



# Assessing local daily temperatures by means of novel analog approaches: a case study based on the city of Augsburg, Germany

Christian Merckenschlager<sup>1,2</sup> · Stephanie Koller<sup>2</sup> · Christoph Beck<sup>1</sup> · Elke Hertig<sup>2</sup>

Received: 24 June 2020 / Accepted: 24 March 2021 / Published online: 16 April 2021  
© The Author(s) 2021

## Abstract

Within the scope of urban climate modeling, weather analogs are used to downscale large-scale reanalysis-based information to station time series. Two novel approaches of weather analogs are introduced which allow a day-by-day comparison with observations within the validation period and which are easily adaptable to future periods for projections. Both methods affect the first level of analogy which is usually based on selection of circulation patterns. First, the time series were bias corrected and detrended before subsamples were determined for each specific day of interest. Subsequently, the normal vector of the standardized regression planes (NVEC) or the center of gravity (COG) of the normalized absolute circulation patterns was used to determine a point within an artificial coordinate system for each day. The day(s) which exhibit(s) the least absolute distance(s) between the artificial points of the day of interest and the days of the subsample is/are used as analog or subsample for the second level of analogy, respectively. Here, the second level of analogy is a second selection process based on the comparison of gridded temperature data between the analog subsample and the day of interest. After the analog selection process, the trends of the observation were added to the analog time series. With respect to air temperature and the exceedance of the 90th temperature quantile, the present study compares the performance of both analog methods with an already existing analog method and a multiple linear regression. Results show that both novel analog approaches can keep up with existing methods. One shortcoming of the methods presented here is that they are limited to local or small regional applications. In contrast, less pre-processing and the small domain size of the circulation patterns lead to low computational costs.

## 1 Introduction

Most of the recent urban climate change studies are based on Regional Climate Models (RCMs) to dynamically downscale large-scale information. The RCM output is then implemented in or provides the boundary conditions for urban climate models in order to assess the future development of the urban heat island UHI (e.g., Town Energy Balance scheme, TEB, in Hamdi et al. (2014); Urban boundary layer climate model, UrbClim, in Lauwaet et al. (2016)). A study which not only uses RCMs but also statistical downscaling (SD) techniques is presented by Früh et al. (2011). On the one hand, the authors use the weather generator “Wetterlagenbasierte

Regionalisierungsmethode WETTREG” (ENKE et al. 2005) which rearranges circulation patterns in order to preserve the previously derived frequency distribution. On the other hand, they use the analog method “Statistical Analogue Resampling Scheme STARS” (ORLOWSKY et al. 2008, LUTZ & GERSTENGARBE 2015) which rearranges circulation patterns in order to preserve the specific linear temperature trend. The generated time series of the predictor fields were then used as input variables for an urban climate model (MUKLIMO).

SD has the advantage that it can provide site specific-information, is computational inexpensive, and can be easily applied to different models for future projections (WILBY et al. 2004). However, one shortcoming of all SD techniques is that the statistical link is assumed to be valid under different forcings and boundary conditions (HERTIG & JACOBEIT, 2008). Wilby et al. (2004) group different statistical downscaling approaches into three main statistical downscaling techniques: weather typing, weather generators, and regression methods. Each of the mentioned methods are used in any manner for urban climate studies but mostly affect the

✉ Christian Merckenschlager  
christian.merckenschlager@geo.uni-augsburg.de

<sup>1</sup> Institute of Geography, University of Augsburg, Alter Postweg 118, 86159 Augsburg, Germany

<sup>2</sup> Faculty of Medicine, University of Augsburg, Alter Postweg 118, 86159 Augsburg, Germany

representation of the spatial distribution of the variable of interest (i.e., second step which is not part of the study) and not the downscaling process of large-scale information. For example, (urban) weather generators were used to assess the effects of different parameters on the UHI (Salvati et al. 2017), the influence of the UHI on buildings energy budget (e.g., BUENO et al. 2013), or the characteristics of the UHI above the urban canopy layer (LE BRAS & MASSON 2015). Regression models, especially multiple linear regression (MLR), were used for detailed analysis of the temperature distribution within urban areas (e.g., STRAUB et al. 2019; KETTERER & MATZARAKIS 2015), but also other approaches like Regression Trees (e.g., MAKIDO et al. 2016) or Random Forests (e.g., HO et al. 2014) are suitable for assessing the characteristics of the UHI.

Here, statistical downscaling approaches based on the analog method were applied. The general idea of analogs is based on the assumption that each weather situation that appears in the future was approximately observed in the past (LUTZ & GERSTENGARBE 2015). But the use of analogs is subjected to restrictions when the pool of training observations is too small (TIMBAL et al. 2003) or the number of classifying predictors is too large (VAN DEN DOOL 1989). However, when the approach meets the requirements, the analog method exhibits the advantages that it yields physically interpretable linkages to surface variables and it is versatile and applicable for extreme event analysis (WILBY et al. 2004). Thus, analog techniques represent marginal aspects well (MARAUN et al. 2019) and, therefore, are applied in different recent studies to assess precipitation or temperature extremes (Castellano and DeGaetano 2016, 2017; RAU et al. 2020; CHATTOPADHYAY et al. 2020).

Many studies use analog approaches for downscaling of, for example, precipitation (e.g., ZORITA & VON STORCH 1999, RADANOVICS et al. 2013; DAYON et al. 2015, HORTON et al. 2017), short wave radiation, and/or temperature (Delle Monache et al. 2011; LUTZ & GERSTENGARBE 2015; CAILLOUET et al. 2016). The requirements for the analog approach in this study are that the analogs represent a realistic picture of the present weather conditions. Moreover, the approach should be applicable to future climate scenarios and should be as straightforward and easily interpretable as possible. An existing analog method which meets the requirements is, for example, the approach of Zorita and Von Storch (1999). Although the method seems to be a relatively old analog approach, many recent studies successfully applied this method (e.g., WAHL et al. 2019, DIAZ et al. 2016, OHLWEIN ET AL. 2012). Thus, the chosen analog method is suitable for comparative analysis.

In the present study, assessments of the future spatial temperature development within the city of Augsburg (Bavaria, Southern Germany) and its surroundings were performed. The assessment process is divided into two steps. Since the

framework conditions of the temperature behavior within the city of Augsburg are determined by the large-scale circulation, first, downscaling is applied to transfer large-scale information on point observations. Second, under consideration of a precise description of the land surface characteristics, a multiple linear regression (MLR) was performed to specify local temperature variability within the city of Augsburg. The present study is focused on the downscaling process. We show that the developed analog approaches are suitable for climate projections and allow for a day-by-day comparison between observation and the analog time series. The general procedure is inspired by the analog setup described in Horton et al. (2017) (but also used in previous studies before, e.g., Marty et al. (2012) and Daoud (2016)) and is divided in four different steps: pre-selection, first level of analogy, subsequent level(s) of analogy, and projection. Here, the novel approaches mainly affect the first level of analogy but also, for some issues (e.g., exceedance of the 90th temperature quantile), the pre-selection process. Different selection criteria (SC) during the analog selection process affect either the pre-selection process (Q90) or the second (subsequent) level of analogy (T2M), whereas different selection variables (SV) only affect the first level of analogy. Since one aim of the project is to describe future heat extremes on a daily scale, single analog situations instead of analog ensembles were considered for the day of interest. Thus, the tails of the temperature distribution were not obliterated, and extremes were entirely reproduced.

Both analog techniques presented here focus on local assessments of climate variables and, thus, are very well suited for urban climate change studies. This means that due to the limitation of the domain size of the circulation patterns, computational costs are very low. For example, with respect to the circulation, the amount of data needed for the NVEC or COG method is only 22% of the data amount used for the approach of Zorita and Von Storch (1999) or multiple linear regression. This leads to a massive reduction of computational costs. Furthermore, the pre-processing of the data is limited to bias correction, detrending and standardization (NVEC), or normalization (COG) in a limited domain and, thus, efficiently applicable, particularly in the scope of climate change studies. The normal vector and the center of gravity were applied to summarize the main features of the prevailing circulation. In combination with the mean of the circulation patterns, both approaches provide indications about the main flow direction of the air masses and, thus, ground-level air temperature. The transfer of the values derived from the NVEC or COG to an artificial coordinate system allows for a clear method to define the analog by only one value (distance between the points of the day of interest and the subsample). In summary, the described advantages make the presented analog approaches suitable for local climate studies like urban climate change analysis.

In Section 2, a brief summary of the observation and reanalysis data used in this study is given. The methods are presented in Section 3. The method section comprises the preparation and pre-selection of the data, the introduction of the two novel analog methods, and post-processing steps to make these methods suitable for studies concerning projections of future climate changes. The results are provided in Section 4. Section 4.1 summarizes the trend analysis of the temperature variables. In Section 4.2, the overall performance of the analog methods was tested by means of the averaged root mean square error (RMSE) considering all standardized variables listed in the data section. The quality of temperature assessments is presented in Section 4.3. Here, the RMSE was calculated by means of the absolute temperature values for each temperature variable and analog method, separately. The agreement of exceeding the 90th quantile is checked by using the Brier Score (Section 4.4). At the end of Section 4, the analog methods are also tested against a further statistical downscaling technique based on multiple linear regression (MLR). For this purpose, the Mean Square Error Skill Score (MSESS) was used for comparative analysis. Finally, the conclusions in Section 5 summarize the present study and highlight the major findings.

## 2 Data

### 2.1 Predictand

The daily temperature data set of the weather station Augsburg-Mühlhausen operated by the German Weather Service (DWD) was used as reference time series (elevation: 461.4m; 10.9420° E, 48.4254° N). The data set contains information on the daily minimum (TMIN), maximum (TMAX), and mean (TMEAN) temperatures 2m above ground and various other variables for the period from 1950 to present. In order to obtain an accurate assessment of the daily temperature development, all three temperature variables were considered during the analog selection process. The data sets were adapted to the length of the ERA5 data set so that the time series contains 40 years of daily temperature data (1979–2018). Since the present study aims at assessing the occurrence and development of heat waves, the assessment period was restricted to the extended summer season from May to September.

After performing different absolute homogeneity tests on a time series representing the seasonal means of the respective temperatures, the Pettitt-Test (PETTITT 1979) assumes an inhomogeneity in the year 2000, and the Buishand Range Test (BUISHAND 1982) as well as the absolute Standard Normal Homogeneity Test (ALEXANDERSSON 1986) assume an inhomogeneity for TMIN in 1999, whereas TMEAN and TMAX exhibit no inhomogeneities. One possible reason for inhomogeneities is assumed to be the consequence of

changes at the measurement station. After removing the mean of the minimum temperatures, Augsburg-Mühlhausen and ERA5 exhibit a general agreement until the late 1990s, followed by a period of increasing differences until mid-2000s. Afterwards, the difference between reanalysis and observation remains on an almost consistent level. A look on the metadata of the station shows that on the one hand, the minimum (and maximum) thermometer was replaced in 1998 and 2006 by a new, but identical in construction, thermometer. On the other hand, the station was moved at the same date from open fields near the airport directly onto the airport grounds. The weather station was then located closer to the landing strip on the northeastern side of the main building of Augsburg airport and, thus, on the lee site regarding the main wind direction before the station was moved again in 2006. Furthermore, ongoing gravel extraction in the direct vicinity could have led to a lowering of the groundwater table and, thus, to drier soil conditions. All these aspects affect the temperature development and could be responsible for the inhomogeneity within the observation time series of TMIN. Especially the lowering of the groundwater table seems to be the most coherent explanation for the increasing spread of TMIN. Before gravel extraction starts, the groundwater had a significant influence on the soil water budget. Since humid soil could store more energy (heat) than dry soils, the difference between observation and reanalysis exhibits a lower (humid) constant level. Within the period 1998–2006, the influence of the groundwater table decreases, and soils are getting drier which could explain the increasing spread. In the years after, the mean groundwater table has only little impact on the soil water budget, and the temperature difference of TMIN between observation and reanalysis remains on a higher (drier) level. This can also explain the fact that only TMIN is affected by an inhomogeneity since this process affects temperature development especially when other influencing factors (e.g., solar radiation) are smaller or absent.

### 2.2 Predictors

In order to describe the prevailing circulation adequately, daily data of the geopotential heights on the 850hPa (HGT.0850) and 700hPa-level (HGT.0700) as well as sea level pressure (SLP) from the ERA5 reanalysis data set provided by the European Centre for Medium-Range Weather Forecasts (Copernicus Climate Change Service (C3S) 2017) were taken into account to determine the analog days. Furthermore, daily minimum, mean, and maximum temperature data at 2m above ground (t2m) as well as the layer thickness (HGT.LT = HGT.0700–HGT.0850) were considered as thermal variables. All data sets were downloaded on a 0.75° × 0.75° spatial resolution for the period 1979–2018. Although a higher resolution is available for ERA5, the resolution was restricted to 0.75° in order to provide a resolution rather

similar to the highest resolution of the Generalized Circulation Models (GCMs) of the Coupled Model Intercomparison Project - Phase 5 (CMIP5).

Since the following methods are limited to a regional to local scale, the domain size has to be selected very carefully. The presented methods are only applicable if a small number of pressure systems are gathered by the chosen domain size so that each derived values of the determinants can exclusively be assigned to a specific state of the atmosphere. Thus, the domain for circulation dynamic predictors and layer thickness is limited to the area from 0.25 to 19.75° E and 40.0 to 59.5° N (center of grid points) resulting in a  $27 \times 27$  grid. This approximately corresponds to the size of a distinct pressure system located in the center of the domain. Therefore, the domain size should not be smaller, but it is quite possible that larger domains could provide good results although it is more likely that two or more pressure systems cancel each other. In this study, only the mentioned domain size was tested. For temperature variables, the grid box containing Augsburg-Mühlhausen weather station and the eight surrounding grid boxes were used to define the grid of interest (Fig. 1). As for the station data, the data is limited to the extended summer season.

### 3 Methods

#### 3.1 Data preparation and pre-selection

Before the analog selection process starts, bias correction and detrending were applied on all data. Since the resolution of General Circulation Models (GCMs) and even reanalysis data is lower than desired, processes governing regional- to local-scale climate are not well represented. The systematic error between observation and climate model or reanalysis data, respectively, represents the bias (MARAUEN 2016). Thus, bias correction is needed to adapt model or reanalysis data to the observation for impact studies or downscaling. By means of bias correction, the analog time series derived from the reanalysis data represent a more realistic picture of the local temperature behavior. The detrending process represents a simple method that allows for projecting climate change by adding the trends during the post-processing of the data. Bias correction and detrending were implemented as follows:

- Bias correction: Based on the difference of the overall across the 9 ERA5 grid boxes and the Augsburg-Mühlhausen time series, the temperatures of ERA5 were adapted to the Augsburg-Mühlhausen temperatures. Thus,

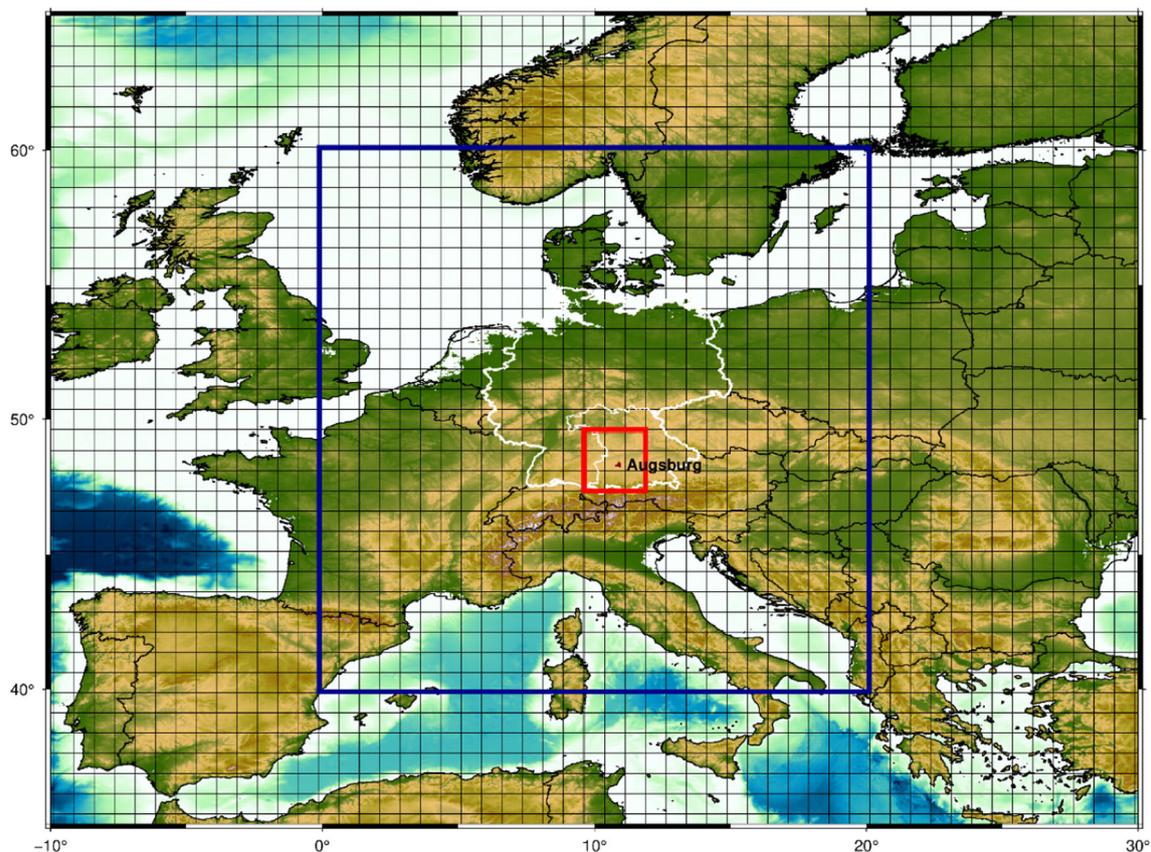


Fig. 1 Domains for the circulation-based predictor variables (blue) and for temperature (red)

the means of all temperature time series of the ERA5 data set correspond to the means of Augsburg-Mühlhausen.

- **Detrending:** All data sets were detrended by means of the linear trends of the seasonal means. The daily values were averaged over the extended summer season so that each variable is represented by one seasonal temperature value. A linear regression based on the least absolute deviations estimation scheme is then applied on the seasonal mean values. The resulting slope represents the long-term seasonal trend. Subsequently the trend is subtracted from each value of the daily time series in order to remove the inter-annual trend but to preserve the intra-seasonal variability.

To identify analogs for each day of the assessment period (see below), specific pre-selection steps were implemented. The idea is to minimize the initial subsample (whole time series) from which analogs were drawn without eliminating the most realistic temperature and circulation patterns. The following points comprise the pre-selection process based on the analog setup described in Horton et al. (2017). For this purpose, time series of different length had to be determined. The different time series are defined as follows:

- **Assessment period (ASP):** The ASP contains all days between May 1 and September 30 of the years 1979–2018. In the following, analogs were determined for each day of the ASP.
- **Analog period (ANP):** The ANP represents the period from which the analog days were drawn. Since atmospheric patterns represent a more realistic picture of the respective variables (e.g., temperature, relative humidity) when they are close to the day of interest, the subsample is restricted to  $\pm 15$  days of the respective day of ASP. Thus, in order to provide subsamples of the same size for each day of the ASP, the length of the analog period has to be extended by 30 days so that the ANP contains all days of the year between April 16 and October 15.
- **Quantile period (QP):** Since the following analysis aims at heat waves, in some parts, the exceedance of the daily 90th temperature quantiles is also considered during the analog selection process. The daily Q90 values are derived from the distribution of the temperature values considering a subsample of  $\pm 2$  days of the day of interest. Since quantile values have to be available for each day of the ANP, the time series has to be further extended by two days in each direction. Thus, the QP contains all days of the year between April 14 and October 17. Each day of the year is now characterized by one quantile value for each temperature variable. If the respective temperature of the day of interest exceeds the quantile value, the day is flagged by 1, else by 0. Thus, a day is characterized by three values, for example, 0|1|1 for a day where only

TMIN lies below Q90 or 0|1|0 where only TMEAN exceeds Q90. If Q90 is considered as selection criteria, analogs were only drawn from a subsample which exhibits the same exceedances of Q90 within  $\pm 15$  days of the day of interest.

### 3.2 Analog approaches (first level of analogy)

The main selection process (first level of analogy) represents two novel approaches, one based on the normal vector (NVEC) and the other on the center of gravity (COG) of the circulation patterns or layer thickness, respectively. By means of the NVEC and COG, the main features of the circulation patterns were summarized. In connection with the means of the circulation patterns, indications about the main wind flow direction and speed as well as air temperature can be derived. But this is only possible when the extent of the circulation patterns is not too large since the greater the number of pressure centers within the domain, the more the explanatory power of both variables would be reduced. Days which exhibit similar circulation patterns were then taken into account for further analysis. Here, the similarity of circulation patterns (layer thickness) is derived from points within an artificial coordinate system defined by the NVEC or COG, respectively. For validation, the data sets were divided into four 10-year sub-periods (1979–1988, 1989–1998, 1999–2008, 2009–2018). For each day of one sub-period (validation period), analogs were drawn from the remaining data (i.e., three sub-periods). This procedure was performed for all 10-year validation periods so that an analog time series is available for each of the sub-periods. Subsequently, the analog time series of the validation periods were merged in order to receive an analog time series comprising all days and years of the original time series. Thus, the results of the analog methods can be compared with the original or detrended time series of Augsburg-Mühlhausen.

To check whether the following approaches are limited to circulation only or whether they also work for thermal variables, layer thickness was also taken into account for the selection of analogs. Thus, we were also able to check whether a selection based on temperature characteristics, only, provides better results for assessing temperatures extremes than an approach with circulation included.

#### 3.2.1 Normal vector

First, the circulation (geopotential heights) or pressure patterns (sea-level pressure) were standardized. An example for one day is given in Fig. 2 (left). The standardized patterns are presented for sea-level pressure and the geopotential heights on the 850hPa and 700hPa level. By means of a linear regression model (based on the least

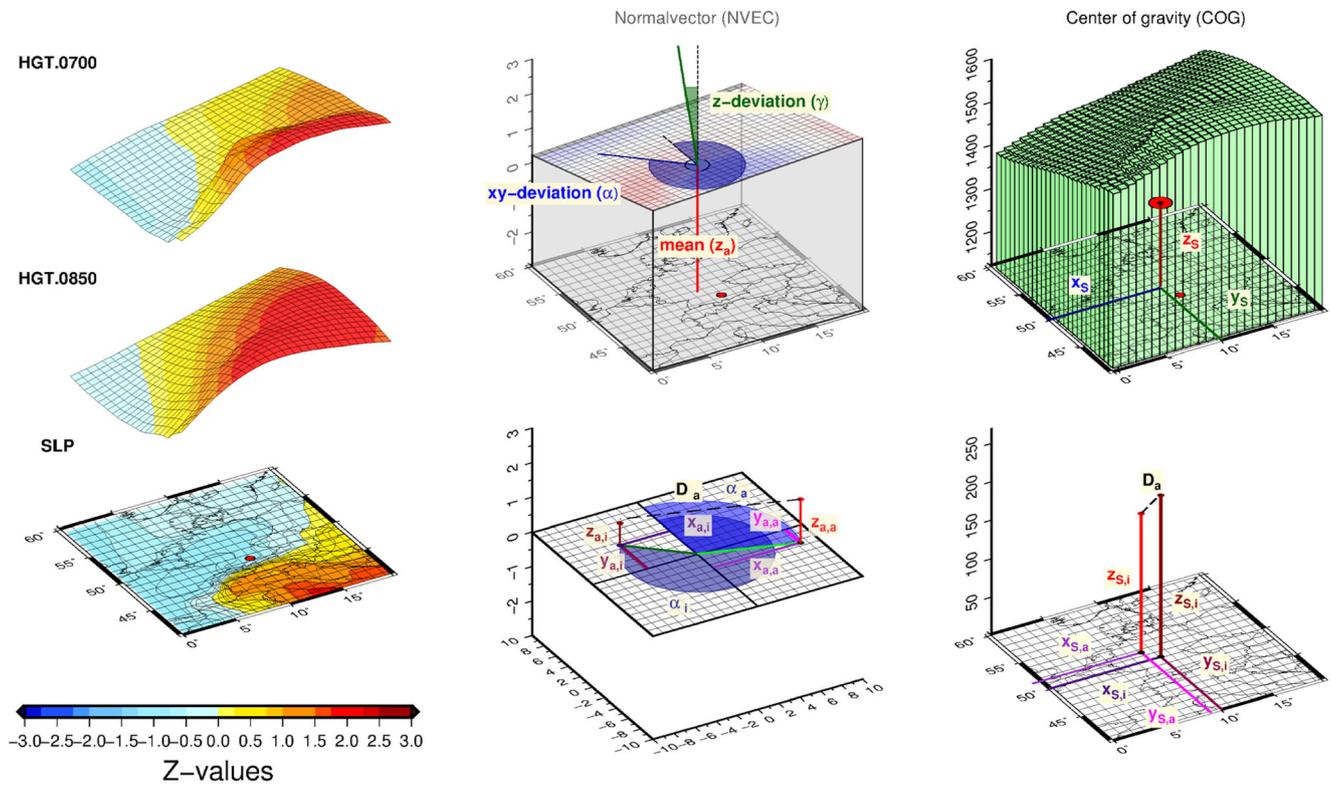


Fig. 2 Graphical overview for the analog methods based on the normal vector (mid) and the center of gravity (right). The standardized values of the geopotential heights and the sea level pressure of the day of interest are depicted on the left

absolute deviation scheme), the standardized patterns were transformed to an even surface (regression plane, RP) which is inclined towards the perpendicular (Fig. 2 top, middle). The inclination of the RP is exactly described by the normal vector  $\vec{n} = (n_1 \ n_2 \ n_3)$  which is orthogonal to the RP. But the normal vector  $\vec{n}$  only describes the deviation to the origin (i.e., perpendicular of an even surface with an inclination=0) and the length is not determined. Thus, the normal vector  $\vec{n}$  provides too little information and is not useful to describe the underlying circulation on its own. However, in combination with the standardized mean of the circulation pattern, the given information by the normal vector  $\vec{n}$  can be used to define an exact point in an artificial coordinate system. Information given by the normal vector are the direction of deviation in  $xy$ -direction ( $\alpha'$ ) and the deviation from the perpendicular (cf. Fig. 2 top, middle). The deviation in  $xy$ -direction ( $\alpha'$ ) is calculated by:

$$\sin(\alpha) = \frac{n_1}{\sqrt{(n_1^2 + n_2^2)}} \tag{1}$$

and

$$\alpha' = \begin{cases} \alpha, & \text{if } n_1 > 0 \wedge n_2 > 0 \\ 360^\circ + \alpha, & \text{if } n_1 < 0 \wedge n_2 > 0 \\ 180^\circ - \alpha, & \text{if } n_2 < 0 \end{cases}$$

The deviation in  $z$ -direction results from

$$\cos(\gamma) = \frac{n_3}{\sqrt{(n_1^2 + n_2^2 + n_3^2)}} \tag{2}$$

For the artificial coordinate system (Fig. 2 bottom, middle), the deviation in  $xy$ -direction ( $\alpha'$ ) is maintained with northern direction as point of origin (0 1 0), whereas the  $z$ -deviation ( $\gamma$ ) will be used to define the length of the artificial vector ( $c = \gamma$ ). In order to avoid scale problems accompanied with the transition between 360 and 0°, angles and lengths are transformed to coordinates within the artificial coordinate system. Thus, the coordinates for  $x_a$  and  $y_a$  for the new artificial point are exactly defined:

$$\begin{aligned} x_a &= \sin(\alpha') \cdot c \\ y_a &= \cos(\alpha') \cdot c \end{aligned} \tag{3}$$

For  $z_a$  the means of the standardized daily circulation or pressure patterns are used so that the circulation for each day is exactly defined by one point  $A = (x_a \ y_a \ z_a)$  in a three-dimensional artificial coordinate system. The consideration of the standardized daily circulation means  $z_a$  prevents that days with high pressure could be assigned to days of low pressure although the inclination against the perpendicular is approximately the same. To find an analog for the respective day, the distance between the artificial point of the day of interest ( $A_i$ ) and the analog subsample ( $A_a$ ) was taken into account. The smaller the distance  $D_a$ , the more the circulation of the analog represents the day of interest.

$$D_a = \sqrt{(x_{a_i} - x_{a_a})^2 + (y_{a_i} - y_{a_a})^2 + (z_{a_i} - z_{a_a})^2} \tag{4}$$

The artificial coordinate system as well as an example for the day of interest (subscript  $i$ ), an analog situation (subscript  $a$ ), and the distance between both points ( $D_a$ ) are given in Fig. 2 (bottom, middle).

### 3.2.2 Center of gravity

The analog approach based on the center of gravity (COG) uses modified absolute values of the daily circulation or pressure variables ( $z$ ) as weights on an imaginary even surface (IES). Since negative weights (i.e., negative values of the respective variable) are not applicable to this method, the use of standardized values is not possible. Furthermore, too high values (e.g., geopotential heights on higher levels) reduce the differences between the center of gravity and the geometric center. In order to enlarge the distance range of the COG, the IES is defined as the absolute minimum value of circulation or pressure variable ( $z_{\min}$ ) of the entire time series. The normalized daily circulation or pressure variables ( $z'_{gb_i}$ ) are calculated as follows:

$$z'_{gb_i} = z_{gb_i} - z_{\min} \tag{5}$$

By means of the normalized values  $z'_{gb_i}$ , the center of gravity is then calculated for each day  $i$  of the time series. Assuming that the density of the weights is equal for all heights and grid boxes, the modified absolute values  $z'_{gb_i}$  can also be considered as cuboid with a specific height on the IES (Fig. 2 top right) and, thus, the center of gravity  $S_{gb_i}$  of each grid box  $gb = 1, \dots, n$  is equal to the geometric center of the cuboid. Therefore, the center of gravity  $S_i$  for a specific day  $i$  results from:

$$\begin{aligned} x_{S_i} &= \frac{\sum_{gb=1}^n \bar{x}_{gb} \cdot V_{gb_i}}{\sum_{gb=1}^n V_{gb_i}}; y_{S_i} = \frac{\sum_{gb=1}^n \bar{y}_{gb} \cdot V_{gb_i}}{\sum_{gb=1}^n V_{gb_i}}; \\ z_{S_i} &= \frac{\sum_{gb=1}^n \bar{z}'_{gb_i} \cdot V_{gb_i}}{\sum_{gb=1}^n V_{gb_i}}; S_i = \begin{pmatrix} x_{S_i} \\ y_{S_i} \\ z_{S_i} \end{pmatrix} \end{aligned} \tag{6}$$

where  $\bar{x}_{gb}$  and  $\bar{y}_{gb}$  represent the distance of the center of each grid box to the origin of the IES. The origin of the IES is defined as the center of the domain chosen for the circulation variables (approx. location of Augsburg, 10° E, 49.75° N). According to the assumption that the density is equal for all heights, the volume ( $V_{gb_i}$ ) of the cuboid represents the specific weight at grid box  $gb$  on day  $i$ , and the least distance (perpendicular) between the center of the cuboid and the IES is given by  $\bar{z}'_{gb_i}$  (i.e.,  $\bar{z}'_{gb_i} = \frac{z'_{gb_i}}{2}$ ). The next step is in accordance with the normal vector approach (Eq. 4). In Fig. 2 (bottom right), the COGs ( $S_i$ ) of the day of interest and a potential analog day are displayed, and the distance  $D_a$  is depicted as dashed lines between both points.

### 3.3 Final selection (second level of analogy)

After the least distances between the artificial points ( $D_a$ ) were determined, a further selection step was applied on an updated subsample. The updated subsample contains a maximum of 5% of the initial subsample which exhibits the least distances between the points within the artificial coordinate system. The averaged root mean square error (RMSE) of the gridded temperature fields (TMIN, TMEAN, and TMAX,  $i = 1, \dots, t$ ) between the day of interest  $x_o$  and the days of the updated subsample  $x_a$  were calculated by row ( $j = 1, \dots, r$ ) and column ( $k = 1, \dots, c$ ). The day which achieves the lowest  $RMSE_{Ana}$  was then taken into account as analog (SC T2M).

$$RMSE_{Ana} = \frac{1}{i \cdot j \cdot k} \sum_{i=1}^t \left( \sum_{j=1}^r (x_{o_{ij}} - x_{a_{ij}})^2 + \sum_{k=1}^c (x_{o_{ik}} - x_{a_{ik}})^2 \right) \tag{7}$$

All variables which are given by the Augsburg-Mühlhausen weather station (e.g., TMIN, TMEAN, and TMAX) were subsequently used as the corresponding values of the analog day. Data which could not be recorded by the weather station, especially the gridded circulation-dynamic variables (i.e., SLP, HGT0850, HGT0700), were adopted from the respective day of the ERA5 data set.

### 3.4 Consideration of trends

For each of the 10-year validation periods, we now have a corresponding analog time series for all variables of interest,

but the analog temperature time series just match the detrended temperature time series of Augsburg-Mühlhausen. Therefore, the derived trends in the beginning of the selection process had to be taken into account to generate the “real” analog time series. Under consideration that the approaches described above should finally be applied to future scenarios (not part of this study) and, thus, to different representative concentration pathways (RCPs), information (trends) had to be taken into account which are not part of the observation (i.e., temperature data of Augsburg-Mühlhausen). For model validation, the calculated trends of the reanalysis data set (here ERA5) were added to the detrended analog temperature time series. For future climate studies, the trends of the temperatures derived from the respective Earth System Model runs under different RCP scenarios can be taken into account for generating the analog time series (not part of this study).

### 3.5 Comparison with existing analog approach

For comparison the results of both methods are checked against the widespread approach of Zorita and Von Storch (1999) which is based on the least Euclidean distance between the new coordinates (PC scores) of the day of interest and the analogs generated by an *s*-mode principal component analysis (*s*-mode PCA). Here, PCA is performed by means of the anomalies of the circulation patterns based on a domain representing 3240 grid boxes on a  $0.75^\circ \times 0.75^\circ$  spatial resolution (20W–40E and 35.5N–64.75N). The method is well established, and many studies are based on the fundamentals of this method (e.g., Wahl et al. (2019), Diaz et al. (2016), Ohlwein and Wahl (2012)). Thus, with respect to the analog approaches described in the present study, the method of Zorita and Von Storch (1999) is well suited for comparative analysis.

All pre- and post-processing steps remain the same for all approaches, and only the first level of analogy which considers the large-scale circulation is replaced by the respective method. The day(s) which exhibit(s) the least Euclidean distances compared to the day of interest represent(s) the analog or 5% subsample, respectively. Thus, all following results are comparable one-by-one for the different approaches.

### 3.6 Comparison with other statistical downscaling approaches

In order to check whether or not the analog approaches can keep up with other statistical downscaling techniques, a multiple linear regression (MLR) was performed by means of the same predictor ensemble used for the analogs. For this purpose, centers of variation were extracted for the circulation and layer thickness by means of an *s*-mode principal component analysis (PCA, c.f. PREISENDORFER 1988) from the original ERA5 reanalysis data set. Since preliminary analysis

showed that SLP provides the best skills, results are presented for the predictor combination of sea level pressure and the respective temperatures of ERA5. In contrast to the analog methods, the data domain for circulation-related predictors was extended to 20W–40E and 35.5N–64.75N representing 3240 grid boxes on a  $0.75^\circ \times 0.75^\circ$  spatial resolution. Subsequently, only those PCs with  $\beta$ -coefficients significant at a level of  $\alpha=0.05$  were selected. For temperature (TMIN, TMEAN, TMAX), only the daily means of the  $3 \times 3$  grid box domain described in Section 2 (see Fig. 1) were used as additional predictor.

For comparison of the analog methods with linear regression, the Mean Square Error Skill Score (MSESS) was taken into account. As reference forecast, the seasonal mean temperatures of a 5-day moving window were taken into account. The derived values were centered to the mid-day of the window so that each day of the year is now represented by one specific temperature value for each temperature variable (TMIN, TMEAN, TMAX). Subsequently, the time series was extended by the number of years represented by the observation so that each time series provides the same number of days.

## 4 Results

### 4.1 Trends

In Table 1, the trends of the seasonal mean temperatures ( $\text{in}^\circ\text{C}/\text{a}$ ) of Augsburg-Mühlhausen (A-MH) and ERA5 are given for the quantile period (QP), the analog period (ANP), and the assessment period (ASP). It shows that TMAX provides rather the same trends within each period, whereas TMIN exhibits even a contrary development. Differences are highest for TMIN ( $0.0498^\circ\text{C}/\text{a}$ ) and TMAX ( $0.0031^\circ\text{C}/\text{a}$ ) of ANP and for TMEAN within ASP ( $0.0201^\circ\text{C}/\text{a}$ ), whereas the smallest differences are provided by ASP for TMAX ( $0.0011^\circ\text{C}/\text{a}$ ).

### 4.2 Evaluation of the analog selection methods

This section presents the results for the detrended analog time series. Results are not only shown with a special focus on temperature but also for different selection criteria. The criterion of the pre-selection steps which remains the same for all approaches is the selection by time of the year ( $\pm 15$  days of the day of interest). If there are no further selection criteria, the results in the following tables are marked as (-). If the exceedance of the 90th quantile is implemented as SC during the pre-selection step, results are marked as (Q90), whereas the results are marked as (T2M) when the temperature fields were used as second level of analogy. When both Q90 and T2M were used as selection criteria, results are marked as (Q90+T2M).

**Table 1** Trends and differences (DIFF) of the annual mean temperatures for TMIN, TMEAN, and TMAX of Augsburg-Mühlhausen (A-MH) and ERA5 for the quantile period QP, analog period ANP, and assessment period ASP (in °C/a)

	QP			ANP			ASP		
	TMIN	TMEAN	TMAX	TMIN	TMEAN	TMAX	TMIN	TMEAN	TMAX
A-MH	-0.0055	0.0309	0.0505	-0.0062	0.0298	0.0502	-0.0052	0.0285	0.0485
ERA5	0.0439	0.0496	0.0532	0.0436	0.0496	0.0533	0.0431	0.0486	0.0496
DIFF	0.0494	0.0187	0.0027	0.0498	0.0198	0.0031	0.0483	0.0201	0.0011

In order to check whether the analog time series represents the actual temperature and circulation adequately, the RMSE was calculated using the analog time series and the detrended original time series of ERA5. To avoid a weighting due to different scales, the RMSE was calculated and averaged by means of the standardized values. The mean overall RMSEs for the different methods (ANO, NVEC, COG), selection criteria (SC: -, T2M, Q90, Q90+T2M), as well as selection variables (SV: SLP, HGT.0850, HGT.0700, HGT.LT) are presented in Table 2. Table 2 shows for all methods that, independent of selection variable, the lowest RMSE is obtained by selection based on T2M. In general, method COG or SV SLP, respectively, provides the best results, and the combination of SC-T2M and SV SLP exhibits the overall lowest RMSE for method COG (0.3915).

In Fig. 3 the density plots of the RMSE are displayed under consideration of SV SLP for the respective selection variable and the temperatures TMIN, TMEAN, and TMAX. It shows

**Table 2** Root mean square error (RMSE) averaged over all standardized variables for different methods (ANO, NVEC, COG)

Method	Selection	SLP	HGT.0850	HGT.0700	HGT.LT
ANO	-	0.5983	0.8382	0.8345	0.5838
	T2M	0.5198	0.4542	0.4644	0.5334
	Q90	0.5554	0.7826	0.7911	0.5813
	Q90+ T2M	0.5656	0.5085	0.5149	0.5654
NVEC	-	0.6478	0.5911	0.5871	0.6576
	T2M	0.4525	0.4407	0.4434	0.4869
	Q90	0.6030	0.5700	0.5696	0.6450
	Q90+ T2M	0.4823	0.4679	0.4693	0.5100
COG	-	0.6435	0.6229	0.6003	0.6266
	T2M	0.3915	0.4355	0.4364	0.5190
	Q90	0.5841	0.5837	0.5775	0.6241
	Q90+ T2M	0.4208	0.4609	0.4623	0.5398

Analysis were performed by different selection criteria (-, T2M, Q90, Q90+T2M) and selection variables (SLP, HGT.0850, HGT.0700, HGT.LT)

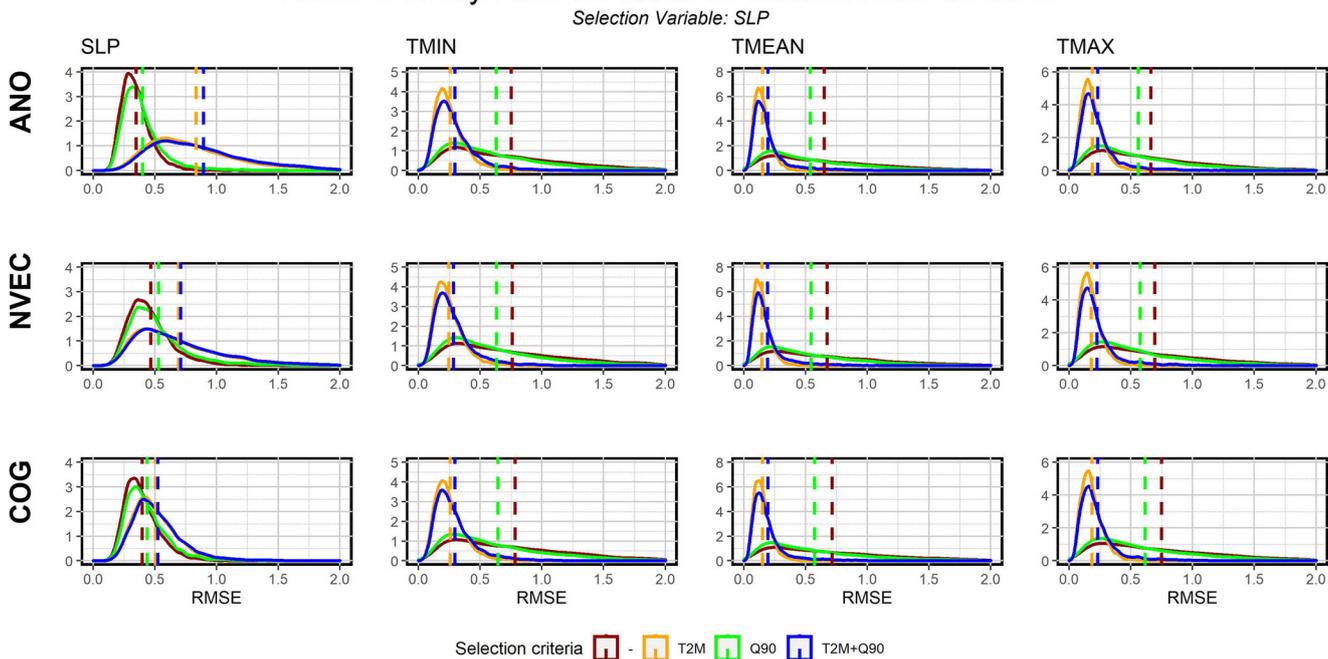
that, independent of method, the RMSE for SLP (temperature) is significantly higher (lower) when SC T2M (yellow and blue) is taken into account. In general, when the first level of analogy is used for reducing the initial subsample and final selection is based on temperatures, the agreement between the temperatures (circulation) of the observation and the analog time series is better (worse) than using the first level of analogy as final selection process. But it is striking that the improvement of the temperatures assessments under consideration of T2M as second level of analogy is much higher than the decline of the skill of the circulation variable(s). This development can also be observed for all SV but, as expected, with smallest differences and on a higher RMSE level for SV HGT.LT (not shown) since layer thickness already contains information about temperature. Thus, considering temperatures within the selection process improves the overall model performance significantly. Due to the fact that SLP exhibits the lowest overall RMSE, the following tables and figures are based on SV SLP, but Table 6 summarizes the best selection variable and criteria with respect to each of the following analyzing steps.

### 4.3 Analog temperature time series

In order to generate the temperature time series for the analogs, the detrended temperatures of Augsburg-Mühlhausen were extracted for each day selected by the respective analog method. Subsequently, the trends derived from the ERA5 re-analysis data set for period ASP were added to the detrended analog time series of Augsburg-Mühlhausen. In Table 3, the RMSE for the different absolute temperature variables are given for all methods and SC under consideration of SV SLP.

Table 3 shows that the best performance is achieved by method ANO when T2M was not considered as selection criteria, whereas NVEC provides the best skills, otherwise. In accordance with Table 2, SC considering temperatures outperform SC without, and SC-T2M provides the lowest overall RMSE for each analog method. However, the method which achieves the best overall performance within the model evaluation process (COG) is rather limited in reproducing the original temperature time series. It is also noticeable that the

### RMSE Density Plots for different variables and methods



**Fig. 3** Density plots (solid lines) and mean (dashed lines) of the RMSE for selection variable SLP and the temperatures (TMIN, TMEAN, TMAX) based on the different analog methods (ANO, NVEC, COG).

The density plots are given for the different selection criteria: no criterion (red), T2M (orange), Q90 (green), and Q90+T2M (blue)

RMSE of TMIN for methods based on SC without temperatures is lower than the RMSE of TMAX for the same setup although the trends of TMIN exhibit a contrasting development. Nevertheless, due to the diverging trends, the overall performance of TMIN is significantly worse than for both other temperature variables.

**Table 3** Root mean square error (RMSE) of TMIN, TMEAN, and TMAX (absolute) for the different analog methods (ANO, NVEC, and COG) and selection criteria (-, T2M, Q90, Q90+T2M) under consideration of selection variable SLP

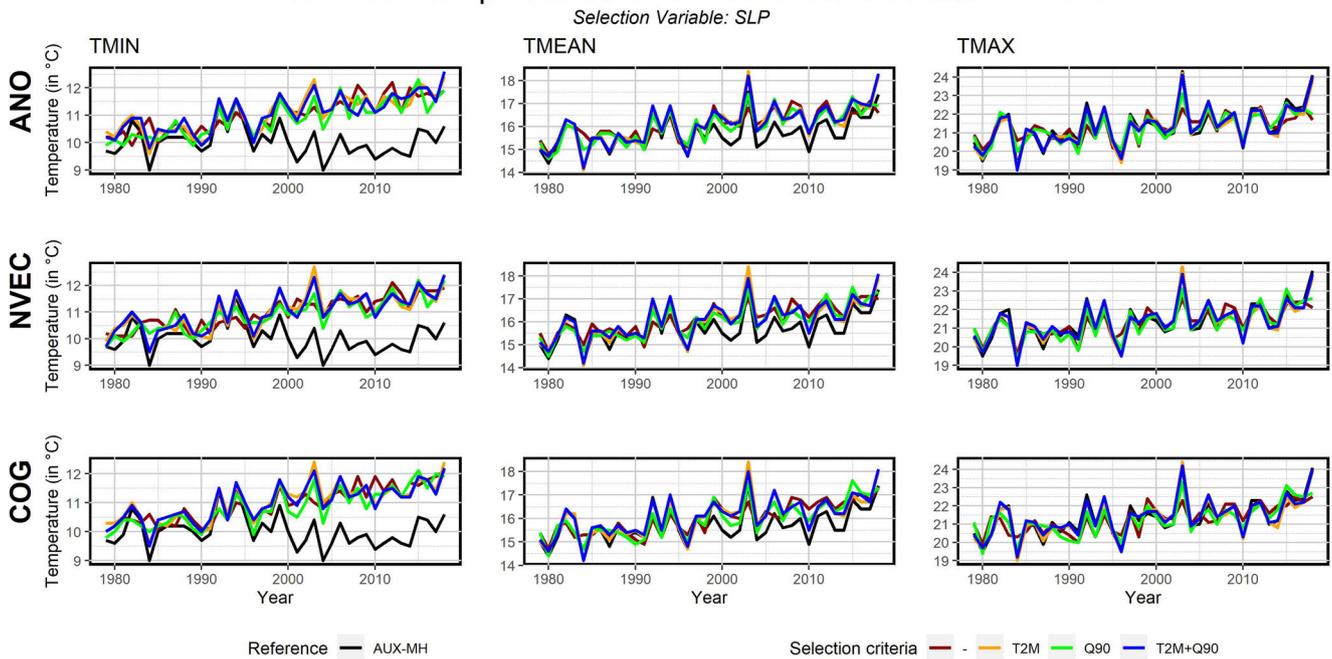
Method	Selection	TMIN	TMEAN	TMAX	$\emptyset$
ANO	-	4.2373	3.6048	4.6259	4.1593
	T2M	2.2544	1.2199	1.4458	1.6400
	Q90	3.5371	2.9393	3.8610	3.4458
	Q90+T2M	2.2589	1.3395	1.6988	1.7657
NVEC	-	4.2627	3.7387	4.8208	4.2741
	T2M	2.2040	1.1616	1.4005	1.5887
	Q90	3.5662	2.9720	3.9265	3.4882
	Q90+T2M	2.1714	1.3070	1.6890	1.7225
COG	-	4.2941	3.8896	5.1421	4.4419
	T2M	2.2520	1.1772	1.4268	1.6187
	Q90	3.6176	3.1206	4.1939	3.6440
	Q90+T2M	2.2004	1.3215	1.7007	1.7409

The seasonal mean temperatures of TMIN, TMEAN, and TMAX for all analog time series based on selection variable SLP are displayed in Fig. 4 for each SC (colors). Additionally, the seasonal mean temperatures are also given for the observation (black). Figure 4 shows that the best match between seasonal temperatures of the analog time series and Augsburg-Mühlhausen is given for TMAX. Accordance between the analog methods and observation is also given for TMIN until the mid-1990s. In the following years, the direction of change is similar, but the magnitude of change is lower for the observation. Thus, a strong discrepancy between the analogs and observation can be recorded during the second half of the time series for TMIN. These differences arise from the diverging trend of the ERA5 data set compared to the local observational trend at Augsburg-Mühlhausen. Since TMIN and TMAX both affect the daily mean temperature, the skill of TMEAN lies in between TMIN and TMAX but closer to TMAX.

#### 4.4 Exceedance of Q90 for analog time series

In order to evaluate the analog time series with respect to the exceedance of the 90th quantile, the daily Q90 values of the observation were used for all data sets. If the temperature exceeds the daily Q90 threshold, the day was flagged as 1, otherwise as 0. Subsequently, the binary time series of the observation and analog methods were analyzed by means of the Brier Score (BS). The BS is a proper scoring function in order to determine the quality of a model which estimates

### Annual mean temperatures for different variables and methods



**Fig. 4** Annual mean temperatures (TMIN, TMEAN, TMAX) of the analog time series for the different analog methods (ANO, NVEC, COG) based on selection variable SLP. The colors depict the different

selection criteria without (red), T2M (orange), Q90 (green), and Q90+T2M (blue). The black line represents the temperature of the observation at Augsburg-Mühlhausen weather station

whether or not an event occurs. Table 4 shows that the overall best performance for each analog method occurs when SC-Q90+T2M was taken into account and SC considering Q90

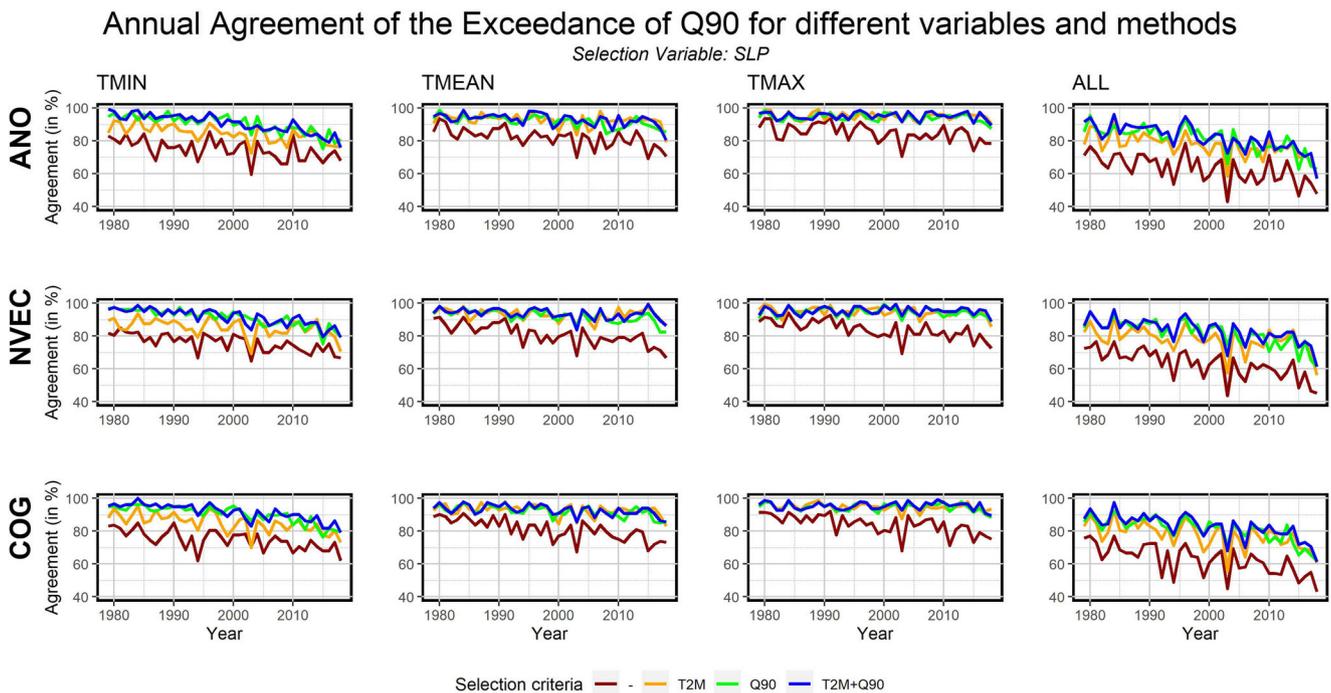
outperforms methods without. Under consideration of the analog approaches based on Q90 as selection criteria, the analog approach NVEC exhibits the lowest BSs for SC-Q90 (0.0742) and SC-Q90+T2M (0.06609). However, in contrast to the different SC, the differences between the analog methods are rather small. The individual examination of the temperatures shows that for SC without (with), Q90 assessments are most consistent for TMAX (TMIN). For SC-Q90+T2M, analog method COG exhibits the lowest BS for TMIN and TMAX but a significantly higher BS for TMEAN so that the overall performance is slightly worse than analog method NVEC.

**Table 4** Brier Score for the exceedance of Q90 of TMIN, TMEAN, and TMAX as well as the overall mean ( $\emptyset$ ) for the different analog methods (ANO, NVEC, and COG) and selection criteria (-, T2M, Q90, Q90+T2M) under consideration of selection variable SLP

Method	Selection	TMIN	TMEAN	TMAX	$\emptyset$
ANO	-	0.2523	0.1806	0.1498	0.1942
	T2M	0.1549	0.0747	0.0510	0.0935
	Q90	0.0962	0.0812	0.0582	0.0785
	Q90+T2M	0.0936	0.0706	0.0484	0.0709
NVEC	-	0.2407	0.1913	0.1605	0.1975
	T2M	0.1484	0.0699	0.0523	0.0902
	Q90	0.0887	0.0784	0.0556	0.0742
	Q90+T2M	0.0851	0.0650	0.0479	0.0660
COG	-	0.2516	0.1892	0.1583	0.1997
	T2M	0.1510	0.0724	0.0516	0.0917
	Q90	0.0969	0.0784	0.0528	0.0760
	Q90+T2M	0.0848	0.0711	0.0466	0.0675
MLR	SLP+T2M	0.0858	0.0493	0.0299	0.0550

Results were also presented for a multiple linear regression (MLR) based on the same predictor set (last row).

Figure 5 shows the annual agreement between analog time series and observation for the exceedance of Q90 for TMIN, TMEAN, TMAX, and the joint temperature variables. For the latter, eight categories are available, from no temperature variable exceeding Q90 (0|0|0) to all temperatures exceeding Q90 (1|1|1). The agreement is calculated as days which match the joint exceedance of the observation divided by the total number of days and is depicted in percent. In accordance with previous results, the assessments of the exceedance of Q90 exhibit the worst seasonal agreement for TMIN, and agreement is getting worse over time. The same development can also be observed for TMEAN and TMAX under consideration of no SC as well as for the joint analysis of Q90 exceedances. For all other SC, the agreement of TMAX (TMEAN) is predominantly within the range of 90–100% (80–100%). A general decrease of the agreement is indicated for TMEAN at the end of the time series but is hardly noticeably for TMAX. For



**Fig. 5** Agreement of the exceedance of Q90 (in %) for TMIN, TMEAN, TMAX, and the joint temperatures (ALL) between the analog time series and the observation at Augsburg-Mühlhausen weather station for the

different analog methods (ANO, NVEC, COG). The colors depict the different selection criteria without (red), T2M (orange), Q90 (green), and Q90+T2M (blue)

the joint analysis of Q90 exceedances (Fig. 5, right column), an agreement of more than 80% can be observed for most of the years until 2014 under consideration of any SC. From 2014 to 2018, the agreement decreases rapidly so that the agreement is over 60% in 2018 for the analog methods with any selection criteria and below 50% for the analog method without. Since the trends of TMIN in reanalysis and observation drift apart, more analog days exhibit an exceedance of Q90 for TMIN at the end of the time series. This fact also concerns days where the exceedance of Q90 of TMEAN and/or TMAX of the models are in accordance with the observation. Thus, even perfect models for TMEAN and TMAX cannot prevent an increasing mismatch between model and observation when the time series of TMIN exhibits different trends. On average, with respect to the entire time series, all analog methods exhibit an agreement of over 80% for SC-Q90 and SC-Q90+T2M with highest values for the analog method NVEC under consideration of SC-Q90+T2M (83.2%).

#### 4.5 Check against multiple linear regression

In Table 5, the MESSs of the daily temperature assessments are presented for all analog methods and the results of the linear regression based on SV-SLP. It shows that the overall performance of all analog methods is worse than the multiple linear regression. The averaged MESS of the linear regression (0.8785) is 13.8% higher than the best performance of the analog methods with T2M as SC. However, with respect to

TMIN, diverging trends also affect model skill of the MLR, but the impact is not as big as for the analog methods.

**Table 5** Mean Square Error Skill Score (MESS) of TMIN, TMEAN, and TMAX as well as the overall mean ( $\emptyset$ ) for the different analog methods (ANO, NVEC, and COG) and selection criteria (-, T2M, Q90, Q90+T2M) under consideration of selection variable SLP

Method	Selection	TMIN	TMEAN	TMAX	$\emptyset$
ANO	-	-1.0710	-0.2219	-0.0228	-0.4386
	T2M	0.4165	0.8601	0.9001	0.7256
	Q90	-0.4363	0.1876	0.2875	0.0129
	Q90+T2M	0.4142	0.8313	0.8621	0.7025
NVEC	-	-1.0860	-0.3143	-0.1108	-0.5037
	T2M	0.4423	0.8731	0.9062	0.7405
	Q90	-0.4600	0.1694	0.2631	-0.0009
	Q90+T2M	0.4587	0.8394	0.8636	0.7206
COG	-	-1.1169	-0.4225	-0.2638	-0.6011
	T2M	0.4178	0.8697	0.9027	0.7301
	Q90	-0.5024	0.0843	0.1593	-0.0863
	Q90+T2M	0.4442	0.8358	0.8617	0.7139
MLR	SLP+T2M	0.7362	0.9376	0.9616	0.8785

Results were also presented for a multiple linear regression (MLR) based on the predictor set SLP and T2M of ERA5 (last row).

With respect to the exceedance of Q90, Table 4 also shows that the overall performance of the MLR is better than all analog approaches, but in comparison to the MSESS of the temperature time series (Table 5), the difference is rather small (0.0110). In general, this confirms the thesis of Wilby et al. (2004) that with respect to extremes, analog models provide good skills. In Fig. 6, the annual agreement for the exceedance of Q90 between observation and model setup is presented. It shows that under consideration of TMIN, the analog approach exhibits higher agreements for the exceedance of Q90 especially in the beginning of the time series. On average, the performance of the analog approach for TMIN (91.5%) is slightly better than the MLR (91.4%). In contrast, the MLR exhibits higher values for TMEAN (+1.6%) and TMAX (+1.8). But with few exceptions, the analog approach achieves consistently values of 90% or higher for both temperature variables. Overall, the exceedance of the joint temperature variables is correctly reproduced in 85.9% (MLR) or 83.2% (NVEC), respectively. Thus, with respect to temperature and the exceedance of Q90, the MLR provides better skills than any of the analog approaches.

### 4.6 Overall performance

In Table 6, the best selection criteria and variables are summarized for each analog method. The skills (SKILL) are based on the same skills given in Sections 4.2, 4.3, and 4.4, i.e., RMSE of the standardized variables (All Var), RMSE of the absolute temperatures (T2M), and the Brier Score (Q90). It shows that for the assessment of all variables and temperature,

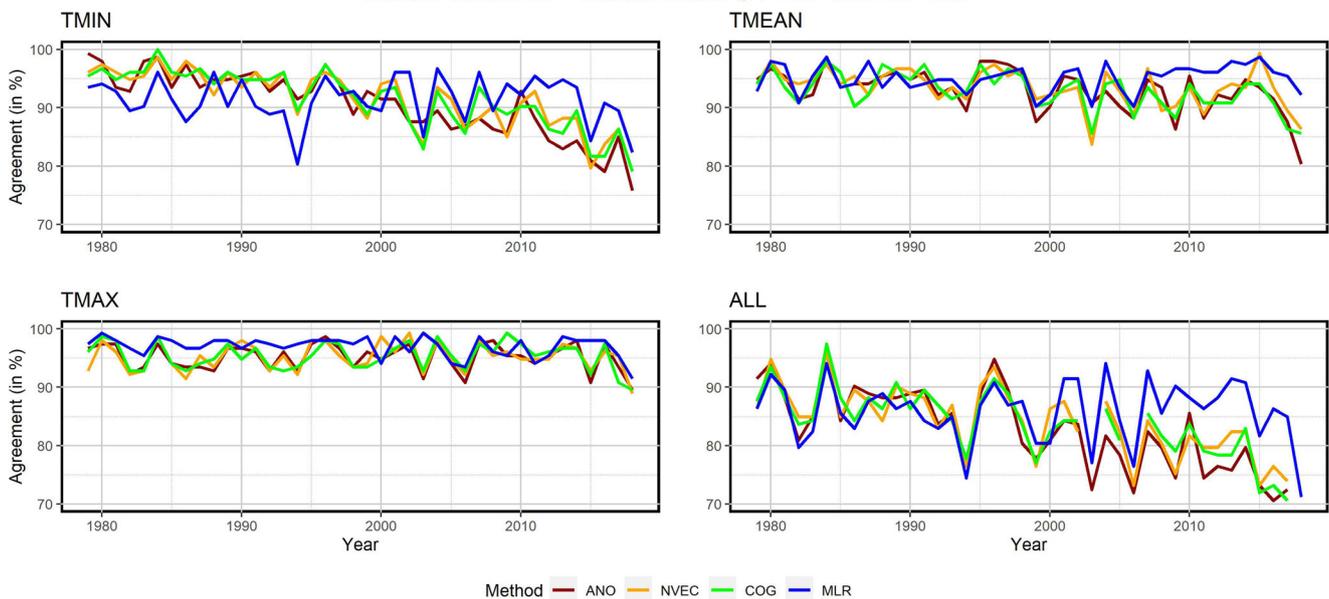
**Table 6** Summary of the best-performing selection variables and selection criteria for the different analog methods (ANO, NVEC, COG) and target variables (TV, All Var, assessing all considered variables; T2M, assessing temperatures; Q90, assessing the exceedance of Q90)

Method	TV	SV	SC	SKILL
ANO	All Var	HGT.0850	T2M	0.4542
	T2M	HGT.0850	T2M	1.6020
	Q90	HGT.0850	Q90+ T2M	0.0656
NVEC	All Var	HGT.0850	T2M	0.4407
	T2M	HGT.0700	T2M	1.5770
	Q90	HGT.0850	Q90+ T2M	0.0657
COG	All Var	SLP	T2M	0.3915
	T2M	HGT.LT	T2M	1.5606
	Q90	HGT.LT	Q90+ T2M	0.0641

SC-T2M provides the best results, whereas SC-Q90+T2M performs best for the assessment of the exceedance of Q90. With respect to the selection variable, HGT.0850 performs best for method ANO and for all variables and Q90 for method NVEC. For COG, HGT.LT provides the best skills for assessing temperatures and the exceedance of Q90, whereas SLP provides better skills for all variables. For all variables and overall temperature assessments, the consideration of Q90 during the pre-selection process is redundant, and only a selection by temperature during the second level of analogy is recommended. If the exceedance of the 90th quantile is one

### Agreement between model and observation of the exceedance of Q90 for different methods

Selection Variable: SLP - Selection Criteria: Q90+T2M - MLR: SLP+T2M



**Fig. 6** Annual agreement for the exceedance of Q90 between model and observation (TMIN, TMEAN, TMAX). The analog method NVEC with SC-Q90+T2M and SV SLP is depicted in blue and the multiple linear regression based on SLP+T2M in green

aspect of the analysis, Q90 should be taken into account during the pre-selection process. Overall, the differences between all analog methods are almost negligible, but, with respect to any target variable, method COG outperforms both other methods.

## 5 Summary and conclusions

The present study presents two novel analog approaches for assessing temperatures and temperature extremes (Q90) for Augsburg-Mühlhausen weather station. The selection process based on the circulation (first level of analogy) is depending on the inclination of the normal vector to the perpendicular of an even surface or the center of gravity, respectively. In combination with the mean value of the circulation patterns, the normal vector/center of gravity was used to calculate points within an artificial coordinate system. The day(s) which exhibit(s) the least distance between the artificial points of the day(s) and the day of interest represent(s) the analog day (sub-sample). Furthermore, different selection criteria were implemented during the pre-selection or second level of analogy to improve the model skills. Finally, the results were compared to another analog method and a multiple linear regression.

Results have shown that independent of target variable(s), selection criteria (SC), or variable (SV), all analog methods provide similar skills. With respect to the different validation skills, the two novel approaches can keep up with an already existing analog method (ZORITA & VON STORCH 1999). But in comparison with a multiple linear regression, all analog methods exhibit a weaker performance. Thus, the clearness and easy interpretability of analogs comes at a price of somewhat lower skill. Overall, under consideration of SV SLP, the best analog method (NVEC SC T2M) achieves an 13.5% lower MSESS for the assessments of temperatures and 0.011 higher Brier Score for the assessments of the exceedance of Q90 (NVEC SC T2M+Q90). For the exceedance of Q90 of TMIN, both novel approaches provide even better skills than the MLR, although the trends for TMIN of Augsburg-Mühlhausen and ERA5 are different in signs. Since the trends directly contribute to the assessments of the analog approaches, differences could not be attenuated. In general, the longer the time series, the more different trends between observation and reanalysis will affect the quality of the assessments.

When applying the novel analog approaches, some aspects have to be kept in mind: both methods focus on a regional to local scale. This means that the domain for the circulation related selection variables should be chosen very carefully, i.e., not too many centers of variation should influence the orientation of the normal vector or position of the center of gravity. Only when the domain size is small enough that the derived values of the determinants can exclusively be assigned

to a specific state of the atmosphere; both methods are appropriate for selecting analogs. For example, if there are two pronounced high-pressure systems located on the northern and southern border of the domain and one low-pressure system in the center of the domain, the orientation of the normal vector could be very similar to a circulation pattern where no significant pressure gradient could be detected since the different pressure centers cancel each other. Here, the domain size is limited to  $20^\circ \times 20^\circ$ , and, thus, the chances that different circulation patterns exhibit the same NVEC or COG are minimized. Only when a pressure system is located in the middle of the domain that a special situation occurs. By means of the orientation of the normal vector, there is no chance to determine whether it is a high- or low-pressure system. For this reason, the mean value of the pressure pattern was additionally considered in order to determine the point within the artificial coordinate system. In general, the domain size has to be defined in a way that the orientation of the normal vector or position of the center of gravity, respectively, in combination with the mean value can capture the different states of the atmosphere.

As the present study showed, limitations can occur when the trends of the observation and of the reanalysis data set exhibit strong differences or different signs. In this study, the daily minimum temperatures showed inhomogeneous behavior around the turn of the millennium with the result that observation and reanalysis are different in sign and, therefore, assessments drift apart over time. But, as could be observed for the MLR, inhomogeneous behavior of predictand or predictor variables is in general a problem for statistical downscaling approaches. The assumption that the relationships are robust and stationary over time is one important limiting factor for statistical downscaling (Gutierrez et al., 2013) but, with respect to climate change studies, can only be assumed. However, both novel analog methods provide good skills when no non-stationarity affects the time series (TMEAN, TMAX).

In summary, the two introduced novel analog approaches are straightforward methods for assessing absolute temperatures or exceedances of temperature thresholds. Furthermore, both methods are transferable on independent data sets and, thus, easily applicable on future climate change studies. Overall, both methods can keep up with an existing analog method and, to some extent, an approach based on multiple linear regression. Since each normal vector or center of gravity should represent a singular pressure pattern only, both approaches are especially suitable for local to regional climate studies. The study focused on urban climate change within the city of Augsburg, and thus the method was only applied on one city. Further analysis is necessary to see if the methods are also applicable to other cities or climate zones, respectively.

**Author contribution** Conceptualization: Christian Merckenschlager. Methodology: Christian Merckenschlager. Formal analysis and investigation: Christian Merckenschlager and Stephanie Koller. Writing—original draft preparation: Christian Merckenschlager. Writing—review and editing: Christoph Beck and Elke Hertig. Funding acquisition: Christoph Beck and Elke Hertig. Supervision: Christoph Beck and Elke Hertig.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under project HE6186/5-1 as well as under project number BE 2406/3-1.

**Availability of data** All data are freely accessible from the data providers mentioned in Section 2 of the manuscript ERA5: [www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5](http://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5) Augsburg-Mühlhausen weather station (DWD): <https://cdc.dwd.de/portal/>

**Code availability** All analyses were performed using the free software environment for statistical computing and graphics R.

## Declarations

**Ethics approval** All authors comply with the guidelines of the journal *Theoretical and Applied Climatology*.

**Consent to participate** All authors agreed to participate in this study.

**Consent for publication** All authors agreed to the publication of this study.

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alexandersson H (1986) A homogeneity test applied to precipitation data. *J Climatol* 6(6):661–675
- Bueno B, Norford L, Hidalgo J, Pigeon G (2013) The urban weather generator. *J Build Perform Simul* 6(4):269–281
- Buishand TA (1982) Some methods for testing the homogeneity of rainfall records. *J Hydrol* 58:11–27
- Copernicus Climate Change Service (C3S) (2017) ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Retrieved from Copernicus Climate Change Service Climate Data Store (CDS). <https://cds.climate.copernicus.eu/cdsapp#!/home>
- Caillouet L, Vidal JP, Sauquet E, Graff B (2016) Probabilistic precipitation and temperature downscaling of the Twentieth Century Reanalysis over France. *Clim Past* 12:635–662
- Castellano CM, DeGaetano AT (2016) A multi-step approach for downscaling daily precipitation extremes from historical analogues. *Int J Climatol* 36(4):1797–1807
- Castellano CM, DeGaetano, AT (2017) Downscaling extreme precipitation from CMIP5 simulations using historical analogs. *J Appl Meteorol Climatol* 56(9):2421–2439
- Chattopadhyay A, Nabizadeh E, Hassanzadeh P (2020) Analog forecasting of extreme-causing weather patterns using deep learning. *J Adv Model Earth Syst* 12(2) e2019MS001958
- Daoud AB, Sauquet E, Bontron G, Obled C, Lang M (2016) Daily quantitative precipitation forecasts based on the analogue method: Improvements and application to a French large river basin. *Atmos Res* 169:147–159
- Dayon G, Boé J, Martin E (2015) Transferability in the future climate of a statistical downscaling method for precipitation in France. *J Geophys Res Atmos* 120(3):1023–1043
- Delle Monache L, Nipen T, Liu Y, Roux G, Stull R (2011) Kalman filter and analog schemes to postprocess numerical weather predictions. *Mon Weather Rev* 139(11):3554–3570
- Diaz HF, Wahl ER, Zorita E, Giambelluca TW, Eischeid JK (2016) A five-century reconstruction of Hawaiian Islands winter rainfall. *J Clim* 29(15):5661–5674
- Enke W, Deuschländer T, Schneider F, Küchler W (2005) Results of five regional climate studies applying a weather pattern based downscaling method to ECHAM4 climate simulation. *Meteorol Z* 14(2):247–257
- Früh B, Becker P, Deuschländer T, Hessel JD, Kossmann M, Mieskes I, Namyslo J, Roos M, Sievers U, Steigerwald T, Turau H, Wienert U (2011) Estimation of climate-change impacts on the urban heat load using an urban climate model and regional climate projections. *J Appl Meteorol Climatol* 50(1):167–184
- Gutierrez J, San-Martin D, Brands S, Manzanar RH (2013) Reassessing statistical downscaling techniques for their robust application under climate change conditions. *J Clim* 26(1):171–188
- Hamdi R, Van de Vyver H, De Troch R, Termonia P (2014) Assessment of three dynamical urban climate downscaling methods: Brussels's future urban heat island under an A1B emission scenario. *Int J Climatol* 34(4):978–999
- Hertig E, Jacobeit J (2008) Assessments of Mediterranean precipitation changes for the 21st century using statistical downscaling techniques. *Int J Climatol* 28(8):1025–1045
- Ho HC, Knudby A, Sirovyak P, Xu Y, Hodul M, Henderson SB (2014) Mapping maximum urban air temperature on hot summer days. *Remote Sens Environ* 154:38–45
- Horton P, Jaboyedoff M, Obled C (2017) Global optimization of an analog method by means of genetic algorithms. *Mon Weather Rev* 145(4):1275–1294
- Ketterer C, Matzarakis A (2015) Comparison of different methods for the assessment of the urban heat island in Stuttgart, Germany. *Int J Biometeorol* 59(9):1299–1309
- Lauwaet D, De Ridder K, Saeed S, Brisson E, Chatterjee F, van Lipzig NP et al (2016) Assessing the current and future urban heat island of Brussels. *Urban Clim* 15:1–15
- Le Bras J, Masson V (2015) A fast and spatialized urban weather generator for long-term urban studies at the city-scale. *Front Earth Sci* 3: 27
- Lutz J, Gerstengarbe FW (2015) Improving seasonal matching in the STARS model by adaptation of the resampling technique. *Theor Appl Climatol* 120(3–4):751–760
- Makido Y, Shandas V, Ferwati S, Sailor D (2016) Daytime variation of urban heat islands: the case study of Doha, Qatar. *Clim* 4(2):32
- Maraun D (2016) Bias correcting climate change simulations—a critical review. *Curr Clim Chang Rep* 2(4):211–220

- Maraun D, Widmann M, Gutiérrez JM (2019) Statistical downscaling skill under present climate conditions: a synthesis of the VALUE perfect predictor experiment. *Int J Climatol* 39(9):3692–3703
- Marty R, Zin I, Obled C, Bontron G, Djerboua A (2012) Toward real-time daily PQPF by an analog sorting approach: application to flash-flood catchments. *J Appl Meteorol Climatol* 51(3):505–520
- Ohlwein C, Wahl ER (2012) Review of probabilistic pollen-climate transfer methods. *Quat Sci Rev* 31:17–29
- Orlowsky B, Gerstengarbe FW, Werner PC (2008) A resampling scheme for regional climate simulations and its performance compared to a dynamical RCM. *Theor Appl Clim* 92(3–4):209–223
- Pettitt AN (1979) A non-parametric approach to the change-point detection. *Appl Stat* 28:126–135
- Preisendorfer RW (1988) In: Mobeley CD (ed) *Principal component analysis in meteorology and oceanography* (Vol. 425). Elsevier, Amsterdam
- Radanovics S, Vidal JP, Sauquet E, Daoud AB, Bontron G (2013) Optimising predictor domains for spatially coherent precipitation downscaling. *Hydrol Earth Syst Sci Eur Geosci Union* 17:4189–4208
- Rau M, He Y, Goodess C, Bárdossy A (2020) Statistical downscaling to project extreme hourly precipitation over the United Kingdom. *Int J Climatol* 40(3):1805–1823
- Salvati A, Palme M, Inostroza L (2017) Key parameters for urban heat island assessment in a Mediterranean context: a sensitivity analysis using the Urban Weather Generator model. In IOP Conference Series: Materials Science and Engineering (Vol. 245, No. 8, p. 082055)
- Straub A, Berger K, Breitner S, Cyrus J, Geruschkat U, Jacobeit J, Kühnbach B, Kusch T, Philipp A, Schneider A, Umminger R, Wolf K, Beck C (2019) Statistical modelling of spatial patterns of the urban heat island intensity in the urban environment of Augsburg, Germany. *Urban Clim* 29:100491
- Timbal B, Dufour A, McAvaney B (2003) An estimate of future climate change for western France using a statistical downscaling technique. *Clim Dyn* 20(7–8):807–823
- Van den Dool HM (1989) A new look at weather forecasting through analogues. *Mon Weather Rev* 117(10):2230–2247
- Wahl ER, Zorita E, Trouet V, Taylor AH (2019) Jet stream dynamics, hydroclimate, and fire in California from 1600 CE to present. *Proc Natl Acad Sci* 116(12):5393–5398
- Wilby, RL, Charles, SP, Zorita, E, Timbal, B, Whetton, P, & Mearns, LO (2004). Guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting material of the Intergovernmental Panel on Climate Change, available from the DDC of IPCC TG CIA, 27.
- Zorita E, Von Storch H (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *J Clim* 12(8):2474–2489

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.