

1 **Event selection for dynamical downscaling: a**  
2 **neural network approach for physically-constrained**  
3 **precipitation events**

4 **J. J. Gómez-Navarro · C. C. Raible · J.**

5 **A. García-Valero · M. Messmer · J. P.**

6 **Montávez · O. Martius**

7 Received: date / Accepted: date

---

J. J. Gómez-Navarro

Department of Physics, University of Murcia, Murcia, Spain

E-mail: jgomeznava@um.es

C. C. Raible

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

E-mail: raible@climate.unibe.ch

J. A. García-Valero

Agencia Estatal de Meteorología (AEMET), Murcia, Spain

E-mail: garciavalero@aemet.es

M. Messmer

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

E-mail: messmer@climate.unibe.ch

J. P. Montávez

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

---

Department of Physics, University of Murcia, Murcia, Spain

E-mail: montavez@um.es

O. Martius

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

E-mail: olivia.romppainen@giub.unibe.ch

## 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 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

49 methods (Maraun et al 2010, 2015). For dynamical downscaling, Regional Cli-  
50 mate Models (RCM) are run over a limited spatial domain at a higher spatial  
51 resolution (e.g. Torma et al 2015; Fantini et al 2016; Giorgi et al 2016; Gómez-  
52 Navarro et al 2018, among many others). However, the computational cost  
53 of such regional simulations is still considerably high. This has motivated the  
54 development of hybrid approaches that take advantage of statistical relation-  
55 ships to extend the results drawn from short RCM simulations over longer  
56 time periods. Martinez et al (2012) developed a statistical-dynamical down-  
57 scaling procedure that relies on Empirical Orthogonal Function analysis to  
58 generate large-scale atmospheric patterns, which are then dynamically down-  
59 scaled. This allowed the researchers to construct regional time series, and was  
60 successfully used to generate realistic regional series of wind with 6-hour res-  
61 olution. Li et al (2011) used a limited number of existing RCM simulations  
62 over North America to fit a linear regression model between the RCM output  
63 and the driving ESM fields. This statistical model was then used, together  
64 with a large set of ESM simulations, to produce a probabilistic projection of  
65 high-resolution temperature change in North America, which even allowed the  
66 researchers to quantify the different sources of uncertainty.

67 Within the RCM community, large on-going initiatives such as The Co-  
68 ordinated Regional Climate Downscaling Experiment (CORDEX) have been  
69 formed to coordinate the computational effort, therefore facilitating and max-  
70 imising the exchange of information derived from these costly simulations.  
71 Nowadays, the resolution in most RCM simulations is about 10 km (e.g.,

72 <https://guidelines.euro-cordex.net>). This resolution is sufficient to demonstrate  
73 the added value of RCM compared to ESM simulations, especially regarding  
74 precipitation processes (Torma et al 2015; Fantini et al 2016; Bowden et al  
75 2016). However, there exists added value of going beyond 10 km, entering the  
76 scale of convection permitting simulations (Ban et al 2014; Giorgi et al 2016;  
77 Zittis et al 2017; Gómez-Navarro et al 2018), in particular in areas of com-  
78 plex topography and during extreme precipitation events (Giorgi et al 2016;  
79 Chan et al 2017). Still, the high computational cost is the bottleneck that has  
80 limited the number of simulations currently available of this nature, and it  
81 may become an unavoidable limitation precluding the RCM community from  
82 taking full advantage of the new CMIP6 ensemble.

83 To overcome this limitation, some researchers have proposed to make a  
84 previous selection of dates to be simulated to avoid the computational cost of  
85 running transient climate simulations. Meredith et al (2018) presented a classi-  
86 fication algorithm based on geopotential height as a mean to select dates with  
87 an elevated potential for extreme precipitation in a narrow river catchment.  
88 This approach enables a clever selection of events to be dynamically down-  
89 scaled that discards situations of fewer interest, and so high-resolution RCM  
90 simulations can be selectively performed saving important computational re-  
91 sources. Felder et al (2018) aimed at simulating worst-case events using a range  
92 of computational models across spatial scales, from an ESM to a damage and  
93 loss model reaching the scale of individual buildings. To keep the physical  
94 consistency among models, the research team selected events within the ESM,

95 using the extreme precipitation (averaged over Switzerland) as criterion and  
96 dynamically downscaled these events to 2 km. The analysis of the downscaled  
97 events showed that this criterion leads to unsatisfactory results, and suggested  
98 that any refinement of the approach shall include more variables from the ESM  
99 used as predictors. Chan et al (2017) selected three large-scale predictors of  
100 extreme precipitation: Mean Sea Level Pressure, 850-hPa relative vorticity and  
101 static stability. They used regression analysis to identify large-scale precursors  
102 of extreme precipitation events in convection-permitting climate simulations,  
103 and found that indeed these three variables have skill in predicting precipita-  
104 tion extremes in simulations both at 12 and 1.5 km spatial resolution.

105 In general, good predictor variables should include the main processes con-  
106 tributing to heavy precipitation on a scale that is captured by the ESM. Heavy  
107 and extreme precipitation requires a steady supply of moisture and a lifting  
108 mechanism that brings the moist air to saturation (e.g., Doswell et al 1996). In  
109 Switzerland moisture is transported towards the Alps from the south prior to  
110 and during regional-scale heavy precipitation events on the Alpine south side  
111 (Martius et al 2006; Winschall et al 2012). During regional-scale heavy precip-  
112 itation on the Alpine north side, sustained and intensive moisture transport  
113 against the orography can occur from the east during Vb weather situations,  
114 from the north and from the west (Piaget et al 2015; Giannakaki and Mar-  
115 tius 2016; Froidevaux and Martius 2016; Messmer et al 2017). The moist air  
116 masses reaching Switzerland are lifted within warm convey belts (Pfahl et al

117 2014), along the orography (Giannakaki and Martius 2016), or in areas of flow  
118 convergence (Giannakaki and Martius 2016).

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

## 126 **2 Data and Methods**

### 127 2.1 Community Earth System Model (CESM)

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

135 We use data from two CESM simulations: i) a 400-year simulation with  
136 perpetual AD 850 conditions, hereafter referred as CESM-control simulation,  
137 ii) a seamless 850-2099 AD simulation driven by reconstructions of external

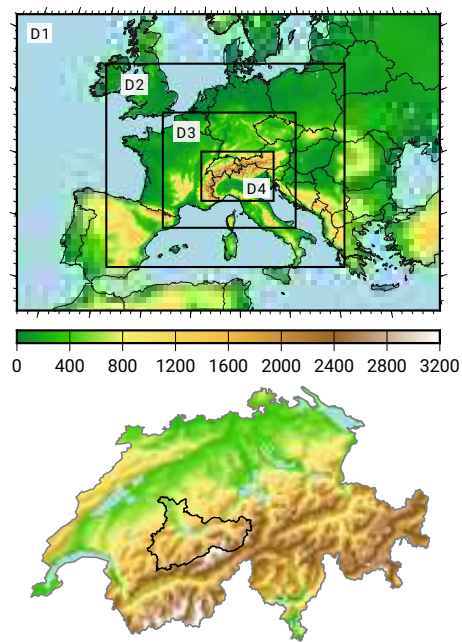
138 forcings for the historical period 850-2005 and RCP8.5 forcing for the future  
139 period 2006-2099 (Lehner et al 2015). In this study, we use this data but  
140 consider it as split in two periods: 1850-2005, hereafter referred as CESM-  
141 historical simulation, and 2006-2099, hereafter CESM-future simulation.

## 142 2.2 Weather Research and Forecasting Model (WRF)

143 The dynamical downscaling of the CESM simulations is performed with the  
144 Weather Research and Forecasting model (WRF), version 3.5 (Skamarock et al  
145 2008). We use a setup with 4 nested domains reaching a spatial resolution of  
146 2 km in its innermost domain spanning the Alpine region entirely (see Fig.  
147 1). This high resolution allows us to explicitly simulate convective processes,  
148 which is of foremost importance in extreme event phenomena, precisely those  
149 that this study tackles (Ban et al 2014; Giorgi et al 2016). A comprehensive  
150 description of the details of these simulations, as well as an evaluation of the  
151 model performance in the particular configuration employed in this study is  
152 presented in Gómez-Navarro et al (2018).

153 The simulations performed with WRF include i) a transient present-climate  
154 simulation that continuously spans the period 1979-2005 and is driven by the  
155 CESM-historical simulation, hereafter referred as WRF-reference period; ii) a  
156 transient simulation over the period 2080-2099 nested to CESM-future simu-  
157 lation, hereafter WRF-future; iii) a number of single-day case studies selected  
158 from the CESM-control, CESM-historical and CESM-future simulations. The





**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.

159 criteria for the selection of dates is described below, and in each case a short  
160 spinup of 12 hours is used.

### 161 2.3 Artificial Neural Networks as Statistical Downscaling Tool

162 An ANN is a mathematical model that acts as a function, relating certain  $n$ -  
163 dimensional input vectors to  $m$ -dimensional output vectors (Schalkoff 1997).  
164 This model is not new in meteorological applications. Dawson and Wilby  
165 (1998) proposed a novel rainfall-runoff model based on ANNs, and used it  
166 to forecast the river flow in two different UK catchments with a skill compara-

167 ble to operational systems. Lee et al (1998) used an ANN to build a model to  
168 forecast precipitation in Switzerland, a region characterised by complex orog-  
169 raphy. ANNs have also found early applications as downscaling technique in  
170 a very similar manner as we aim here (Zorita and Storch 1999). The reader  
171 is referred to the former references for a more comprehensive explanation of  
172 the algorithm, as we just briefly outline its most important aspects in the  
173 following.

174 An ANN is composed of various layers, each of which contains so-called  
175 neurons, that can be regarded as computation units. A network contains at  
176 least an input layer (with  $n$  neurons, the dimension of the input vector) and an  
177 output layer (with  $m$  neurons, the dimension of the output vector). In between,  
178 there can be a number of so-called hidden layers, each of which including  
179 a variable number of neurons. Each neuron is connected to all neurons in  
180 the following layers through connections that are characterised by a weight.  
181 Given an input, the network calculates a unique output that results from a  
182 relatively simple and therefore computationally inexpensive calculation that  
183 involves the input vector and all the weights, which act as the parameters of  
184 the model. Therefore, the calibration of the network consists of finding the set  
185 of weights that optimise the output of the network for a given metric. We use  
186 supervised learning, in which the training of the network consists of using a  
187 number of input-output couples from the reference period, i.e. the predictors  
188 and the predictand, to find the set of weights that minimise the difference  
189 between the transient RCM and the output of the ANN, averaged for the

190 whole pool of samples that conform the reference period. The search of an  
191 optimal solution is computationally moderately demanding, and is based on  
192 the backward propagation algorithm, that can be viewed as an application of  
193 the chain rule in differentiation (Schalkoff 1997).

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

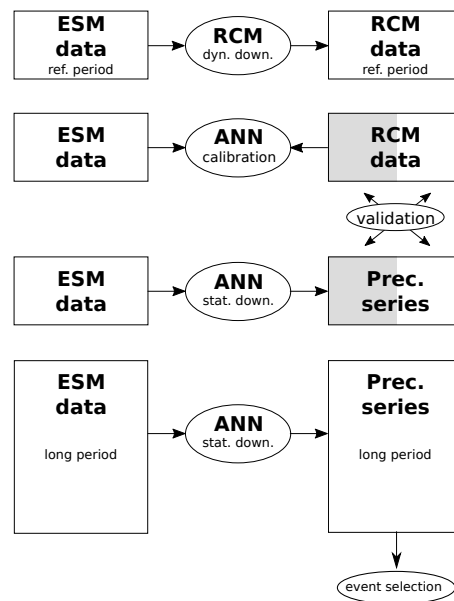
203 Finally, we use the ANN from the R package neuralnet ([https://CRAN.R-](https://CRAN.R-project.org/package=neuralnet)  
204 [project.org/package=neuralnet](https://CRAN.R-project.org/package=neuralnet)). The geometry of the network and the num-  
205 ber of variables used as input predictors for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

## 208 2.4 Strategy for the Selection of Events

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

- 211 1. The coarse fields from an ESM are dynamically downscaled with an RCM.  
212 This is a computationally demanding step, so this reference period is in-  
213 admissibly short for most climate applications.
- 214 2. An ANN is trained to learn the relationship between the large-scale vari-  
215 ables in the ESM and the daily precipitation simulated by the RCM in an  
216 orographically complex region.
- 217 3. The calibration and validation periods are swapped to carefully assess the  
218 performance of the ANN with independent data during the reference pe-  
219 riod.
- 220 4. Finally, the calibrated ANN can be used to statistically downscale a longer  
221 ESM simulation. This way, the ANN tries to emulate the series we would  
222 obtain running the RCM for the longer period in case it would be com-  
223 putationally feasible. This series shall be used in the selection of dates  
224 candidates to extreme events.

225 To demonstrate the feasibility of this approach, we implement it in the  
226 following sections. We use the WRF-reference simulation comprehensively de-  
227 scribed in Gómez-Navarro et al (2018). The arguments for the selection of  
228 predictors are presented in Sec. 3.1, followed by a range of tests that allow  
229 to determine the optimal ANN geometry in Sec. 3.2. A comprehensive cali-  
230 bration using various statistics associated to the forecast of extreme events is  
231 presented in Sec. 3.3. Finally, in Sec. 4 we apply the obtained ANN during a  
232 perturbed climate period the ANN was not calibrated for. This allows us to



**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.

233 assess the limitations derived from the stationary hypothesis implicit in many

234 statistical downscaling exercises.

---

### 235 **3 Event selection based on ANN during the reference period**

#### 236 **3.1 Large scale variables – potential predictors of extreme events**

237 Similarly to Chan et al (2017), we first gain insight on how local extreme  
238 events in precipitation over the Aare catchment (bottom panel in Fig. 1) are  
239 connected to the large scale atmospheric dynamics in the WRF-CESM cou-  
240 pled system. For this, we use a composite analysis applied to different variables  
241 using the 1-day extreme precipitation events, defined here as those days ex-  
242 ceeding the 95th percentile of daily precipitation and applied to each season  
243 separately. All days in the WRF-reference simulation are filtered out accord-  
244 ing to this criterion and then averaged. The selection of the variables is based  
245 on previous studies (Martius et al 2006; Winschall et al 2012; Messmer et al  
246 2017): sea level pressure (SLP), geopotential height at 850 and 500 hPa (Z850  
247 and Z500), integrated water vapour and vapour transport (IWV and IWVT,  
248 respectively) and precipitation (PREC). The annual cycle is removed from  
249 each variable to obtain anomalies. Note that one limitation of this analysis  
250 is that these composites reflect the mean large-scale flow patterns. Thus, for  
251 individual events the large-scale flow and the moisture transport can substan-  
252 tially deviate from these composites as presented by Giannakaki and Martius  
253 (2016) who identified several relevant flow patterns associated with heavy pre-  
254 cipitation events in Northern Switzerland.

255 Composites for selected variables for winter are presented in Fig. 3. Ex-  
256 treme daily precipitation events are related to a west-east oriented pressure

257 anomaly dipole with high pressure centred between the Iberian Peninsula and  
258 the Bay of Biscay and low pressure over eastern Europe (Fig. 3a). This pres-  
259 sure dipole has a barotropic vertical structure (Fig. 3b). The height thickness  
260 between 500-hPa and 850-hPa, a measure of temperature anomalies in this  
261 layer, indicates a dipole with warmer air over western and colder air over east-  
262 ern Europe (Fig. 3c). The strongest pressure gradient is located over Germany  
263 and Switzerland and suggests strong north to northwesterly winds over this  
264 region. Consistently extreme daily precipitation events in winter are associ-  
265 ated with a strong north to northwesterly integrated water vapour transport  
266 (Fig. 3e). As expected, CESM generates a positive precipitation anomaly over  
267 Switzerland when sampling over extreme precipitation events in the WRF  
268 simulation (Fig. 3d). The autumn patterns (not shown) resemble the winter  
269 ones while the spring ones are rotated counter clockwise by 10 degrees with  
270 respect to the winter pattern for all but the precipitation pattern (therefore  
271 not shown).

272 The corresponding composites for summer are shown in Fig. 4. Extreme  
273 daily precipitation events are associated with a surface low pressure system  
274 centred over Austria and the Czech Republic. At 500 hPa, two low pressure  
275 minima are found. A stronger one located over northern France and a weaker  
276 one over south-eastern Europe. The thickness between 500-hPa and 850-hPa  
277 shows warm air over western Europe, whereas over the British Isles and eastern  
278 Europe a cold anomaly is present (Fig. 4c). The moisture fluxes over Switzer-  
279 land are weaker than in winter, and the main source of humidity in these

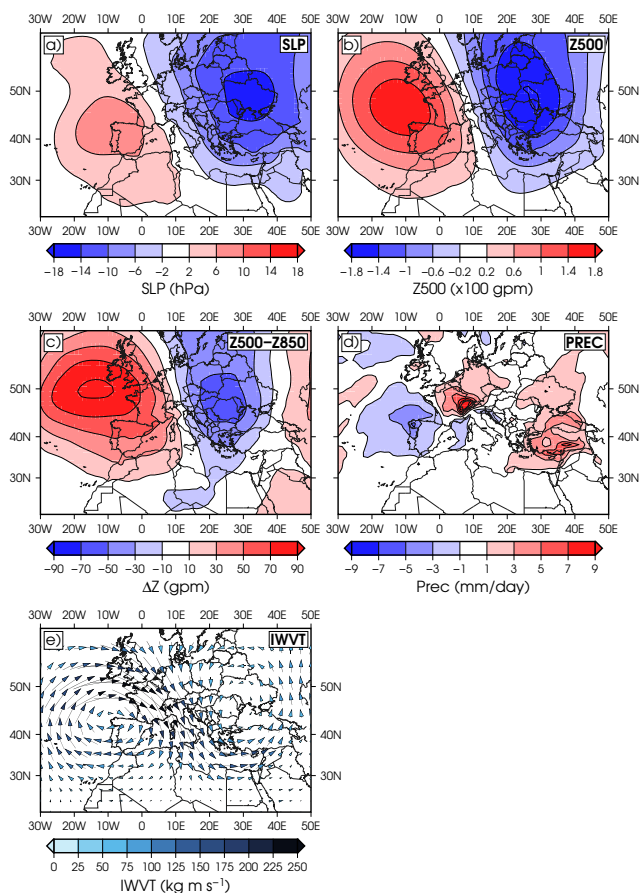
280 situations is the Mediterranean Sea, in good agreement with results of Mess-  
281 mer et al (2017). A positive precipitation anomaly is found over Switzerland  
282 in the ESM (Fig. 4d), which again shows the link between the precipitation  
283 simulated by the RCM and its driving dataset, i.e. the relationship exploited  
284 by Felder et al (2018).

285 In summary, we find that extreme precipitation events identified in the  
286 WRF simulation are related to large scale circulation patterns of the driving  
287 CESM simulation. The flow patterns vary depending on the season. In winter  
288 (and similarly for autumn and spring) a west east dipole pattern with low  
289 pressure at the east becomes an important predictor of extreme events. In  
290 summer a low pressure system centred over Austria and the Czech Republic  
291 indicates a so-called Vb-cyclone situation (e.g., van Bebber 1891; Stucki et al  
292 2012; Messmer et al 2015, 2017). This information is used below to define  
293 meteorological indices that are exploited by the ANN.

### 294 3.2 Network geometry and predictor variables

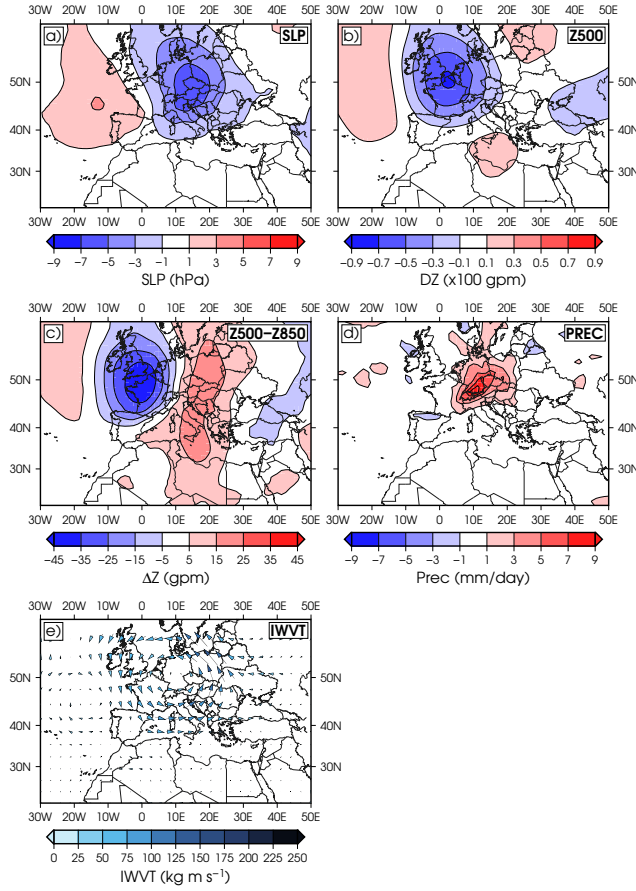
295 Once the variables candidate to be used as predictors are identified, we need to  
296 determine the geometry of the ANN, which includes the number of variables  
297 considered, but also the number and size of neurons in the hidden layers.  
298 The geometry of an ANN used in this study consists of an input, an output  
299 and a single hidden layer. The output layer of the model is the predictand,  
300 and consists of a single number, i.e. the daily mean precipitation in the Aare  
301 catchment. Thus, we need a single neuron in the output layer. The input layer





**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-hPa and 850-hPa geopotential height, (d) precipitation and (e) water vapour transport vertically integrated up to 700 hPa.

302 contains the predictors, and has as many neurons as the number of variables  
 303 considered for the downscaling. For the sake of simplicity, we set one single  
 304 hidden layer, whose size is variable. Its number of neurons is not determined  
 305 arbitrarily, but exhaustive tests are carried out to identify in each season the



**Fig. 4** As Fig. 3 but for summer, i.e. JJA months.

306 optimal number of neurons of the hidden layer that yields the highest skill of  
 307 the ANN (steps 1 and 2 in Fig. 5).

308 A number of variables obtained from the ESM are considered as input for  
 309 the network. The full list is shown in Table 1, and is based on previous lit-  
 310 erature (e.g., Martius et al 2006; Chan et al 2017) as well as the composite  
 311 analysis described in Sec. 3.1 (note that we use different variables according to  
 312 each season). Although in principle all variables are considered by the model

313 to avoid prejudices and gain in generality, a first test allows to establish which  
314 variables seem more closely related to the output variable, i.e., precipitation  
315 over the Aare catchment (see diagram labelled 1 in Fig. 5). To do so, each vari-  
316 able is used separately to build a simple ANN where it is the only input, and  
317 a single neuron exists in the hidden layer. In each case, the ANN is calibrated  
318 using half of the reference period, i.e. 1979-1992, and the fitted ANN is used  
319 to produce a prediction for the other half, i.e. 1993-2005. This is then com-  
320 pared to the expected output, i.e. the dynamically downscaled precipitation,  
321 and their mutual correlation is calculated, which serves as metric to build the  
322 aforementioned ranking of the variables. In this step, other metrics than cor-  
323 relation were considered, such as Root Mean Square Error or Mean Averaged  
324 Bias. The results indicate modest sensitivity to the choice of the metric, so  
325 correlation was finally the metric used.

326 In a second step (see diagram labelled 2 in Fig. 5), the ranking is used as  
327 the base to calibrate more complex ANNs that include a growing number of  
328 variables and their interactions, and allow to determine the optimal geometry.  
329 To do so, we loop in the variables in the order defined by the previous ranking,  
330 including them in a growing set of predictors. In each step  $i$  of this loop,  
331 another loop is considered in  $j$ , the number of neurons in the hidden layer,  
332 which varies between 1 and  $i$ . Therefore, a total of  $\frac{N^2+N}{2}$  ANNs, where  $N$  is  
333 the total number of variables in Table 1, are tested for each season. As before,  
334 the period 1979-1992 is used for calibration and 1993-2005 for validation. From  
335 all the combinations of number of variables and neurons, the optimal one, in

336 the sense of maximising the correlation, is chosen as ultimate ANN geometry  
337 to perform the downscaling. Note that the role of calibration and validation  
338 periods is exchanged to complete the full reference period, allowing the cross  
339 validation of the results without circularity.

340 These tests result in a distinct combination of predictors and number of  
341 neurons in the hidden layer for each season separately. These configurations, to-  
342 gether with the correlations obtained during the validation are shown in Table  
343 2. Note that the number of variables does not grow monotonically, but reaches  
344 an optimal number for each season. Similarly, the inclusion of more neurons  
345 in the hidden layer does not necessarily improve the ANN. As expected, the  
346 precipitation of the ESM emerges as an important variable predicting precip-  
347 itation in the target region, but in all cases the addition of further large-scale  
348 predictors improves the performance of the ANN. Indeed, in autumn zonal  
349 wind at 850 hPa is more skilful predicting local precipitation than the precip-  
350 itation of the ESM. In agreement with the results in Sec. 3.1, the variables  
351 most relevant for DJF, MAM and SON are similar, being wind at various levels  
352 very illuminating predictors of extreme events. This is indicative of the close  
353 relationship between extreme events in this region and the moisture provided  
354 by the large-scale circulation. Summer again stands out as a distinct season,  
355 where the moisture provided by the Mediterranean Sea associated to low pres-  
356 sure centres in central Europe (note the presence of IWVTXFRANCE and  
357 SLPCZECH, in the most relevant variables for JJA in Table 2), as well as in-

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

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 Sec. 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.

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

**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, SLP-GRAD, 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

358 stability measures (KI), emerge as important key variables to predict summer  
 359 precipitation in this region.

### 360 3.3 Results of the event selection in the reference period

361 To assess the performance of the event selection based on ANNs, we compare  
 362 it with the skill achieved by the simple approach employed by Felder et al  
 363 (2018). They used only extreme precipitation in the CESM-control simulation  
 364 averaged over the Switzerland (bottom map of Fig. 1) as a criterion to identify  
 365 dates which potentially deliver extreme precipitation over the Aare catchment.  
 366 The selected dates correspond to the four most extreme precipitation events

367 for each season in CESM-control, i.e. 16 cases in total, and were dynamically  
368 downscaled with WRF. Their results demonstrated that the precipitation ob-  
369 tained for these cases in winter and summer over the target region was far  
370 lower than expected, and motivated the new method presented in this study.  
371 The results by Felder et al (2018) are extended here to the four seasons and  
372 analysed in more detail. The downscaled precipitation in all the selected cases  
373 is relatively large, exceeding all but one case the 90th percentile (see Table 3).  
374 Still, the precipitation obtained for these events is lower than the maximum  
375 precipitation during the reference period in nearly all cases. Given the length  
376 of the CESM-control simulation that was used to search for extremes, the se-  
377 lection of dates aimed at providing physically consistent precipitation events  
378 with return periods of up to 400 years. However, the method failed in this  
379 regard. Only 1 event in summer leads to precipitation that can be considered  
380 as extreme in a 400-year frame, whereas for 10 out of 15 cases the estimated  
381 return period is below 10 years.

382 The skill of event selection based on precipitation within the ESM alone  
383 can be further assessed using the whole WRF-reference simulation. This way,  
384 we can compute various skill metrics between the precipitation predicted by  
385 the ESM over the target region and the one simulated by WRF during the  
386 complete reference period 1979-2005 (left column of Fig. 6). The precipitation  
387 within the ESM alone is a poor predictor of precipitation at local scale. Con-  
388 sidering all days, correlations are around 0.4 in all seasons (note that the Hit  
389 Rate trivially converges to 1 at low percentiles by its very definition). How-

	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

**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. 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).



390 ever, when higher percentiles are considered, the skill rapidly deteriorates in  
391 all seasons. Above the 90th percentile, the correlation ranges between 0.3 (in  
392 winter) and 0.1 (in summer), and the Hit Rate is about 0.3 in all seasons.  
393 For percentiles above the 99th percentile, the Hit Rate rapidly approaches  
394 zero and becomes insignificantly different from random chance in all seasons.  
395 SEDI is stable through most of the percentiles, as it is expected for the prop-  
396 erties that define this index (Ferro and Stephenson 2011). Again it shows that  
397 the selection of events has certain skill in moderate percentiles, but it rapidly  
398 deteriorates towards rarer events.

399 We focus now on the approach based on ANNs. The optimal geometry  
400 for each season is fixed as indicated in Table 2 and determined by the tests  
401 in Sec. 3.2. Next, we apply the approach represented by the diagram 3 in  
402 Fig. 5. This is, an ANN for each season is calibrated during the first half of  
403 the WRF-reference period, and used to forecast the daily precipitation in the  
404 Aare catchment during the second half. These periods are exchanged to en-  
405 sure that the full period is statistically downscaled using independent data for  
406 the calibration. The validation, based on the comparison between the WRF  
407 output and the optimal ANN for each season, is summarised in the results  
408 of the middle column of Fig. 6. Correlations are systematically higher than  
409 those for the more simple method in all seasons and percentiles. Considering  
410 all days, correlation is about 0.8 in spring and autumn, with slightly higher  
411 and lower values in winter and summer, respectively. As before, the skill de-  
412 creases towards higher percentiles, but in clear contrast to the performance

413 of the simple approach by Felder et al (2018), in all seasons but summer the  
414 correlation is nearly constant up to the highest percentiles. The results are  
415 worse in summer, where the ANN monotonically decreases its performance  
416 towards higher percentiles. Still, in this season the ANN demonstrates valu-  
417 able skill up to the 99th percentile, where correlation reaches a critical value  
418 of 0.3. This is again in contrast to the simple approach (left column of Fig.  
419 6) showing no skill above the 90th percentile. The Hit Rate also shows the  
420 seasonal differences. The ANN's ability to select the right dates is best in win-  
421 ter, with probabilities of detecting an event that actually happen above 0.6 in  
422 percentiles above 99th, and worst in summer, when it drops to 0.1. But even  
423 in that case, and unlike in the simpler approach, the ANN is able to capture  
424 the 99th percentile events better than pure chance. In both, spring and au-  
425 tumn, there is a somewhat unexpected improvement in the metrics beyond  
426 the 90th percentile that is provoked by the ability of the ANN to capture the  
427 precise ordering of the most extreme cases (not shown), and implies that the  
428 ANN is able to adequately select between 60% and 80% of the extreme events  
429 in these seasons. SEDI remains remarkably stable, and even grows towards  
430 higher percentiles in all seasons but summer. The values are systematically  
431 higher across all percentiles than those of the simple approach, and the skill  
432 is comparable to that obtained for the 12h to 24h precipitation forecasts of  
433 the ECMWF, where the seasonal cycle in the forecasting performance is also  
434 observed (North et al 2013).

435 The ability of the ANN to predict the correct extreme events during the  
436 reference period is further evaluated by comparing the magnitude of the precip-  
437 itation in observed and predicted extremes. This is presented through boxplots  
438 in Fig. 7. This figure shows, for each season, the distribution of precipitation  
439 in the days when it is observed to be above the 99th percentile (black), and  
440 compares it to the distribution when the days are those predicted by the ANN  
441 (blue). First, we note that the left tail of the distribution is longer in the pre-  
442 dicted events (compare black and blue boxplots across seasons). This is the  
443 expected behaviour, as we already know from the analysis above that about  
444 50% of the events are incorrectly attributed to extreme events in this per-  
445 centile. Still, in all seasons but summer the median is well captured, as well as  
446 the right tail of the distribution, i.e. the absolute most extreme events during  
447 the period. In agreement with the analysis above, in summer the predictive  
448 skill is the lowest. Only around 10% of the cases are correctly predicted (see  
449 Hit Rate in Fig. 6) and these correspond to two marked outliers. In this re-  
450 gard, it noteworthy that all the absolute maxima for each season during the  
451 reference period have been captured by the ANNs.

#### 452 **4 Detection of extremes in climate change projections**

453 For the ANNs to be successful, and more generally for any statistical downscal-  
454 ing tool, the climate must be stationary between the period used to calibrate  
455 and the one where the model is applied. But the actual climate is not com-  
456 pletely stationary, which can limit the skill of these methods. We have taken

457 the opportunity provided by the WRF-future simulation to evaluate the per-  
458 formance of statistical downscaling under the unfavourable circumstances of a  
459 climate severely perturbed with respect to the one used to calibrate the model.

460 In detail, the optimal ANN calibrated during the WRF-reference period  
461 has been used to downscale the CESM-future simulation during the period  
462 2080-2099. This prediction is compared with the WRF-future simulation fol-  
463 lowing the same protocol than in the Sec. 3.3. The result of this analysis is  
464 presented in the right column of Fig. 6. Compared to the reference period, the  
465 ANN generally presents lower predictive skill. In winter, correlation ranges  
466 between 0.75 and 0.6, whereas it is above 0.8 in the reference period. The  
467 Hit Rate is very similar up to the 80th percentile, but above this point the  
468 probability of predicting extreme events drops faster, reaching 0.4 for the 99th  
469 percentile (compare this with the value about 0.7 during the reference pe-  
470 riod). This worsening of the skill is also visible in lower SEDI values in the  
471 highest percentiles. A similar behaviour is found in spring, albeit with slightly  
472 lower correlations, closely resembling the results during the reference period.  
473 The Hit Rate is very similar in both periods until the 90th percentile, where  
474 the increase of the Hit Rate observed in the reference period is absent in the  
475 future period, thus also SEDI is reduced. In summer, the ANN presents the  
476 lowest correlations, but they are remarkably similar to those during the ref-  
477 erence period, showing that the reduction of skill of the ANN under future  
478 climate conditions is not very pronounced for summer. Indeed, above 80th  
479 percentile the Hit Rate is higher under future climate conditions, leading to

480 values comparable to other seasons. In autumn, correlations are strongly re-  
481 duced compared to the reference period, and are close to those for summer.  
482 This difference is also found in the Hit Rate above 80th percentile, which leads  
483 to the lowest Hit Rate of all seasons in the future simulation, and the strongest  
484 reduction of SEDI. This is in clear contrast to the behaviour of the reference  
485 period.

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

499 Fig. 7 represents the distribution of extremes observed (orange) and pre-  
500 dicted by the ANNs (red) under climate change conditions. As with the ref-  
501 erence period, the longer left tails are expected as a consequence of the non-  
502 perfect Hit Rate. Unlike in the reference period, a systematic bias stands out,

503 i.e., a general underestimation of precipitation, illustrated by the fact that the  
504 median for the ANN-selected cases is below the 25th percentile for the WRF-  
505 future simulation. Regarding the most extreme events, in winter and spring  
506 the ANNs are able to predict the most extreme events, whereas in summer  
507 and autumn the two most extreme events are not identified with the ANNs,  
508 respectively.

509 The signal of climate change on extreme events should be sought in a  
510 shift in the blue and red boxplots in Fig. 7, respectively. However, instead  
511 of a systematic shift towards more severe extremes, which could be expected  
512 according to other studies (e.g. Seneviratne et al 2012; Rajczak et al 2013,  
513 2016; Messmer et al 2017) as well as basic thermodynamic relationships (e.g.  
514 O’Gorman 2015), we notice a rather stationary behaviour with modest changes  
515 attributable to sampling uncertainty. Further, the analysis of the distribution  
516 of these extremes in consecutive periods shows a lack of trend towards more  
517 severe or frequent extremes (not shown). Therefore, the ANNs suggest a lack  
518 of sensitivity of extreme precipitation events to climate change in all seasons.

519 This contrasts with the results we can draw by comparing the WRF-  
520 reference and WRF-future runs (see the horizontal shift in the black vs. orange  
521 boxplots in Fig. 7). Using the full transient runs, we find more severe extremes  
522 in winter, summer and to a lesser extent in autumn, with a strong opposite  
523 behaviour in spring. This behaviour is concurrent with an overall increase in  
524 precipitation not only in the extremes but also on average in these seasons  
525 (not shown), which better agrees with the sensitivity of extremes to a warm-

526 ing climate reported for the aforementioned studies. Similar conclusions can  
527 be drawn from the application of Extreme Value Theory to these data. Fig.  
528 8 depicts the return level plots for daily precipitation in each season in the  
529 WRF-reference and WRF-future simulations. In winter and summer, the re-  
530 turn levels are systematically higher in the WRF-future simulation than in  
531 WRF-reference. The lower climate change signal described above for autumn  
532 can be understood under the light of this analysis as a mixed behaviour be-  
533 tween the events with return levels below and above 5 years. In contrast to  
534 winter and autumn, the events with longer return periods are ameliorated by  
535 climate change in autumn according to these simulations, although the uncer-  
536 tainty in this range is large due to the modest number of events that support  
537 this conclusion. Finally, as described above respect to Fig. 7, spring stands out  
538 in the return level plot as an anomalous season, where climate change seems  
539 to reduce the occurrence of extreme events.

540 Based on the discussion above, the apparent lack of sensitivity to climate  
541 change identified by the ANNs has to be attributed, at least in all seasons  
542 but autumn, to a limitation of statistical downscaling regarding its ability  
543 to identify extremes under perturbed climate conditions, rather than to an  
544 outcome of the CESM-WRF simulations. This can introduce subtle and non-  
545 systematic biases that largely affect the study of trends, as the comparison of  
546 the two transient simulations demonstrates.

---

## 547 **5 Conclusions**

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

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



569 Meredith et al (2018), our method relies on existing RCM simulations rather  
570 than observations.

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

585 Finally, we use an existing high-resolution climate change projection to  
586 evaluate the sensitivity of the method to the non-stationarity of actual cli-  
587 mate. The ANNs trained during the reference period have been tested under  
588 the RCP 8.5 scenario, searching for events above the 99th percentile using  
589 the CESM-future simulation as input for the statistical model, and comparing  
590 the output with WRF-future. The events selected after the application of the  
591 ANNs are overall extreme, with the majority of events above the expected

---

592 percentile. Still, the performance of the ANN method is lower than during the  
593 reference period, which can be attributed to the fact that the model exploits  
594 relationships between variables learnt for different climate conditions. A num-  
595 ber of events are erroneously identified as severe, although this is expected and  
596 agrees with the Hit Rates obtained during the reference period, and demon-  
597 strates the ability of the ANNs to predict extreme events with remarkable  
598 performance even in climate conditions very different to the ones used to cali-  
599 brate the model. The analysis of the response of the severity of these events to  
600 climate change evidences no trend. This could be erroneously interpreted as a  
601 lack of sensitivity of extreme precipitation under climate change in the simu-  
602 lations, which would be in contradiction with other studies (e.g. Seneviratne  
603 et al 2012; Rajczak et al 2013). Indeed, the comparison of the transient simu-  
604 lations WRF-reference and WRF-future rules out this possibility, and instead  
605 hints to a limitation of the ANNs to capture the thermodynamic mechanisms  
606 responsible for this trend. Thus, these results suggest that statistical meth-  
607 ods like ANN, which rely on the assumption of stationarity of the statistical  
608 relationships for reference and change climate states, may be unsuitable for  
609 correctly identifying trends.

---

## 610 **6 Appendix: Skill metrics**

### 611 **6.1 Correlation**

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

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

631 by independently bootstrapping both series with repetition (dashed curves in  
632 Fig. 6).

## 633 6.2 Hit Rate $F$

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

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

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

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

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

648 A detail to be determined is how to define whether an event happened  
649 or not in either the observations or the predicted dataset. For instance, if a

650 given threshold of precipitation is fixed for both datasets, it might be that the  
651 total number of events above such threshold differs between the two datasets,  
652 leading to a systematic bias, defined as:

$$B = \frac{a + b}{a + c}. \quad (3)$$

653 Values of  $B$  other than 1 indicate a systematic bias between the observations  
654 and the predicted dataset. However, this bias is meaningless to us, as we are  
655 not interested in the given values of precipitation provided directly by the  
656 ANN, but in their ranking of most extreme values, which will be ultimately  
657 used to select the events to be downscaled dynamically. Therefore, we carry  
658 out a form of hedging to the data that consists of working with quantiles. This  
659 is, for a given a percentile  $p$ , we obtain the corresponding quantiles separately  
660 for the statistical and dynamical downscaling series (as they are in general  
661 different if the ANN is biased). Then, we define that an event happened in  
662 one of the series when the precipitation in a given day is above its respective  
663 quantile. Summing the number of events, leads to the numbers  $a$ ,  $b$ ,  $c$  and  $d$   
664 of the contingency table, which ultimately determines  $H$  for a given percentile  
665  $p$ . As describe above, this calculation is repeated for  $p$  ranging between 1 and  
666 99.

### 667 6.3 Symmetric Extremal Dependence Index

668 The Symmetric Extremal Dependence Index (SEDI) was proposed by Ferro  
669 and Stephenson (2011) as an alternative metric to evaluate the skill in predict-

ing 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 proce-  
dure as for  $H$ .

**Acknowledgements** We thank the Mobiliar Lab for Natural Risks for supporting the  
study. The simulations are performed on the supercomputing architecture of the Swiss Na-  
tional Supercomputing Centre (CSCS). JJGN acknowledges the CARM for the funding  
provided both through the Seneca Foundation (project 20022/SF/16), as well as the "Juan  
de la Cierva-Incorporación" program (IJCI-2015-26914). MM acknowledges support by the  
SNF (Early Postdoc.Mobility).

## References

- Ban N, Schmidli J, Schär C (2014) Evaluation of the convection-resolving regional climate  
modeling approach in decade-long simulations. *Journal of Geophysical Research: Atmo-  
spheres* 119(13):7889–7907, DOI 10.1002/2014JD021478
- van Bebbber WJ (1891) Die Zugstrassen der barometrischen Minima nach Bahnenkarten der  
Deutschen Seewarte für den Zeitraum von 1870-1890. *Meteorol Zeitschr* 8:361–366
- Bowden JH, Talgo KD, Spero TL, Nolte CG (2016) Assessing the added value of dy-  
namical downscaling using the standardized precipitation index 2016:1–14, DOI  
10.1155/2016/8432064, URL <http://www.hindawi.com/journals/amete/2016/8432064/>

- 692 Chan SC, Kendon EJ, Roberts N, Blenkinsop S, Fowler HJ (2017) Large-scale predictors  
693 for extreme hourly precipitation events in convection-permitting climate simulations  
694 31(6):2115–2131, DOI 10.1175/JCLI-D-17-0404.1
- 695 Dawson CW, Wilby R (1998) An artificial neural network approach to rainfall-runoff mod-  
696 elling. *Hydrological Sciences Journal* 43(1):47–66, DOI 10.1080/02626669809492102,  
697 URL <https://doi.org/10.1080/02626669809492102>
- 698 Doswell CA, Brooks HE, Maddox RA (1996) Flash flood forecasting: An ingredients-  
699 based methodology. *Weather and Forecasting* 11(4):560–581, DOI Doi 10.1175/1520-  
700 0434(1996)011;0560:Fffaib;2.0.Co;2, URL ;Go to ISI;://WOS:A1996VZ46100009
- 701 Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016) Overview  
702 of the coupled model intercomparison project phase 6 (CMIP6) experimental design and  
703 organization 9(5):1937–1958, DOI 10.5194/gmd-9-1937-2016
- 704 Fantini A, Raffaele F, Torma C, Bacer S, Coppola E, Giorgi F, Ahrens B, Dubois C, Sanchez  
705 E, Verdecchia M (2016) Assessment of multiple daily precipitation statistics in ERA-  
706 interim driven med-CORDEX and EURO-CORDEX experiments against high resolu-  
707 tion observations pp 1–24, DOI 10.1007/s00382-016-3453-4
- 708 Felder G, Gómez-Navarro JJ, Zischg AP, Raible CC, Röthlisberger V, Bozhinova D, Martius  
709 O, Weingartner R (2018) From global circulation to local flood loss: Coupling models  
710 across the scales 635:1225–1239, DOI 10.1016/j.scitotenv.2018.04.170
- 711 Ferro CAT, Stephenson DB (2011) Extremal dependence indices: Improved verification mea-  
712 sures for deterministic forecasts of rare binary events 26(5):699–713, DOI 10.1175/WAF-  
713 D-10-05030.1
- 714 Froidevaux P, Martius O (2016) Exceptional integrated vapour transport toward orogra-  
715 phy: an important precursor to severe floods in switzerland. *Quarterly Journal of the*  
716 *Royal Meteorological Society* 142(698):1997–2012, DOI 10.1002/qj.2793, URL ;Go to  
717 ISI;://WOS:000380941100010
- 718 García-Valdecasas Ojeda M, Gámiz-Fortis SR, Castro-Díez Y, Esteban-Parra MJ (2017)  
719 Evaluation of WRF capability to detect dry and wet periods in Spain using drought  
720 indices: WRF Capability to Detect Drought. *Journal of Geophysical Research: Atmo-*

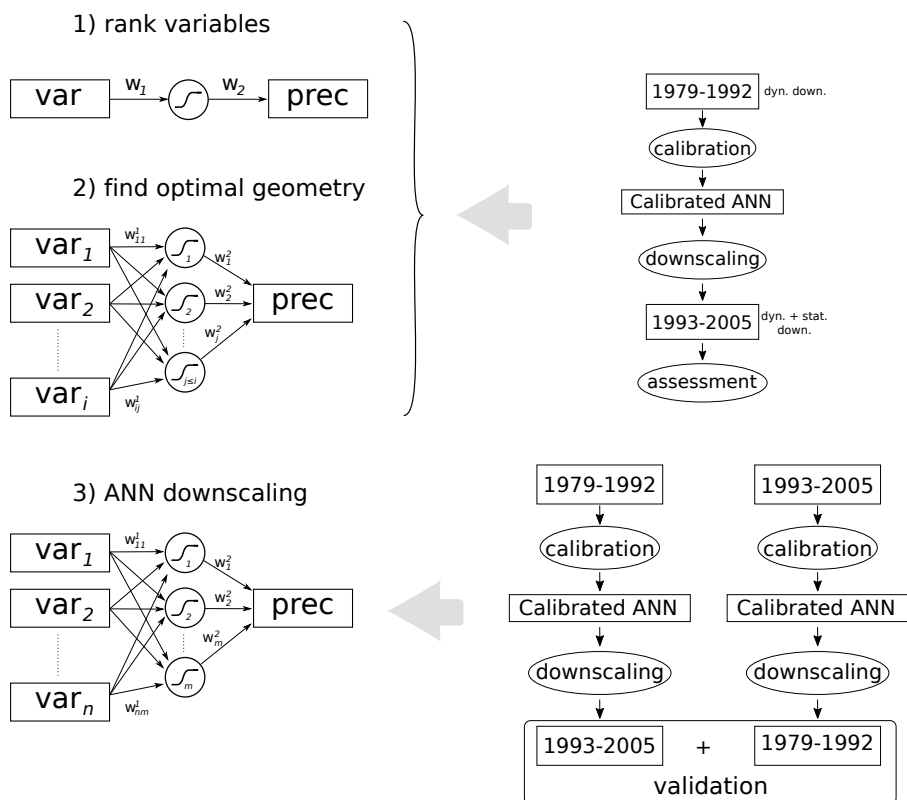
- spheres 122(3):1569–1594, DOI 10.1002/2016JD025683
- Gent PR, Danabasoglu G, Donner LJ, Holland MM, Hunke EC, Jayne SR, Lawrence DM, Neale RB, Rasch PJ, Vertenstein M, Worley PH, Yang ZL, Zhang M (2011) The community climate system model version 4. *Journal of Climate* 24:4973–4991, DOI 10.1175/2011JCLI4083.1
- Giannakaki P, Martius O (2016) Synoptic-scale flow structures associated with extreme precipitation events in northern Switzerland. *International Journal of Climatology* 36(6):2497–2515, DOI 10.1002/joc.4508, URL <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), DOI 10.18637/jss.v072.i08, URL <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 9(8):584–589, DOI 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 11(6):2231–2247, DOI 10.5194/gmd-11-2231-2018
- 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. *Bulletin of the American Meteorological Society* 94:1339–1360, DOI 10.1175/BAMS-D-12-00121.1
- Lee S, Cho S, Wong PM (1998) Rainfall prediction using artificial neural networks. *Journal of Geographic Information and Decision Analysis* 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 System Dynamics* 6(2):411–434, DOI 10.5194/esd-6-411-2015



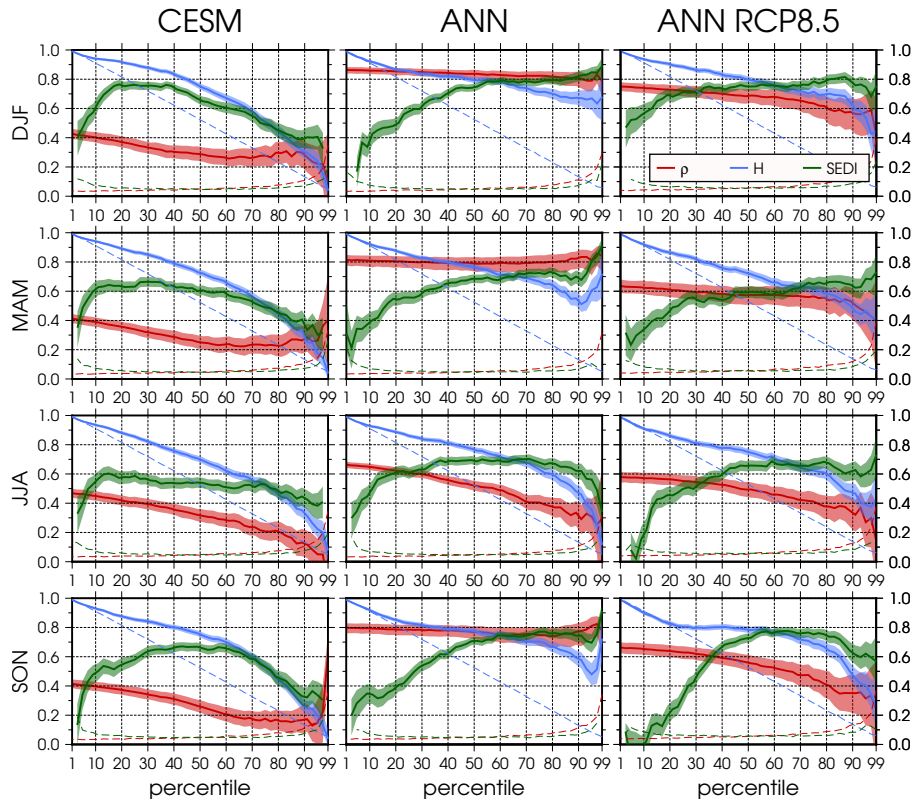
- 749 Li G, Zhang X, Zwiers F, Wen QH (2011) Quantification of uncertainty in high-resolution  
750 temperature scenarios for North America. *Journal of Climate* 25(9):3373–3389, DOI  
751 10.1175/JCLI-D-11-00217.1
- 752 Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S,  
753 Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG,  
754 Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate change:  
755 Recent developments to bridge the gap between dynamical models and the end user.  
756 *Reviews of Geophysics* 48(3), DOI 10.1029/2009RG000314
- 757 Maraun D, Widmann M, Gutiérrez JM, Kotlarski S, Chandler RE, Hertig E, Wibig J, Huth  
758 R, Wilcke RA (2015) VALUE: A framework to validate downscaling approaches for  
759 climate change studies. *Earth's Future* 3(1):1–14, DOI 10.1002/2014EF000259
- 760 Martinez Y, Yu W, Lin H (2012) A New Statistical–Dynamical Downscaling Procedure  
761 Based on EOF Analysis for Regional Time Series Generation. *Journal of Applied Me-*  
762 *teorology and Climatology* 52(4):935–952, DOI 10.1175/JAMC-D-11-065.1
- 763 Martius O, Zenklusen E, Schwieler C, Davies HC (2006) Episodes of alpine heavy precipita-  
764 tion with an overlying elongated stratospheric intrusion: A climatology. *International J*  
765 *Climatol* 26:1149–1164
- 766 Meredith EP, Rust HW, Ulbrich U (2018) A classification algorithm for selective dy-  
767 namical downscaling of precipitation extremes. *Hydrology and Earth System Sciences*  
768 22(8):4183–4200, DOI <https://doi.org/10.5194/hess-22-4183-2018>
- 769 Messmer M, Gómez-Navarro JJ, Raible CC (2015) Climatology of vb cyclones, physical  
770 mechanisms and their impact on extreme precipitation over central europe. *Earth Sys-*  
771 *tem Dynamics* 6(2):541–553, DOI 10.5194/esd-6-541-2015
- 772 Messmer M, Gómez-Navarro JJ, Raible CC (2017) Sensitivity experiments on the response  
773 of vb cyclones to sea surface temperature and soil moisture changes. *Earth System*  
774 *Dynamics* 8(3):477–493, DOI 10.5194/esd-8-477-2017
- 775 North R, Trueman M, Mittermaier M, Rodwell MJ (2013) An assessment of the SEEPS  
776 and SEDI metrics for the verification of 6 h forecast precipitation accumulations: As-  
777 sessment of SEEPS and SEDI for 6 h precipitation accumulations 20(2):164–175, DOI

- 778 10.1002/met.1405
- 779 O’Gorman PA (2015) Precipitation Extremes Under Climate Change. *Current Climate*  
780 *Change Reports* 1(2):49–59, DOI 10.1007/s40641-015-0009-3
- 781 Panziera L, Gabella M, Germann U, Martius O (2018) A 12-year radar-based clima-  
782 *tology of daily and sub-daily extreme precipitation over the swiss alps. Interna-*  
783 *tional Journal of Climatology* 38(10):3749–3769, DOI 10.1002/joc.5528, URL [iGo to](#)  
784 [ISI://WOS:000440826000001](#)
- 785 Pfahl S, Madonna E, Boettcher M, Joos H, Wernli H (2014) Warm conveyor belts in the  
786 *era-interim dataset (1979-2010). part ii: Moisture origin and relevance for precipita-*  
787 *tion. Journal of Climate* 27(1):27–40, DOI Doi 10.1175/Jcli-D-13-00223.1, URL [iGo to](#)  
788 [ISI://WOS:000329276000002](#)
- 789 Piaget N, Froidevaux P, Giannakaki P, Gierth F, Martius O, Riemer M, Wolf G, Grams  
790 *CM (2015) Dynamics of a local alpine flooding event in october 2011: moisture source*  
791 *and large-scale circulation. Quarterly Journal of the Royal Meteorological Society*  
792 *141(690):1922–1937, DOI 10.1002/qj.2496, URL [iGo to ISI://WOS:000360203800034](#)*
- 793 Rajczak J, Pall P, Schär C (2013) Projections of extreme precipitation events in regional  
794 *climate simulations for Europe and the Alpine region. Journal of Geophysical Research:*  
795 *Atmospheres* 118(9):3610–3626, DOI 10.1002/jgrd.50297
- 796 Rajczak J, Kotlarski S, Schär C (2016) Does quantile mapping of simulated precipita-  
797 *tion correct for biases in transition probabilities and spell-lengths? Journal of Climate*  
798 *29(5):1605–1615, DOI 10.1175/JCLI-D-15-0162.1*
- 799 Requena AI, Mediero L, Garrote L (2013) A bivariate return period based on copu-  
800 *las for hydrologic dam design: accounting for reservoir routing in risk estimation*  
801 *17(8):3023–3038, DOI 10.5194/hess-17-3023-2013, URL [https://www.hydrol-earth-syst-](https://www.hydrol-earth-syst-sci.net/17/3023/2013/)*  
802 [sci.net/17/3023/2013/](https://www.hydrol-earth-syst-sci.net/17/3023/2013/)
- 803 Salvadori G, De Michele C, Durante F (2011) On the return period and design in  
804 *a multivariate framework 15(11):3293–3305, DOI 10.5194/hess-15-3293-2011, URL*  
805 <https://www.hydrol-earth-syst-sci.net/15/3293/2011/>
- 806 Schalkoff RJ (1997) *Artificial neural networks, vol 1. McGraw-Hill New York*

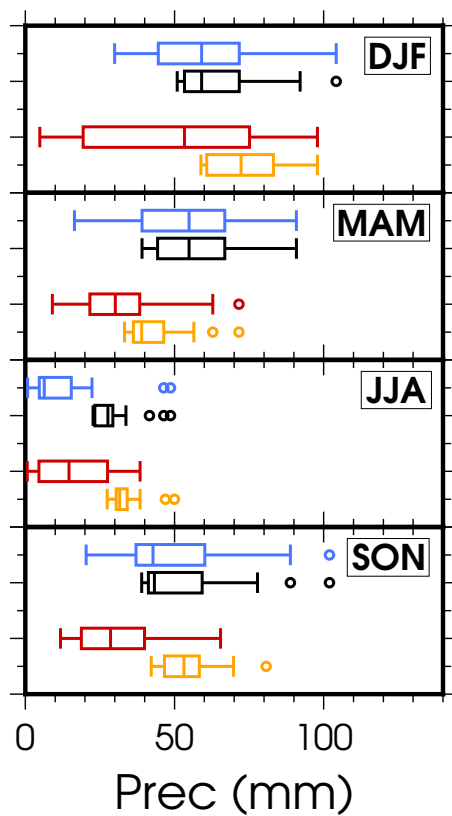
- 807 Seneviratne SI, Nicholls N, Easterling D, Goodess CM, Kanae S, Kossin J, Luo Y, Marengo  
808 J, McInnes K, Rahimi M, Reichstein M, Sorteberg A, Vera C, Zhang X (2012) Changes  
809 in climate extremes and their impacts on the natural physical environment. In: Field  
810 CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, Mastrandrea MD, Mach KJ,  
811 Plattner GK, Allen SK, Tignor M, Midgley PM (eds) *Managing the Risks of Extreme  
812 Events and Disasters to Advance Climate Change Adaptation. A Special Report of  
813 Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC)*,  
814 Cambridge University Press, pp 109–230
- 815 Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, Powers JG (2008) A  
816 description of the advanced research wrf version 3. Tech. Rep. TN-475+STR, National  
817 Center for Atmospheric Research
- 818 Stephenson DB (2000) Use of the “odds ratio” for diagnosing forecast skill 15(2):221–232,  
819 DOI 10.1175/1520-0434(2000)015;0221:UOTORF;2.0.CO;2
- 820 Stucki P, Rickli R, Brönnimann S, Martius O, Wanner H, Grebner D, Luterbacher J (2012)  
821 Weather patterns and hydro-climatological precursors of extreme floods in Switzerland  
822 since 1868. *Meteorologische Zeitschrift* 21(6):531–550, DOI 10.1127/0941-2948/2012/368
- 823 Torma C, Giorgi F, Coppola E (2015) Added value of regional climate modeling over areas  
824 characterized by complex terrain—precipitation over the alps 120(9):3957–3972, DOI  
825 10.1002/2014JD022781
- 826 Winschall A, Pfahl S, Sodemann H, Wernli H (2012) Impact of north atlantic evapora-  
827 tion hot spots on southern alpine heavy precipitation events. *Quarterly Journal of  
828 the Royal Meteorological Society* 138(666):1245–1258, DOI 10.1002/qj.987, URL [Go  
829 to ISI://WOS:000306859800009](#)
- 830 Zittis G, Bruggeman A, Camera C, Hadjinicolaou P, Lelieveld J (2017) The added value  
831 of convection permitting simulations of extreme precipitation events over the eastern  
832 mediterranean 191:20–33, DOI 10.1016/j.atmosres.2017.03.002
- 833 Zorita E, Storch HV (1999) The analog method as a simple statistical downscaling technique:  
834 comparison with more complicated methods 12:2474–2489



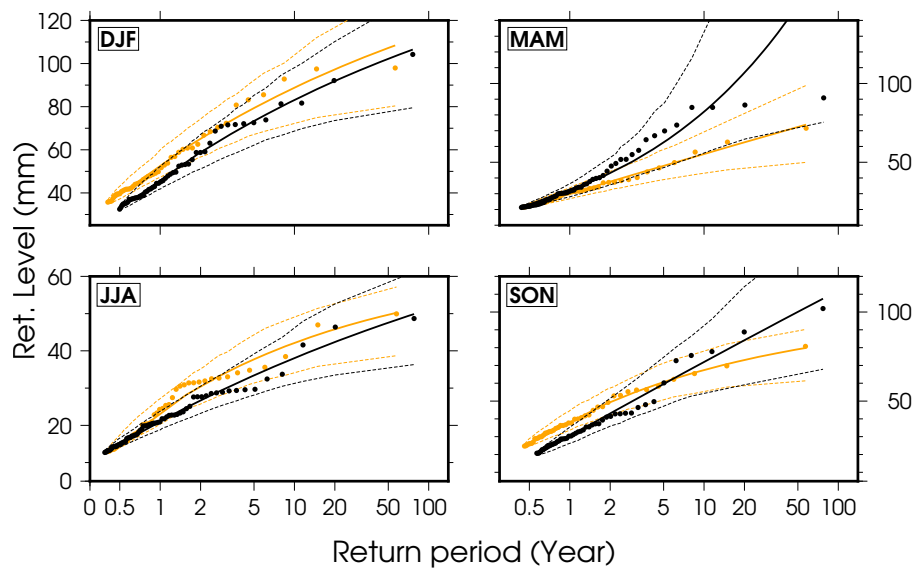
**Fig. 5** Steps to find the optimal ANN geometry, which is then used to produce a statistical downscaling that can be used to validate the skill of the ANN during the WRF-reference period. 1) all variables considered are tested individually with a trivial ANN to build a ranking. Each ANN is calibrated 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.



**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 Sec. 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 bootstrapping the sample with repetition. Dashed lines represent the threshold to reject the null hypothesis of skill by random chance at the 95% confidence level.



**Fig. 7** Distribution of daily precipitation in the Aare catchment. Each boxplot represents different datasets and periods. Black and blue 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 corresponds to the days predicted by the ANN. Red and orange represent the same information but for the WRF-future simulation (2080-2099), discussed in Sec. 4: orange represents the actual extreme events observed in the WRF-future simulation, whereas red correspond to the individual cases predicted by the ANNs.



**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>).