANN is biased). Then, we define that an event happened in one of the series when the precipitation in a given day is above its respective quantile. Summing the number of events, leads to the numbers a, b, c and d of the contingency table, which ultimately determines H for a given percentile p. As describe above, this calculation is repeated for p ranging between 1 and 99.

Symmetric Extremal Dependence Index

The Symmetric Extremal Dependence Index (SEDI) was proposed by Ferro and Stephenson (2011) as an alternative metric to evaluate the skill in predicting rare events that supersedes a number of drawbacks of more simple metrics,

such as H. It is still based on the calculation of a contingency table, and as such it is defined as a function of a, b, c and d:

$$\text{SEDI} = \frac{\log F - \log H - \log(1 - F) + \log(1 - H)}{\log F + \log H + \log(1 - F) + \log(1 - H)}. \quad (4)$$

SEDI has the advantage of being base rate independent, non degenerate and asymptotically equitable (Ferro and Stephenson 2011). The calculation of SEDI for different percentiles p has been performed following the same procedure as for H.

References

Ban N, Schmidli J, Schär C (2014) Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. J Geophys Res Atmos 119(13):7889–7907. https://doi.org/10.1002/2014JD021478

Bowden JH, Talgo KD, Spero TL, Nolte CG (2016) Assessing the added value of dynamical downscaling using the standardized precipitation index 2016:1–14. https://doi.org/10.1155/2016/8432064. http://www.hindawi.com/journals/amete/2016/8432064/

Chan SC, Kendon EJ, Roberts N, Blenkinsop S, Fowler HJ (2017) Large-scale predictors for extreme hourly precipitation events in convection-permitting climate simulations. J Clim 31(6):2115–2131. https://doi.org/10.1175/JCLI-D-17-0404.1

Dawson CW, Wilby R (1998) An artificial neural network approach to rainfall-runoff modelling. Hydrol Sci J 43(1):47–66. https://doi.org/10.1080/02626669809492102

Doswell CA, Brooks HE, Maddox RA (1996) Flash flood forecasting: an ingredients-based methodology. Weather Forecast 11(4):560–581. https://doi.org/10.1175/1520-0434(1996)011<0560:Fffaib>2.0.Co;2

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016) Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. Geosci Model Dev 9(5):1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fantini A, Raffaele F, Torma C, Bacer S, Coppola E, Giorgi F, Ahrens B, Dubois C, Sanchez E, Verdecchia M (2016) Assessment of multiple daily precipitation statistics in ERA-interim driven med-CORDEX and EURO-CORDEX experiments against high resolution observations. Clim Dyn. https://doi.org/10.1007/s00382-016-3453-4

Felder G, Gómez-Navarro JJ, Zischg AP, Raible CC, Röthlisberger V, Bozhinova D, Martius O, Weingartner R (2018) From global circulation to local flood loss: coupling models across the scales. Sci Total Environ 635:1225–1239. https://doi.org/10.1016/j.scitotenv.2018.04.170

Ferro CAT, Stephenson DB (2011) Extremal dependence indices: improved verification measures for deterministic forecasts of rare binary events. Weather Forecast 26(5):699–713. https://doi.org/10.1175/WAF-D-10-05030.1

Froidevaux P, Martius O (2016) Exceptional integrated vapour transport toward orography: an important precursor to severe floods in Switzerland. Q J R Meteorol Soc 142(698):1997–2012. https://doi.org/10.1002/qj.2793

García-Valdecasas Ojeda M, Gámiz-Fortis SR, Castro-Díez Y, Esteban-Parra MJ (2017) Evaluation of WRF capability to detect dry and wet periods in Spain using drought indices: WRF capability to detect drought. J Geophys Res Atmos 122(3):1569–1594. https://doi.org/10.1002/2016JD025683

Gent PR, Danabasoglu G, Donner LJ, Holland MM, Hunke EC, Jayne SR, Lawrence DM, Neale RB, Rasch PJ, Vertenstein M, Worley PH, Yang ZL, Zhang M (2011) The community climate system model version 4. J Clim 24:4973–4991. https://doi.org/10.1175/2011JCLI4083.1

Giannakaki P, Martius O (2016) Synoptic-scale flow structures associated with extreme precipitation events in northern Switzerland. Int J Climatol 36(6):2497–2515. https://doi.org/10.1002/joc.4508. https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.4508

Gilleland E, Katz RW (2016) extRemes 2.0: an extreme value analysis package in r 72(8). https://doi.org/10.18637/jss.v072.i08. http://www.jstatsoft.org/v72/i08/

Giorgi F, Torma C, Coppola E, Ban N, Schär C, Somot S (2016) Enhanced summer convective rainfall at alpine high elevations in response to climate warming. Nat Geosci 9(8):584–589. https://doi.org/10.1038/ngeo2761

Gómez-Navarro JJ, Raible CC, Bozhinova D, Martius O, García Valero JA, Montávez JP (2018) A new region-aware bias-correction method for simulated precipitation in areas of complex orography. Geosci Model Dev 11(6):2231–2247. https://doi.org/10.5194/gmd-11-2231-2018

Hewitson BC, Daron J, Crane RG, Zermoglio MF, Jack C (2014) Interrogating empirical–statistical downscaling. Clim Change 122(4):539–554. https://doi.org/10.1007/s10584-013-1021-z

Hurrell JW, Holland MM, Gent PR, Ghan S, Kay JE, Kushner PJ, Lamarque JF, Large WG, Lawrence D, Lindsay K, Lipscomb WH, Long MC, Mahowald N, Marsh DR, Neale RB, Rasch P, Vavrus S, Vertenstein M, Bader D, Collins WD, Hack JJ, Kiehl J, Marshall S (2013) The community earth system model: a framework for collaborative research. Bull Am Meteorol Soc 94:1339–1360. https://doi.org/10.1175/BAMS-D-12-00121.1

Lee S, Cho S, Wong PM (1998) Rainfal prediction using artificial neural networks. J Geogr Inf Decis Anal 2(2):233–242

Lehner F, Joos F, Raible CC, Mignot J, Born A, Keller KM, Stocker TF (2015) Climate and carbon cycle dynamics in a CESM simulation from 850 to 2100 CE. Earth Syst Dyn 6(2):411–434. https://doi.org/10.5194/esd-6-411-2015

Li G, Zhang X, Zwiers F, Wen QH (2011) Quantification of uncertainty in high-resolution temperature scenarios for North America. J Clim 25(9):3373–3389. https://doi.org/10.1175/JCLI-D-11-00217.1

Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. Rev Geophys 48:3. https://doi.org/10.1029/2009RG000314

Maraun D, Widmann M, Gutiérrez JM, Kotlarski S, Chandler RE, Hertig E, Wibig J, Huth R, Wilcke RA (2015) VALUE: a framework to validate downscaling approaches for climate change studies. Earth's Future 3(1):1–14. https://doi.org/10.1002/2014EF000259

Martinez Y, Yu W, Lin H (2012) A new statistical–dynamical downscaling procedure based on EOF analysis for regional time series generation. J Appl Meteorol Climatol 52(4):935–952. https://doi.org/10.1175/JAMC-D-11-065.1

Martius O, Zenklusen E, Schwierz C, Davies HC (2006) Episodes of alpine heavy precipitation with an overlying elongated stratospheric intrusion: a climatology. Int J Climatol 26:1149–1164

Meredith EP, Rust HW, Ulbrich U (2018) A classification algorithm for selective dynamical downscaling of precipitation extremes. Hydro Earth Syst Sci 22(8):4183–4200. https://doi.org/10.5194/hess-22-4183-2018

Messmer M, Gómez-Navarro JJ, Raible CC (2015) Climatology of vb cyclones, physical mechanisms and their impact on extreme

precipitation over central europe. Earth Syst Dyn 6(2):541–553. https://doi.org/10.5194/esd-6-541-2015

Messmer M, Gómez-Navarro JJ, Raible CC (2017) Sensitivity experiments on the response of vb cyclones to sea surface temperature and soil moisture changes. Earth Syst Dyn 8(3):477–493. https://doi.org/10.5194/esd-8-477-2017

North R, Trueman M, Mittermaier M, Rodwell MJ (2013) An assessment of the SEEPS and SEDI metrics for the verification of 6 h forecast precipitation accumulations: assessment of SEEPS and SEDI for 6 h precipitation accumulations. Meteorol Appl 20(2):164–175. https://doi.org/10.1002/met.1405

O'Gorman PA (2015) Precipitation extremes under climate change. Curr Clim Change Rep 1(2):49–59. https://doi.org/10.1007/s40641-015-0009-3

Panziera L, Gabella M, Germann U, Martius O (2018) A 12-year radar-based climatology of daily and sub-daily extreme precipitation over the Swiss alps. Int J Climatol 38(10):3749–3769. https://doi.org/10.1002/joc.5528

Pfahl S, Madonna E, Boettcher M, Joos H, Wernli H (2014) Warm conveyor belts in the era-interim dataset (1979–2010). Part II: Moisture origin and relevance for precipitation. J Clim 27(1):27–40. https://doi.org/10.1175/Jcli-D-13-00223.1

Piaget N, Froidevaux P, Giannakaki P, Gierth F, Martius O, Riemer M, Wolf G, Grams CM (2015) Dynamics of a local alpine flooding event in October 2011: moisture source and large-scale circulation. Q J R Meteorol Soc 141(690):1922–1937. https://doi.org/10.1002/qj.2496

Rajczak J, Pall P, Schär C (2013) Projections of extreme precipitation events in regional climate simulations for Europe and the Alpine region. J Geophys Res Atmos 118(9):3610–3626. https://doi.org/10.1002/jgrd.50297

Rajczak J, Kotlarski S, Schär C (2016) Does quantile mapping of simulated precipitation correct for biases in transition probabilities and spell-lengths? J Clim 29(5):1605–1615. https://doi.org/10.1175/JCLI-D-15-0162.1

Requena AI, Mediero L, Garrote L (2013) A bivariate return period based on copulas for hydrologic dam design: accounting for reservoir routing in risk estimation 17(8):3023–3038. https://doi.org/10.5194/hess-17-3023-2013. https://www.hydrol-earth-syst-sci.net/17/3023/2013/

Salvadori G, De Michele C, Durante F (2011) On the return period and design in a multivariate framework 15(11):3293–3305. https://doi.org/10.5194/hess-15-3293-2011. https://www.hydrol-earth-syst-sci.net/15/3293/2011/

Schalkoff RJ (1997) Artificial neural networks, vol 1. McGraw-Hill, New York

Seneviratne SI, Nicholls N, Easterling D, Goodess CM, Kanae S, Kossin J, Luo Y, Marengo J, McInnes K, Rahimi M, Reichstein M, Sorteberg A, Vera C, Zhang X (2012) Changes in climate extremes and their impacts on the natural physical environment. In: Field CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, Mastrandrea MD, Mach KJ, Plattner GK, Allen SK, Tignor M, Midgley PM (eds) 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 (IPCC). Cambridge University Press, Cambridge, pp 109–230

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, Powers JG (2008) A description of the advanced research WRF version 3. Tech. Rep. TN-475+STR, National Center for Atmospheric Research

Stephenson DB (2000) Use of the "odds ratio" for diagnosing forecast skill. Weather Forecast 15(2):221–232. https://doi.org/10.1175/1520-0434(2000)015<0221:UOTORF>2.0.CO;2

Stucki P, Rickli R, Brönnimann S, Martius O, Wanner H, Grebner D, Luterbacher J (2012) Weather patterns and hydro-climatological precursors of extreme floods in Switzerland since 1868. Meteorol Z 21(6):531–550. https://doi.org/10.1127/0941-2948/2012/368

Torma C, Giorgi F, Coppola E (2015) Added value of regional climate modeling over areas characterized by complex terrain-precipitation over the alps. J Geophys Res Atmos 120(9):3957–3972. https://doi.org/10.1002/2014JD022781

van Bebber WJ (1891) Die Zugstrassen der barometrischen Minima nach Bahnenkarten der Deutschen Seewarte für den Zeitraum von 1870–1890. Meteorol Z 8:361–366

Winschall A, Pfahl S, Sodemann H, Wernli H (2012) Impact of north atlantic evaporation hot spots on southern alpine heavy precipitation events. Q J R Meteorol Soc 138(666):1245–1258. https://doi.org/10.1002/qj.987

Zittis G, Bruggeman A, Camera C, Hadjinicolaou P, Lelieveld J (2017) The added value of convection permitting simulations of extreme precipitation events over the eastern mediterranean. Atmos Res 191:20–33. https://doi.org/10.1016/j.atmosres.2017.03.002

Zorita E, Storch HV (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J Clim 12:2474–2489