

New Approach for Bias Correction and Stochastic Downscaling of Future Projections for Daily Mean Temperatures to a High-Resolution Grid

QIFEN YUAN AND THORDIS L. THORARINSDOTTIR

Norwegian Water Resources and Energy Directorate, and Norwegian Computing Center, Oslo, Norway

STEIN BELDRING, WAI KWOK WONG, AND SHAOCHUN HUANG

Norwegian Water Resources and Energy Directorate, Oslo, Norway

CHONG-YU XU

University of Oslo, Oslo, Norway

(Manuscript received 10 April 2019, in final form 1 September 2019)

ABSTRACT


In applications of climate information, coarse-resolution climate projections commonly need to be downscaled to a finer grid. One challenge of this requirement is the modeling of subgrid variability and the spatial and temporal dependence at the finer scale. Here, a postprocessing procedure for temperature projections is proposed that addresses this challenge. The procedure employs statistical bias correction and stochastic downscaling in two steps. In the first step, errors that are related to spatial and temporal features of the first two moments of the temperature distribution at model scale are identified and corrected. Second, residual space–time dependence at the finer scale is analyzed using a statistical model, from which realizations are generated and then combined with an appropriate climate change signal to form the downscaled projection fields. Using a high-resolution observational gridded data product, the proposed approach is applied in a case study in which projections of two regional climate models from the Coordinated Downscaling Experiment–European Domain (EURO-CORDEX) ensemble are bias corrected and downscaled to a $1 \text{ km} \times 1 \text{ km}$ grid in the Trøndelag area of Norway. A cross-validation study shows that the proposed procedure generates results that better reflect the marginal distributional properties of the data product and have better consistency in space and time when compared with empirical quantile mapping.

1. Introduction

Climate change impacts often are realized at local to regional scales, resulting in impact models such as hydrological models, forest growth models, and crop models requiring tailored information on future climate at fine spatial and temporal scales (IPCC 2014; Hanssen-Bauer et al. 2017). In particular, many of these impact models are conducted on a very fine spatial grid and at daily time scale (e.g., Beldring et al. 2003). Future climate information commonly derives from coupled atmosphere–ocean general circulation models (GCMs) that currently provide neither unbiased nor local-

regional-scale information. Regional climate models (RCMs), with a spatial resolution of 10–15 km, provide a partial bridge for the spatial scale gap. Although ensembles of RCMs are able to capture basic features of regional climate variability in space and time (Kotlarski et al. 2014), their output may still contain substantial errors, partly inherited from the driving GCM (Rummukainen 2010; Hall 2014).

Impact studies are generally performed by comparing results for a reference climate to those obtained under a projected future climate. Where high-resolution gridded climate input data are required, the reference results are commonly based on gridded data products derived from lower-dimensional observations such as a network of surface observation stations (e.g., Lussana et al. 2018a,b). These data products come with their own inherent biases that are difficult to correct due to a lack of data. For an accurate assessment of climate impact, one goal is thus

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Qifen Yuan, qiyu@nve.no

DOI: 10.1175/JAMC-D-19-0086.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy \(www.ametsoc.org/PUBSReuseLicenses\)](https://www.ametsoc.org/PUBSReuseLicenses).

to generate high-resolution realizations of future climate the properties of which differ from those describing the reference climate only in terms of the expected climate change between the two time periods. In particular, statistical aspects such as the space–time variability and dependence at the finer scale should be realistically represented (Wood et al. 2004; Beldring et al. 2008).

To generate tailored climate information for various impact studies, postprocessing methods are almost routinely performed on climate model outputs. In a recent monograph on the subject, Maraun and Widmann (2018) identify three classes of statistical postprocessing methods. *Model output statistics* (MOS) approaches apply a statistical transfer function between simulated and observed data and are employed for both bias correction and downscaling. Depending on the specific needs of the climate information user, a wide variety of such methods are in use, ranging from simple mean adjustment to flexible, potentially multivariate quantile mapping methods (Maraun et al. 2010; Piani and Haerter 2012; Vrac et al. 2012; Vrac and Friederichs 2015; Cannon 2016; Vrac 2018). For downscaling, *perfect prognosis* (PP) methods establish a statistical link between large-scale predictors and local-scale predictands typically in a regression framework, while *weather generators* (WGs) are stochastic models that explicitly model marginal and higher-order structures. WGs are widely used for generating weather time series at stations (Semenov and Barrow 1997), with some extensions to multisite (Wilks 1999, 2009) and multivariate (Kilsby et al. 2007) settings.

One common issue with MOS methods applied to downscaling is that they are not able to capture spatial and temporal variability at the finer scale (e.g., Maraun et al. 2017; Maraun and Widmann 2018). The transfer functions derived from the historical period are transformations of the stochasticity at the model scale that are often not realistic at the required finescale. Hence, including stochastic components into the bias-correction procedure is imperative to account for local-scale variability (Maraun et al. 2017). Recently proposed stochastic downscaling methods have proven skillful in modeling the small-scale variability of precipitation occurrence and intensity across sets of point locations (Wong et al. 2014; Volosciuk et al. 2017). The impact models considered in our application commonly require high-resolution gridded input and thus approaches that scale to high-dimensional and spatially coherent settings.

Further, it has been argued that methods based on PP assumptions where it is assumed that daily based coarse-scale information can be used to predict the probability distribution at the local-scale are not appropriate for

free-running model simulations such as the RCMs from the Coordinated Downscaling Experiment (CORDEX; Jacob et al. 2014). For full-field downscaling without PP assumptions, techniques of shuffling the time series produced by univariate bias correction have been proposed (e.g., “Schaake shuffle”; Clark et al. 2004), both for temporal (Vrac and Friederichs 2015) and for multisite and multivariate reordering (Vrac 2018). The shuffling techniques impose historical rank correlation structure on the bias-corrected data. They have, in some instances, been shown to underrepresent the dependence structure (Vrac 2018). Moreover, the size of the shuffled dataset is restricted to the size of the observational dataset.

Alternatively, multisite WGs that explicitly model the finescale stochasticity are able to generate spatially and temporally coherent fields and thus have shown potential for full-field downscaling (Wilks 2010, 2012). This approach, however, has been typically applied where parameters are calibrated first at single locations and then interpolated onto a grid consisting of a small set of grid points (Wilks 2009), which is not straightforward to work with when gridded data products are available and have been used to train the impact models. Besides, they are primarily constructed for generating daily precipitation, whereas daily mean temperature has its own properties (Huybers et al. 2014) and is an equally important input to, for example, hydrological models (Xu 1999). Our objective is thus to propose a full-field downscaling approach for daily mean temperature that explicitly accounts for the finescale variability and dependence in both space and time.

In this paper, we introduce a two-stage statistical postprocessing procedure that bias corrects and down-scales RCM simulations to a high-resolution grid where we combine MOS and WG to utilize the advantages of both approaches. In the first stage, a MOS approach is applied to bias correct RCM output at the model scale by comparing it against upscaled gridded data product. Daily mean temperatures are generally considered well represented by a Gaussian distribution (e.g., Piani et al. 2010). Here, we apply a transfer function where the parameters of the Gaussian distribution vary across space and time to account for seasonal and geographic changes in temperatures. Second, we construct a WG to simulate pseudo-observations that replicate the properties of a finescale gridded data product under a stationary climate. Using a separable space–time correlation structure, the method is able to efficiently generate high-dimensional realizations. We then impose these realizations with appropriate climate change signal derived from the RCM simulations. The approach can thus be thought of as a delta change method that preserves space–time consistency. For comparison, we also perform the

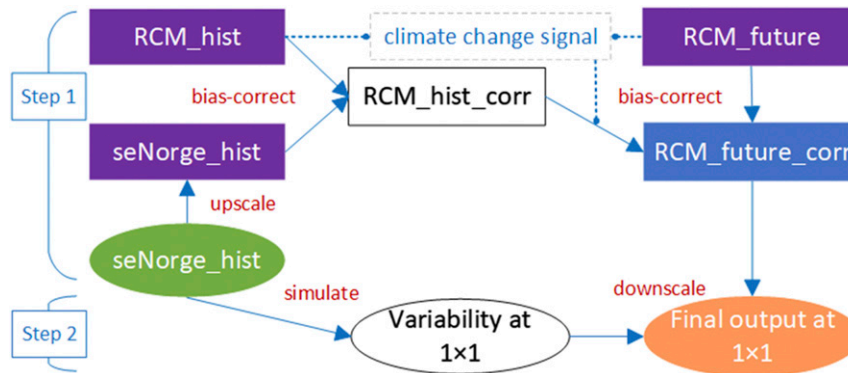


FIG. 1. Proposed general framework for postprocessing of climate model output.

analysis with the commonly used empirical quantile mapping (EQM) method (e.g., Piani et al. 2010; Gudmundsson et al. 2012), which is a MOS approach.

The remainder of the paper is organized as follows. In section 2 we first describe the data and the study area. A description of the proposed postprocessing procedure follows, as well as a brief discussion of evaluation methods. Results are presented in section 3, and section 4 provides discussion and conclusions.

2. Data and methods

We propose a two-step postprocessing approach for statistical bias correction and stochastic downscaling as demonstrated in Fig. 1. First, biases at the coarser RCM scale are identified and corrected, where the climate change signal simulated by RCM is preserved. Then, we estimate the space–time residual variability at the finer data product scale using a statistical model, and, by simulating from this model, we are able to generate a set of realizations of a stationary future climate possessing the same space–time structures as the historical data product. Based on these, we compare three different approaches for adding layers of the climate change signal to obtain the final output. A description of the data and study area is given below, followed by a detailed description of each step in the postprocessing procedure.

a. Data and study area

We apply our methodology to daily mean temperature simulations from two RCMs from the CORDEX–European Domain (EURO-CORDEX-11) ensemble. One combines the COSMO Climate Limited Area Model (CCLM) from the Potsdam Institute for Climate Research (Rockel et al. 2008) with boundary conditions from the CNRM-CM5 Earth system model (referred to as RCM1 in the following text) developed by the French National Centre for Meteorological Research (Voldoire

et al. 2013), whereas the other (referred to as RCM2) combines the CCLM model with boundary conditions from the MPI Earth system model developed by the Max Planck Institute for Meteorology (Giorgetta et al. 2013). The RCM simulations are conducted over the European domain at a spatial resolution of 0.11° or about 12.5-km grid resolution (Jacob et al. 2014). In the historical period up to 2005 the outputs are simulated on the basis of recorded emissions and are thus comparable to observed climate.

For observational reference data, we use the seNorge gridded data product, version 2.1, produced by the Norwegian Meteorological Institute (Lussana et al. 2018b; available at http://thredds.met.no/thredds/catalog/seNorge/seNorge2_1/TEMP1d/catalog.html). The data result from an optimal spatial interpolation method applied to measurements at a network of weather and climate stations with the number varying from 150 to 450 for the period from 1957 to the present and are available at a spatial resolution of 1 km over an area covering the mainland Norway and an adjacent strip along the Norwegian border. For bias correcting the RCM output, we upscale the seNorge data to the RCM grid by taking a weighted average over all seNorge grid cells found within each RCM grid cell, where the weights are area ratios of the seNorge cells to that RCM cell.

For the study area, we consider the Trøndelag area in central Norway; see Fig. 2. The area comprises 695 RCM grid cells and 109 514 seNorge grid cells. The bias correction is performed over the entire study domain while the statistical downscaling focuses on nine hydrological catchments within the domain (see Fig. 2 and Table 1). Two catchments Krinsvatn and Oeyungen have maritime climate while the rest have continental climate. For each catchment, the downscaling is performed over all seNorge grid cells within the RCM grid cells that cover the catchment. The spatial dimensions of the downscaling areas thus vary between approximately 940 and 5500 grid cells at 1-km resolution. Both historical RCM simulations

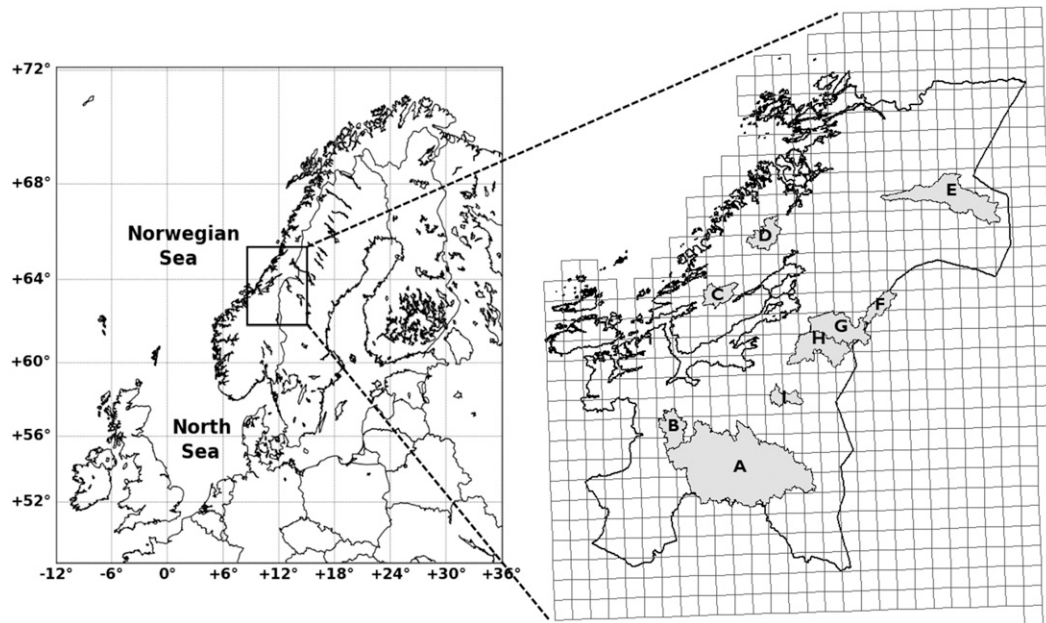


FIG. 2. Location of the study area—Trøndelag in central Norway. For the RCM bias correction, the entirety of Trøndelag and a small part of neighboring Sweden with a total area covering by 695 RCM grid cells (rectangular-like polygons) and 109 514 seNorge grid cells (within the polygons; not shown) are used. For the stochastic downscaling, nine hydrological catchments within Trøndelag with catchment areas from 143 to 3086 km² (shaded in gray) are used; see also Table 1.

and seNorge observations are available over the time period 1957–2005. We use the time period 1957–86 as a training period to estimate the parameters of the postprocessing approaches and perform an out-of-sample evaluation over the remaining 19 years 1987–2005. As a result, the training period consists of 10 950 days and the test period comprises 6935 days.

In addition, we use explanatory variables, or covariates, to describe the spatial variations in the statistical characteristics of the daily mean temperature distributions. We consider latitude, longitude, and elevation as potential geographic covariates. Elevation information for the seNorge data is obtained from a digital elevation model that is based on a 100-m-resolution terrain model from the Norwegian Mapping Authority (Mohr 2009). We upscale these data in the same manner as the daily mean temperatures to obtain the elevation at the RCM scale. Note that this is not equal to the orography information provided by EURO-CORDEX.

b. Bias correction

For bias correcting the RCM output, we perform weighted upscaling of the seNorge data product as described in section 2a. The RCM-simulated gridcell values represent area averages; the upscaled seNorge data should thus be comparable to the RCM output in distribution. We follow, for example, Piani et al. (2010)

and assume that temperature can be modeled by a Gaussian distribution. However, rather than modeling each month separately, the parameters of the distribution are assumed to change smoothly across time and space.

Specifically, denote by Y_{rt} the daily mean temperature in grid cell $r \in \{1, \dots, R\}$ at time $t \in \{1, \dots, T\}$, where R denotes the number of grid cells and T is the number of days in a given RCM-scale dataset. We then set

$$Y_{rt} \sim N(\mu_{rt}, \sigma_{rt}^2), \quad (1)$$

where

TABLE 1. Characteristics of the nine hydrological catchments in Trøndelag, Norway considered in the stochastic downscaling (ID = catchment identifier used herein).

Catchment	ID	Size (km ²)	Downscaling area (km ²)	Median elev (m MSL)
Gaulfoss	A	3086	5479	734
Aamot	B	283	1112	460
Krinsvatn	C	206	1108	349
Oeyungen	D	239	952	295
Trangen	E	852	2327	558
Veravatn	F	175	1101	514
Dillfoss	G	483	1863	506
Hoeggaas	H	495	1853	505
Kjeldstad	I	143	940	578

$$\mu_{rt} = f_1^\mu(\mathbf{c}_r) + f_2^\mu(t) + f_3^\mu(t) \tag{2}$$

$$\log(\sigma_{rt}) = f_1^\sigma(\mathbf{c}_r) + f_2^\sigma(t), \tag{3}$$

with

$$f_1^\zeta(\mathbf{c}_r) = \alpha_{11} + \alpha_{12}c_{r1} + \alpha_{13}c_{r2} + \alpha_{14}c_{r3}, \tag{4}$$

$$f_2^\zeta(t) = \alpha_{21} \cos\left[\frac{2\pi d(t)}{365}\right] + \alpha_{22} \sin\left[\frac{2\pi d(t)}{365}\right] + \alpha_{23} \cos\left[\frac{4\pi d(t)}{365}\right] + \alpha_{24} \sin\left[\frac{4\pi d(t)}{365}\right], \tag{5}$$

$$f_3^\mu(t) = \alpha_3 y(t), \tag{6}$$

for $\zeta \in \{\mu, \sigma\}$. Here, f_1 models the spatially varying baseline of the two moments with $\mathbf{c}_r = (c_{r1}, c_{r2}, c_{r3})$ being latitude, longitude, and mean elevation of grid cell r . Seasonal changes in the moments are captured by f_2 , where $d(t)$ returns the calendar day of time point t and f_3 describes potential linear trend in the first moment, with $y(t)$ returning the calendar year normalized so that α_3 describes the trend in degrees per decade.

The model specified in Eqs. (1)–(6) has 17 coefficients, that is, 9 coefficients for the first moment and 8 coefficients for the second moment. A two-step analysis of three RCM-scale datasets where the mean and the residuals were analyzed separately found all the coefficients significant. For the bias correction we estimate all 17 coefficients for each dataset simultaneously by numerically obtaining the maximum likelihood estimator (MLE) using the function “lmvar()” from the R (R Core Team 2018) package “lmvar” (Posthuma Partners 2018). Subsequently, we adjust the estimated model parameters for the RCM simulations in the out-of-sample test period based on the estimates from the upscaled seNorge data and the RCM data in the training period, and the RCM-simulated changes from the training period to the test period. In particular, the correction (“Corr”) is similar to Eq. (12.7) in Maraun and Widmann (2018), with two exceptions: First, the mean and standard deviation terms are estimates by Eqs. (2) and (3) varying across space and time and, second, the standardized RCM anomalies (i.e., dividing by its own standard deviation) are rescaled by the square root of the variance of the upscaled seNorge data plus the RCM-simulated change in the variances between the two periods. See the appendix for further details.

For comparison, we consider two simple bias-correction methods commonly used as benchmarks (e.g., Räisänen and Rätty 2013) where only the mean of the RCM output is corrected using one common correction term across the entire domain (“Simple”) or independently for each grid cell (“LocalSimple”). These methods explicitly

preserve the change in the long-term mean simulated by RCM; also see the appendix.

c. Stochastic downscaling

1) STATIONARY SPACE–TIME HIGH-RESOLUTION MODEL

We model the space–time variability at the finer 1-km scale by a stochastic model that assumes stationarity and space–time separability in the residuals. To warrant these rather strict assumptions—assumed for computational feasibility—we estimate the model independently for each catchment, allowing for, for example, changes in the space–time variability across different climatic zones.

Let X_{st} denote the daily mean temperature at time $t \in \{1, \dots, T\}$ in the training period and finescale grid cell $s \in \{1, \dots, S\}$ for a given catchment. We fit a model of the form given in Eqs. (1)–(6) above to this dataset and generate the corresponding residuals

$$Z_{st} = \frac{X_{st} - \hat{\mu}_{st}}{\hat{\sigma}_{st}}, \tag{7}$$

where $\hat{\mu}_{st}$ and $\hat{\sigma}_{st}$ denote the estimated mean and standard deviation at time point t and location s , respectively.

We assume that the residual field \mathbf{Z}_t varies around a mean value of zero, so that the time series $\bar{Z}_t = (1/S)\sum_{s=1}^S Z_{st}$ represents a reasonable approximation of the true temporal dependence. Although appearing stationary, the time series $\{\bar{Z}_t\}$ does not seem to have Gaussian marginals as shown in Fig. 3. In particular, the marginals have a positive skewness in the warmer months and a negative skewness in the colder months. We thus employ a copula approach to estimate the temporal correlation (Nelsen 2007) where we combine split normal marginals and an autoregressive moving average (ARMA) structure to account for both the marginal skewness and the temporal correlation. The split normal distribution (Wallis 2014), sometimes called the two-piece normal distribution, is a three parameter generalization of the normal distribution that allows for asymmetry in the tails in that a separate scale parameter is used for each of the two tails. For comparison, we have also investigated a simpler model with Gaussian marginals. However, this results in substantially reduced performance, see section 3b(1).

Next, we assume that the spatial residuals follow a multivariate normal distribution with a mean vector zero and a variance matrix specified by a stationary and isotropic covariance function of the exponential type so that the spatial correlation between two grid cells s and $s' \in \{1, \dots, S\}$ is determined by their Euclidean distance $\|s - s'\|$ (e.g., Cressie and Wikle 2015). That is, we define a residual model of the form

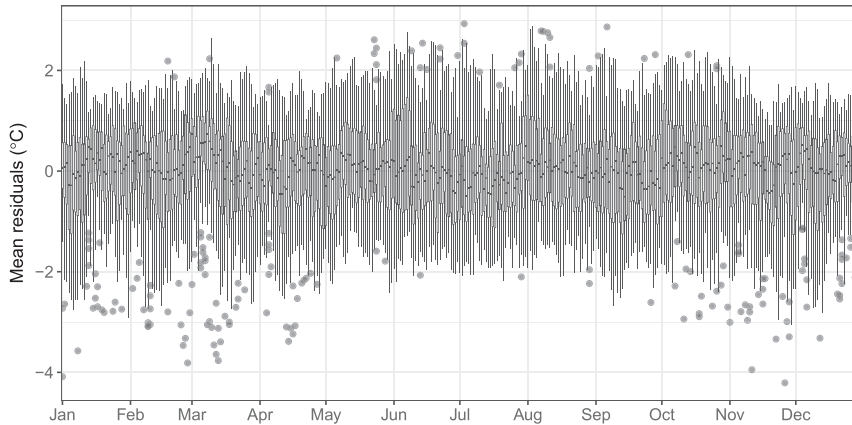


FIG. 3. Boxplots of the mean residuals $\bar{Z}_t = (1/S)\sum_{s=1}^S Z_{st}$ where the mean is taken over all grid cells in catchment A, which is the largest in our study area. The boxplot for each calendar day consists of all values for that calendar day in the training period 1957–86, with outliers denoted by dots.

$$Z_{st} = \eta_t + \nu_{st}, \tag{8}$$

$$\eta_t \sim \text{SN}(\mu_t, \sigma_{1t}, \sigma_{2t}), \tag{9}$$

$$U_t = \Phi^{-1}[F_{\text{SN}}(\eta_t)] \sim \text{ARMA}(p, q), \tag{10}$$

$$\nu_t \sim N(\mathbf{0}, \Sigma_t) \text{ and} \tag{11}$$

$$\text{Cov}(\nu_{st}, \nu_{s't'}) = \theta_{0t} \mathbb{1}\{\|s - s'\| = 0\} + \theta_{1t} \exp(-\|s - s'\|/\theta_{2t}). \tag{12}$$

Here, we denote the split normal distribution by SN and the normal distribution by N , F_{SN} stands for the split normal distribution function, Φ^{-1} denotes the quantile function of the normal distribution, Cov denotes the covariance function, and $\mathbb{1}\{\|s - s'\| = 0\}$ is the indicator function, which is equal to 1 if $\|s - s'\| = 0$ and 0 otherwise.

Daily-varying parameter estimates for the split normal distribution are obtained by numerically optimizing the likelihood function “logs_2pnorm()” from the R package “scoringRules” (Jordan et al. 2018). Subsequently, an ARMA model of order (p, q) is estimated using the function “auto.arima()” from the R package “forecast” (Hyndman et al. 2018). This approach searches for the best autoregressive integrated moving average (ARIMA) model that is reduced to an ARMA model in the case in which the time series is assumed to be stationary, here determined by an augmented Dickey–Full test. The method then determines the values of p and q and estimates the associated parameters. We found that $p = 2$ and $q = 3$ fits best for catchment A, $p = 3$ and $q = 2$ fits best for catchment F, and $p = q = 2$ works best for the rest.

To estimate the parameters of the covariance function given in Eq. (12), we employ semivariogram function given by

$$\gamma_{\theta_t}(h) = \theta_{0t} + \theta_{1t}[1 - \exp(-h/\theta_{2t})] \tag{13}$$

with $h = \|s - s'\|$. To account for potential seasonal changes in the spatial correlation structure, we obtain separate estimates for each month and, subsequently, fit a smooth function through each set of estimates to obtain smoothly changing daily estimates.

Denote by \mathcal{S} the set of all time points from a given month, with $|\mathcal{S}|$ being the number of days in this set, and denote by $\mathcal{S}(h)$ the set of gridcell pairs that have distances within some small interval approximately centered around h , with $|\mathcal{S}(h)|$ being the number of pairs in the set. We then estimate the covariance parameters by fitting the semivariogram function given by Eq. (13) to the empirical semivariogram

$$\hat{\gamma}(h, \mathcal{S}) = \frac{1}{2|\mathcal{S}(h)||\mathcal{S}|} \sum_{(s,s') \in \mathcal{S}(h)} \sum_{t \in \mathcal{S}} (\nu_{st} - \nu_{s't})^2. \tag{14}$$

Here, we employ the R package “spacetime” (Pebesma 2012) to organize our spatiotemporal residuals and “gstat” (Pebesma 2004; Gräler et al. 2016) to calculate empirical semivariograms and perform the fitting using decreasing weights on the pairs that are further apart.

Given parameter estimates, we can then simulate a set of residuals Z_{st}^* for $s \in \{1, \dots, S\}$ and $t \in \{1, \dots, T\}$ in four steps, where we use the asterisk indicator to denote simulated variables:

- 1) simulate $U_t^* \sim \widehat{\text{ARMA}}(p, q)$,
- 2) set $\eta_t^* = \hat{F}_{\text{SN}}^{-1}[\Phi(U_t^*)]$,
- 3) simulate $\nu_t^* \sim N(\mathbf{0}, \hat{\Sigma}_t)$, and
- 4) set $Z_{st}^* = \eta_t^* + \nu_{st}^*$.

Here, the split normal variables are simulated using “qsplnorm()” from the R package “fanplot” (Abel 2015)

and the multivariate normal simulation is carried out by “mvrnorm()” from the R package “MASS” (Venables and Ripley 2002).

2) ADDING A CLIMATE CHANGE SIGNAL

We obtain a realization of a stationary climate for the test period corresponding to the mean climate in the training period by setting $X_{st}^* = Z_{st}^* \hat{\sigma}_{st} + \hat{\mu}_{st}^*$, where Z_{st}^* is obtained with the simulation algorithm above, $\hat{\sigma}_{st}$ is the standard deviation estimate based on Eq. (3) and also used in Eq. (7), and $\hat{\mu}_{st}^*$ is the mean estimate based on Eq. (2) and also used in Eq. (7) without the trend component centered so that the average over all time points $\bar{\mu}_{st}^*$ fulfills $\bar{\mu}_{st}^* = \bar{\mu}_{st}$. We call this realization “Xstar,” and it serves as a reference for the other methods described below.

In the realization “XstarTrend” that is defined by $X_{st}^{\text{Trend}} = Z_{st}^* \hat{\sigma}_{st} + \hat{\mu}_{st}^{\text{Trend}}$, adjustments are made to the baseline and the linear trend components of the mean to reflect the RCM-simulated changes from the training period to the test period [see Eqs. (A5) and (A6) in the appendix]. Here, we assume that the long-term average changes at RCM scale directly carry over to the seNorge scale using the information from the RCM grid cell that has the largest intersection area with the seNorge grid cell. We further investigated adding RCM-simulated changes in the seasonality of the mean. However, this resulted in substantially reduced agreement between our model and the out-of-sample data.

We also generate the realization “XstarTrendVar” with $X_{st}^{\text{TrendVar}} = Z_{st}^* \hat{\sigma}_{st}^{\text{Var}} + \hat{\mu}_{st}^{\text{Trend}}$ in which adjustment is made to both the mean and the variance [see Eqs. (A5), (A6), and (A9) in the appendix]. While the mean is adjusted as before, the variance at the seNorge grid level is adjusted such that changes relative to the training period at the upscaled RCM grid level match those in the corrected RCM output.

d. Reference method

We compare the final results from our method with EQM (e.g., Piani et al. 2010; Gudmundsson et al. 2012), a widely adopted method for bias correcting and downscaling RCM outputs to a finer grid. The EQM method utilizes the empirical cumulative distribution function (eCDF) for variables at both scales. In a first step, we regrid the RCM output to the seNorge grid using a simple nearest-neighbor method. Then, we derive a transfer function matching the RCM-scale eCDF with the seNorge-scale eCDF. The eCDFs are approximated using tables of empirical percentiles with fixed interval of 0.1 spanning the probability space [0, 1]. Spline interpolation is performed for the values in between these percentiles and to extrapolate beyond the highest and lowest observed values. In the training period, we

derive 12 calendar-month-specific transfer functions for each seNorge grid cell. These transfer functions are assumed to be valid for use in the test period. And we apply them to adjust the RCM output quantile by quantile so that they yield a better match with the seNorge data. To perform the EQM, we employ the R package “qmap,” version 1.0–4 (Gudmundsson 2016).

e. Evaluation methods

We assess the performance of the postprocessing methods by comparing projections with out-of-sample data. We compare the marginal distributions in each grid cell using eCDFs over all time points in the test period, the temporal autocorrelation and the spatial correlation in each catchment.

We compare two marginal distributions F and G using the integrated quadratic distance (IQD; Thorarinsdottir et al. 2013),

$$\text{IQD}(F, G, \omega) = \int_{-\infty}^{+\infty} [F(x) - G(x)]^2 \omega(x) dx, \quad (15)$$

where ω denotes a nonnegative weight function that can be designed to focus on a particular part of the distributions. For assessing results at the RCM scale, we use the unweighted version with $\omega_1 \equiv 1$. For the finer scale, we consider four different weighting options, the unweighted version as well as three weights that focus on the tails and the center of the distributions. Specifically, we set

$$\omega_2(x) = 1\{x \geq G^{-1}(0.95)\}, \quad (16)$$

$$\omega_3(x) = 1\{G^{-1}(0.45) \leq x \leq G^{-1}(0.55)\}, \quad \text{and} \quad (17)$$

$$\omega_4(x) = 1\{x \leq G^{-1}(0.05)\}, \quad (18)$$

where, for example, $1\{x \geq u\}$ denotes the indicator function that is equal to 1 if $x \geq u$ and 0 otherwise and G denotes the data eCDF. A lower IQD value indicates a better correspondence between F and G . For each comparison, we report average IQD values across all grid cells, for example, in a catchment,

$$\frac{1}{S} \sum_{s=1}^S \text{IQD}(F_s, G_s, \omega), \quad (19)$$

together with uncertainty bounds obtained by bootstrapping.

For distributions F and G with finite first moments, the IQD is the score divergence of the continuous ranked probability score (CRPS), which is a proper scoring rule (Gneiting and Raftery 2007). It thus fulfills a similar propriety condition and can be used to rank competing

methods (Thorarinsdottir et al. 2013). In fact, using the IQD in Eq. (15) will result in the same model rankings as computing the average CRPS over all the observations in G . However, we find that using the IQD provides improved interpretability as the lowest possible IQD value is zero if $F = G$ while the lowest possible CRPS value depends on the unknown true data distribution.

At the seNorge scale, the assessment of temporal and spatial correlation structures is carried out separately in each catchment. For the temporal correlation, we aggregate the daily gridded data into a single time series and, subsequently, calculate the autocorrelation up to a certain lag using the function “Acf()” from the R package forecast (Hyndman et al. 2018). For the spatial correlation, we calculate the empirical semivariogram for each month using the same R functions as described in section 2c(1).

3. Results

a. Bias correction at model scale

The marginal performance of the three bias-correction methods at RCM model scale is shown in Fig. 4. There is notable difference in performance between the two raw RCM outputs with RCM2 more compatible with the upscaled seNorge data. A considerable decrease in the IQD from “Raw” to Simple means that both RCMs fail to capture the correct long-term average over the whole study area, as expected from free-running climate models. Additional improvement can be achieved by local mean correction (LocalSimple). For RCM1, the proposed bias-correction method Corr further improves the compatibility with the data product. For RCM2, however, Corr performs worse than LocalSimple, indicating that additional correction of the variance, the linear trend and seasonality of the mean has a slightly adverse effect for RCM2. The bootstrapped 90% confidence intervals for the mean IQD values are nonoverlapping for all comparisons, indicating that the differences in performance are significant.

To investigate this aspect further, consider the estimated mean baseline and trend for a single grid cell shown in Fig. 5. The upscaled data product has a slightly negative trend in the training period 1957–86 and a positive trend in the test period 1987–2005 with the overall mean temperature in the test period 0.9°C higher than that in the training period. While RCM1 has a baseline estimate that is around 2°C colder than the data product, the trend estimates of the two datasets are similar, resulting in a bias-corrected RCM output that is overall 0.55°C colder than the data product with a slightly slower warming rate. The raw output from RCM2, on the other hand, has opposite trends compared to the data product in both time periods, resulting

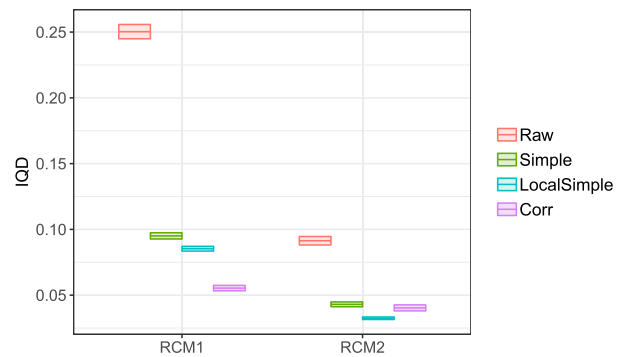


FIG. 4. Marginal performance of RCM raw output and three bias-correction methods aggregated over the RCM grid cells in the study area, as measured by the IQD, with a lower value indicating better performance. The marginal distribution over all days in 1987–2005 is compared with the corresponding distribution derived from the upscaled seNorge data product. Raw output and each bias-correction method are indicated by color. The middle line of a crossbar indicates the average IQD value across the grid cells, and the lower and upper bounds indicate a 90% score uncertainty obtained with 100 000 bootstrap samples.

in a bias-corrected trend that further exaggerates the model errors. As a result, the empirical distribution function over all time points in the test period for the Corr method has a much larger spread than that for the data product or that obtained under LocalSimple.

b. Bias correction and downscaling

1) MARGINAL PERFORMANCE

The marginal performance at the finescale is assessed in Fig. 6. From the example in Fig. 5, we see that the mean climate varies between the two time periods. The results here similarly show that adding climate change information from the RCMs substantially improves the WG realization centered on the mean climate in the training period (denoted X_{star}). For each RCM, the difference between $X_{starTrend}$ and $X_{starTrendVar}$ is small, the mean-only correction of $X_{starTrend}$ commonly showing minimally better performance. This may be explained partly by the fact that there is little difference between the variances of the two time periods and partly by the fact that the coarse-resolution variance of the RCMs does not perfectly relate to the finescale variance of the data product. We obtain consistently better results using the climate change information from RCM1 than RCM2. This is in line with the trend estimation results shown in Fig. 5, despite the bias-corrected RCM2 showing better overall marginal performance at the model scale (cf. Fig. 4).

The proposed two-step postprocessing approach shows consistently better marginal performance than EQM under an assessment of the full distribution and when focusing on the lower tail. When focusing on the

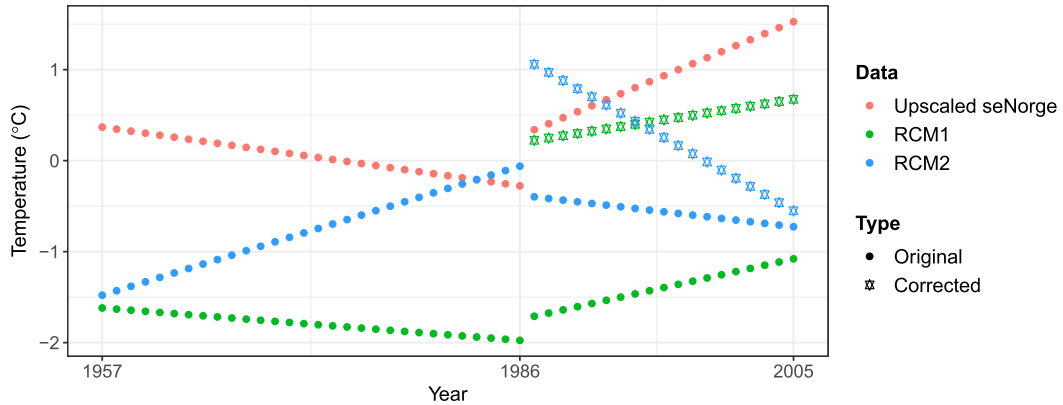


FIG. 5. Combined baseline [Eq. (4)] and linear trend [Eq. (6)] components of the estimated mean for one RCM grid cell in the study area, for the upscaled seNorge data product and the two RCMs over (left) the training period 1957–86 and (right) the test period 1987–2005, where also the corrected estimates of the two RCMs are indicated. The estimates are standardized such that the overall mean of the data product in the training period is equal to 0.

upper tail, or the central part of the distribution, the differences between the methods are smaller and the method ranking varies substantially across the catchments. Notably, the EQM results are much better for RCM2 than for RCM1. Furthermore, the EQM performance is less stable across the different catchments with some indications of worse performance in the inland catchments A and E through I.

As mentioned in section 2c(1), we have also investigated a slightly simpler two-step postprocessing procedure where the temporal residual series is assumed to follow an ARMA(p, q) model with Gaussian marginals. This simplification results in significantly reduced performance, adding approximately 0.008 to the average IQD value of the full distribution per catchment (results not shown). For RCM2, EQM performs better than this

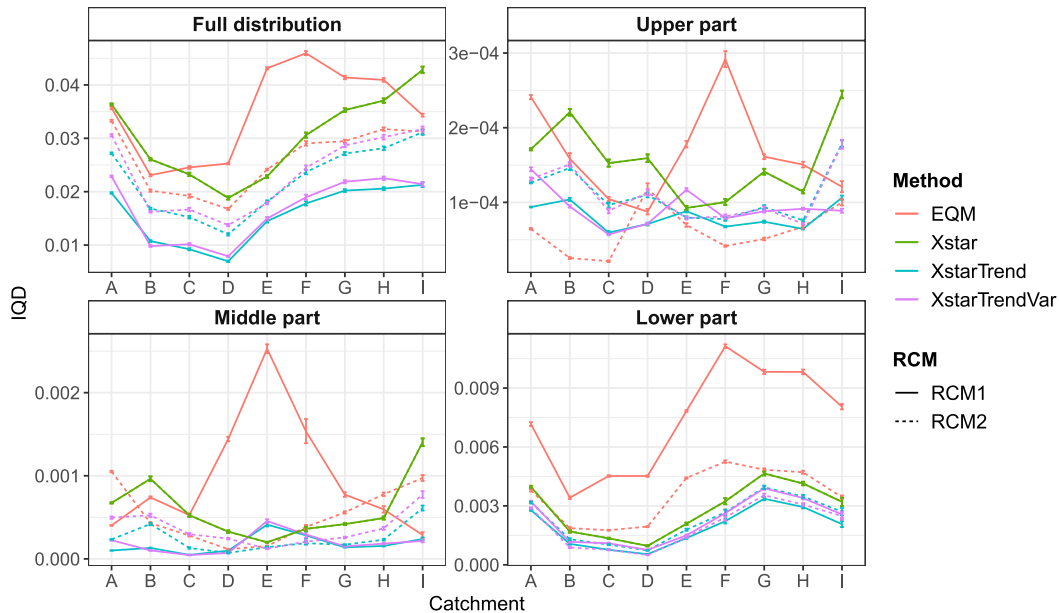


FIG. 6. IQD values for marginal comparison of the daily seNorge data product and postprocessed RCM model output for 1987–2005 aggregated over the grid cells in each catchment. A lower value indicates a better performance. The postprocessing method is indicated by the color, and the RCM is indicated by the line type. (top left) The full distributions are compared; also shown are comparisons focusing on (top right) the upper part [Eq. (16)], (bottom left) the middle part [Eq. (17)], and (bottom right) the lower part [Eq. (18)] of the distributions. The 90% score uncertainty bounds as obtained with 100 000 bootstrap samples are indicated with vertical error bars for each catchment and method.

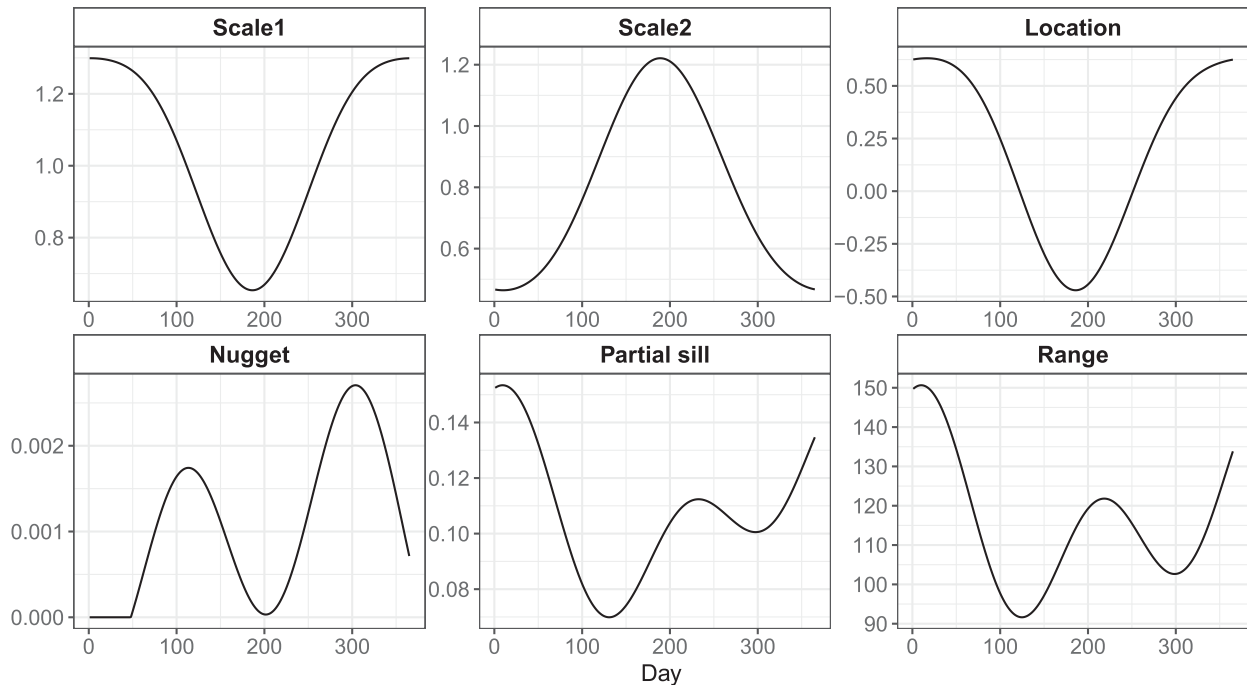


FIG. 7. Parameter estimates in catchment A (Gaulfoss) in the training period 1957–86 for the residual models in Eqs. (9) and (12), showing (top) the two scale parameters σ_{1t} and σ_{2t} ($^{\circ}\text{C}$) and the location parameter μ_t ($^{\circ}\text{C}$) of the split normal distribution in Eq. (9) and (bottom) the parameters of the exponential covariance function in Eq. (12): nugget θ_{0t} ($^{\circ}\text{C}^2$), partial sill θ_{1t} ($^{\circ}\text{C}^2$), and range θ_{2t} (km).

simplified two-step approach in seven out of the nine catchments.

Below, we further assess the spatial and temporal characteristics of the XstarTrend method applied to RCM1 and EQM applied to RCM2.

2) SPATIOTEMPORAL DEPENDENCE STRUCTURE

Parameter estimates for the split normal residual model in Eq. (9) and the spatial covariance function in Eq. (12) in the largest catchment A, Gaulfoss, are given in Fig. 7. The parameters of the split normal distribution follow a seasonal pattern that can be deduced from the data plot in Fig. 3, with the scale parameter for the lower tail σ_{1t} being higher in winter and lower in summer and the opposite holding for σ_{2t} , which is the scale parameter for the upper tail. While the location parameter estimates are, by construction, approximately mean zero over the entire year, these also follow a seasonal pattern with negative values in summer and positive values in winter. The spatial covariance function similarly exhibits a seasonal pattern. The spatial correlation has the highest range in winter followed by summer, while the range is smaller in spring and fall when, instead, the nugget parameter takes positive values.

To assess the temporal dependence, we estimate the autocorrelation function of the average daily temperature series in each catchment, see Fig. 8. The results are

very similar across the catchments: as to be expected for a temperature average over a larger area, the raw RCM output has a substantially higher autocorrelation than the finer-scale seNorge data product. Even if this is somewhat corrected in EQM, the results are not quite comparable to seNorge. The XstarTrend postprocessing inherits its temporal dependence structure mostly from the seNorge data product in training period, resulting in temporal dependence very similar to that of the data product in the test period.

For assessing the spatial dependence structure, Fig. 9 shows the empirical semivariograms for the seNorge data product, the XstarTrend method applied to RCM1 and EQM applied to RCM2 in the winter and summer months at the largest catchment A, Gaulfoss. While all three methods are comparable in the summer, the spatial dependence in winter is better modeled by XstarTrend than EQM. Due to its continental climate, the spatial dependence in the temperature at Gaulfoss is quite different for the two seasons. The y value attained when the semivariogram starts to level off, called *sill* in geostatistics, measures the total variance of the variable within the spatial domain. The temperatures are more variable in winter, leading to a larger sill (cf. Fig. 7). This feature is properly captured by XstarTrend, and largely overestimated by EQM. The distance at which the semivariogram first flattens out, called the *range*, is

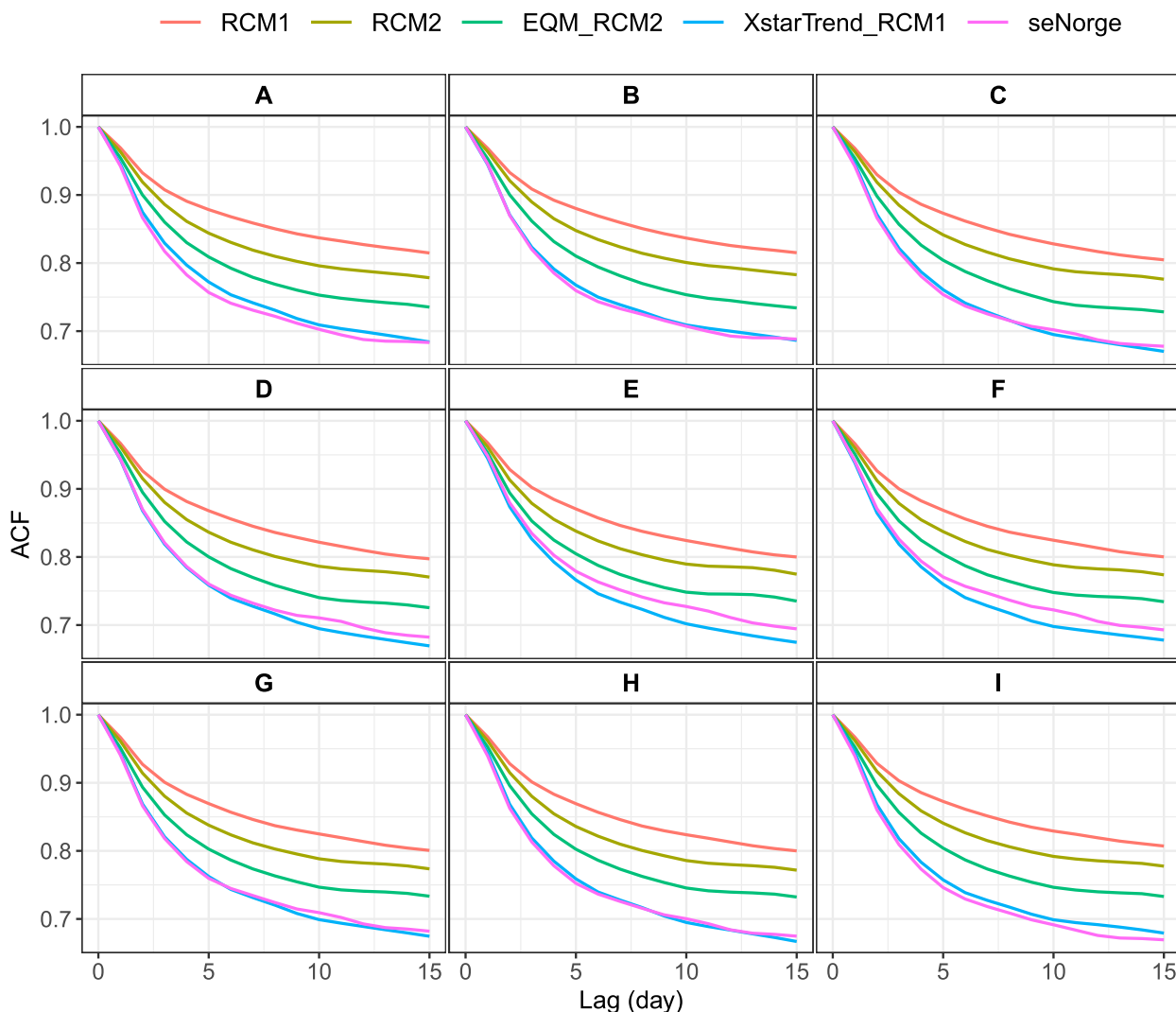


FIG. 8. The temporal dependence in the nine catchments, measured by an autocorrelation function of the average time series over the daily fields from 1987 to 2005, for the raw RCM output and two downscaling methods is compared with that from the seNorge data product.

typically around 30–35 km in summer and somewhat longer in winter, potentially because of dominance of continental arctic air masses in the region. In winter, the semivariogram values given by XstarTrend level off similarly as the seNorge data, whereas those by EQM show substantial differences for distances longer than 30 km. In general, the spatial dependence patterns vary somewhat between the catchments (results not shown). In particular, EQM and XstarTrend return essentially identical spatial patterns in the two maritime climate catchments C and D.

Figure 10 shows examples of cold and warm January days from the seNorge data product and the two post-processing methods. EQM relies on the nearest-neighbor method for interpolation between the two spatial scales and is applied independently for each

RCM grid cell. It can be seen that the resulting daily temperature fields have artificial boundaries corresponding to the RCM grid cells, whereas those from XstarTrend do not have such boundaries and show a spatial consistency closer to the seNorge temperature fields.

4. Discussion and conclusions

We propose a two-step statistical postprocessing procedure that bias corrects and downscales RCM simulations to a high-resolution grid. Our objective is to develop a full-field downscaling method for daily mean temperature that explicitly accounts for the fine-scale variability and dependence in both space and time. Employing two RCMs from the EURO-CORDEX ensemble and the high-resolution gridded observational

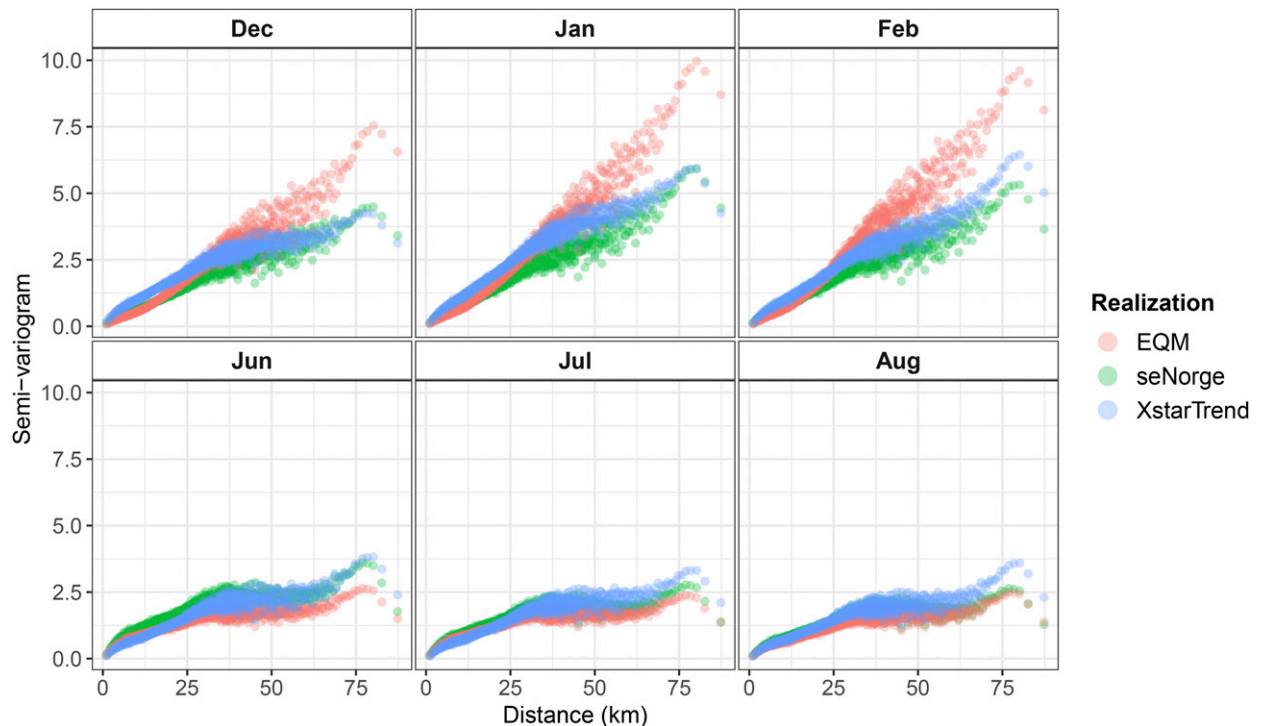


FIG. 9. Spatial dependence in catchment A (Gaulfoss) in 1987–2005 by month, as measured by an empirical semivariogram that is based on the daily temperature fields. The plots show empirical semivariograms derived from the seNorge data product and two downscaled results (EQM applied to RCM2 and XstarTrend applied to RCM1).

data product seNorge, we apply the procedure in the Trøndelag area of Norway, and find that the generated results are closer to the gridded reference data in terms of marginal, temporal and spatial properties than an empirical quantile mapping approach.

Our specific implementation separates statistical bias correction and stochastic downscaling. In a first step, to overcome the representativeness issue of the RCM simulations (Maraun and Widmann 2018) for local-scale climate, we follow Volosciuk et al. (2017) and perform bias correction only at the model scale. Then we follow, for example, Piani et al. (2010) and assume a Gaussian distribution for daily mean temperature. Here, using a model that parameters vary smoothly in space and time, we are able to account for the spatial and day-to-day variation of the two moments as well as a potential linear trend. Calibration is performed once for the full training dataset in each catchment. Other post-processing methods, however, often calibrate a model at single locations (e.g., Volosciuk et al. 2017) for individual months (e.g., the EQM applied in current study) or seasons (e.g., Vrac et al. 2012; Wong et al. 2014). Such separation of the data in space and time may overlook systematic variations with topography or seasons. Furthermore, they are typically unable to estimate a single

long-term linear trend for the whole domain, for example, separating it from the seasonal variations, and, as a consequence, modify the trend when correcting other properties (Maraun and Widmann 2018).

In a second step, we model the space–time variability at the finer scale using residuals generated on a high-resolution grid for limited areas (hydrological catchments in our case). Even for daily mean temperature that can be assumed Gaussian, simultaneous modeling of the space–time dependence is not straightforward. For computational feasibility and flexibility, we assume stationarity and space–time separability in the model. For the spatial dependence, we employ a similar approach as Wilks (2009) and specify a parametric covariance function of the exponential type with parameters smoothly changing to describe spatial structure variations throughout the year. Alternative, more advanced approaches here include the Matérn covariance model (e.g., Lindgren et al. 2011) or models based on nonparametric approaches such as principal component analysis (C. Heinrich et al. 2019, unpublished manuscript, <https://arxiv.org/abs/1907.09716>). For the temporal dependence, we found that the daily mean residuals have negative skewness in winter and positive skewness in summer (found also in, e.g., Huybers et al. 2014), which if not accounted for would lead to reduced

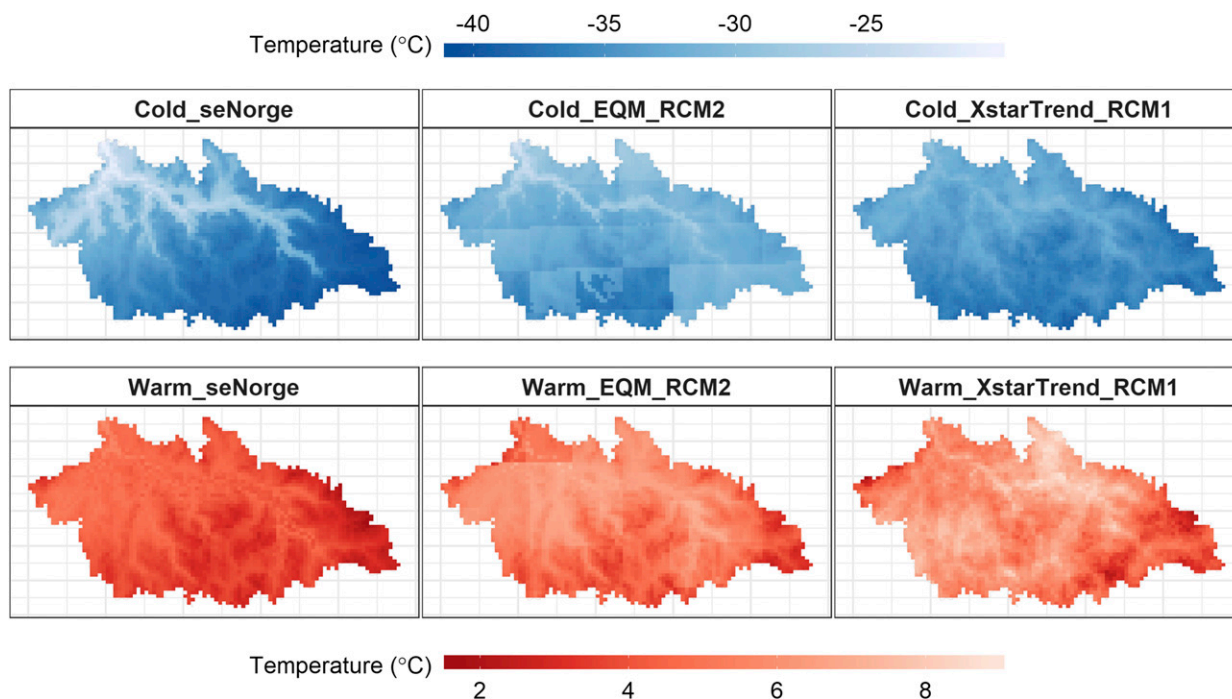


FIG. 10. Examples of the (top) coldest and (bottom) warmest temperature fields in catchment A (Gaulfoss) over all January days in the test period 1987–2005. The examples shown are from (left) the seNorge data product, (center) EQM applied to RCM2, and (right) XstarTrend applied to RCM1.

performance of the downscaling methods. Our solution is to combine a split normal distribution for the asymmetry with an ARMA model for the temporal correlation, a computationally feasible approach to non-Gaussian modeling of the temporal process.

The gap between the bias correction and the stochastic modeling is bridged by adding climate change signal derived from the model scale to the finescale WG realizations. Climate change signals in the mean and the variance can be selectively added to form the final results of the proposed procedure. Here, we compared three options: using just the stationary climate (Xstar), adjusting only the mean (XstarTrend), and adjusting both mean and variance (XstarTrendVar). To assess the agreement between the generated results and the gridded data in the test period, we employ the IQD for evaluation of the marginal aspect, the autocorrelation function for temporal dependence, and the empirical semivariogram for spatial dependence. We find that in all the catchments in our study area and under both RCMs, XstarTrend and XstarTrendVar perform better than EQM in terms of marginal distribution and temporal dependence, while properly representing spatial dependence. In addition, we found that the skill of an RCM at the coarser scale may not necessarily carry over to the finer scale and agree with [Maraun and Widmann \(2018\)](#) that it is important to assess the skill of the

climate model output in terms of the information to be used.

The parameter estimation in our approach is performed using methods such as maximum likelihood estimation where parameter uncertainty is not accounted for. While Bayesian inference methods, that return assessments of the parameter uncertainty, have been used for similar models (e.g., [Hewitt et al. 2018](#)), they tend to be computationally significantly more complex and may have issues with scaling to high dimensions. In practice, available information regarding parameter uncertainty would allow us to slightly vary the parameters used in the WG simulations at finescale, in the marginal space-time distributions and for the trend estimates. This would somewhat increase the variability between individual simulations, in particular regarding the trend as the space-time variation in the other components already induces variability. Such an approach would also be appropriate for combining information from multiple sources, a discussion we consider outside of the scope of the current paper.

The example in [Fig. 5](#) illustrates the importance of comparing the trend estimate of the RCM and the calibration data in the training period before applying a trend correction to the RCM output. The example furthermore shows that the training and the test period used in our study are not sufficiently long for a robust

estimation of local trends. The large differences between the estimated trends in the training and the test period indicate that what is here considered trend is likely to be partly due to natural variability, so that large differences between the various datasets are to be expected. In practice, when applying our method to future climate projections, one should thus use the entire available historic period of 49 years for the calibration.

From the example in Fig. 10, we see that the EQM reference method appears to have some aliasing effects at the edges of the RCM grid cells. These effects might possible be improved by using a spatially smoothing interpolation such as bilinear interpolation rather than the nearest-neighbor approach applied here. On the other hand, under bilinear interpolation, the area averages would not match the raw RCM outputs in the corresponding RCM grid cell. Optimally, one would furthermore apply a spatially consistent transfer function. However, the multivariate EQM approaches proposed thus far (e.g., Cannon 2016) are not able to deal with such a high-dimensional setting.

Acknowledgments. This work was supported by the Research Council of Norway through project 255517 “Post-processing Climate Projection Output for Key Users in Norway.” We thank three anonymous reviewers for their comments, which helped to improve the paper.

APPENDIX

Details of Bias-Correction Methods

For bias correcting the RCM simulations in the test period, we correct the errors that are related to spatial and temporal features of the two parameters of the Gaussian model specified in Eqs. (1)–(6). In the method Corr, we set

$$\hat{z}_r^{\text{te}} = (y_r^{\text{te}} - \hat{\mu}_r^{\text{te}}) / \hat{\sigma}_r^{\text{te}} \quad \text{and} \quad (\text{A1})$$

$$\hat{y}_r^{\text{te,corr}} = \hat{\mu}_r^{\text{te,corr}} + \hat{z}_r^{\text{te}} \hat{\sigma}_r^{\text{te,corr}}, \quad (\text{A2})$$

where “te” denotes the test period. That is, we replace the MLEs of $\hat{\mu}_r^{\text{te}}$ and $\hat{\sigma}_r^{\text{te}}$ with their corrected counterparts $\hat{\mu}_r^{\text{te,corr}}$ and $\hat{\sigma}_r^{\text{te,corr}}$.

To derive $\hat{\mu}_r^{\text{te,corr}}$, we first define the long-term average over a given period for a given RCM cell as

$$\begin{aligned} \bar{\mu}_r &:= \frac{1}{T} \sum_{t=1}^T [\hat{f}_1^\mu(\mathbf{c}_r) + \hat{f}_2^\mu(t) + \hat{f}_3^\mu(t)] \\ &= \hat{f}_1^\mu(\mathbf{c}_r) + \hat{\alpha}_3 \bar{n}, \end{aligned} \quad (\text{A3})$$

where \hat{f}_1^μ , \hat{f}_2^μ , and \hat{f}_3^μ are the three components of the first moment specified in Eqs. (4)–(6) with coefficients equal

to their MLEs. That is, $\bar{\mu}_r$ is the sum of the local baseline and a regional trend increment. In Eq. (6), we index the calendar years by an arithmetic sequence $y(t)$ with initial value 0 and increments of 0.1, so that $y(t) \in \{0, 0.1, \dots, 2.9\}$ for the 30-yr training period and $y(t) \in \{0, 0.1, \dots, 1.8\}$ for the 19-yr test period. It follows that $\bar{n} = 1.45$ for the training period and $\bar{n} = 0.9$ for the test period. With these settings, we formulate the corrected $\bar{\mu}_r$ in the test period as the value based on the upscaled seNorge data $\bar{\mu}_r^o$ combined with the RCM-simulated change from the training period to the test period $\bar{\mu}_r^{\text{te}} - \bar{\mu}_r^{\text{tr}}$,

$$\begin{aligned} \bar{\mu}_r^{\text{te,corr}} &= \bar{\mu}_r^o + \bar{\mu}_r^{\text{te}} - \bar{\mu}_r^{\text{tr}} \\ &= \hat{f}_1^{\mu,o}(\mathbf{c}_r) + \hat{f}_1^{\mu,\text{te}}(\mathbf{c}_r) - \hat{f}_1^{\mu,\text{tr}}(\mathbf{c}_r) \\ &\quad + 1.45(\hat{\alpha}_3^o - \hat{\alpha}_3^{\text{tr}}) + 0.9\hat{\alpha}_3^{\text{tr}}, \end{aligned} \quad (\text{A4})$$

where the second equality follows from Eq. (A3).

The RCM simulated trend in the test period is corrected as

$$\hat{f}_3^{\mu,\text{te,corr}}(t) = \hat{\alpha}_3^{\text{te,corr}} y(t), \quad (\text{A5})$$

where $\hat{\alpha}_3^{\text{te,corr}} = \hat{\alpha}_3^o + \hat{\alpha}_3^{\text{te}} - \hat{\alpha}_3^{\text{tr}}$. The corrected baseline function of the first moment is obtained using the result of Eq. (A3), that is, by subtracting the increment due to the trend from the corrected long-term average in Eq. (A4),

$$\begin{aligned} \hat{f}_1^{\mu,\text{te,corr}}(\mathbf{c}_r) &= \bar{\mu}_r^{\text{te,corr}} - 0.9\hat{\alpha}_3^{\text{te,corr}} \\ &= \hat{f}_1^{\mu,o}(\mathbf{c}_r) + \hat{f}_1^{\mu,\text{te}}(\mathbf{c}_r) - \hat{f}_1^{\mu,\text{tr}}(\mathbf{c}_r) \\ &\quad + 0.55(\hat{\alpha}_3^o - \hat{\alpha}_3^{\text{tr}}). \end{aligned} \quad (\text{A6})$$

The corrected seasonality of the first moment is

$$\hat{f}_2^{\mu,\text{te,corr}}(t) = \hat{f}_2^{\mu,o}(t) + \hat{f}_2^{\mu,\text{te}}(t) - \hat{f}_2^{\mu,\text{tr}}(t). \quad (\text{A7})$$

By combining the results above, we have the corrected first moment for a given RCM grid cell r at time t ,

$$\hat{\mu}_r^{\text{te,corr}} = \hat{f}_1^{\mu,\text{te,corr}}(\mathbf{c}_r) + \hat{f}_2^{\mu,\text{te,corr}}(t) + \hat{f}_3^{\mu,\text{te,corr}}(t). \quad (\text{A8})$$

For the standard deviation $\hat{\sigma}_r^{\text{te,corr}}$, we simply set

$$\hat{\sigma}_r^{\text{te,corr}} = [(\hat{\sigma}_r^o)^2 + (\hat{\sigma}_r^{\text{te}})^2 - (\hat{\sigma}_r^{\text{tr}})^2]^{1/2}. \quad (\text{A9})$$

For comparison, we perform two simple, nonparametric methods of correcting the mean biases of the RCM simulations in the test period. Both methods explicitly preserve the change in the long-term average simulated by RCM. In the method Simple, we first calculate three domainwise long-term averages: a^o from the upscaled seNorge data in the training period, a^{tr} from the RCM raw outputs in the training period, and a^{te} from the RCM

raw outputs in the test period. We then remove a^{te} from the raw output y_r^{te} and add back the corrected average $a^o + a^{\text{te}} + a^{\text{tr}}$ for the test period. Formally,

$$\begin{aligned} y_r^{\text{te,Simple}} &= y_r^{\text{te}} - a^{\text{te}} + (a^o + a^{\text{te}} - a^{\text{tr}}) \\ &= y_r^{\text{te}} + a^o - a^{\text{tr}}. \end{aligned} \quad (\text{A10})$$

In the second method, the same procedure is applied to correct the mean biases for individual RCM grid cells, hence the name LocalSimple,

$$y_r^{\text{te,LocalSimple}} = y_r^{\text{te}} + a_r^o - a_r^{\text{tr}}. \quad (\text{A11})$$

REFERENCES

- Abel, G. J., 2015: Fanplot: An R package for visualising sequential distributions. *R J.*, **7**, 15–23, <https://doi.org/10.32614/RJ-2015-002>.
- Beldring, S., K. Engeland, L. A. Roald, N. R. Sælthun, and A. Voksø, 2003: Estimation of parameters in a distributed precipitation-runoff model for Norway. *Hydrol. Earth Syst. Sci.*, **7**, 304–316, <https://doi.org/10.5194/hess-7-304-2003>.
- , T. Engen-Skaugen, E. J. Førland, and L. A. Roald, 2008: Climate change impacts on hydrological processes in Norway based on two methods for transferring regional climate model results to meteorological station sites. *Tellus*, **60A**, 439–450, <https://doi.org/10.1111/j.1600-0870.2008.00306.x>.
- Cannon, A. J., 2016: Multivariate bias correction of climate model output: Matching marginal distributions and intervariable dependence structure. *J. Climate*, **29**, 7045–7064, <https://doi.org/10.1175/JCLI-D-15-0679.1>.
- Clark, M., S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, 2004: The Schaake shuffle: A method for reconstructing space time variability in forecasted precipitation and temperature fields. *J. Hydrometeorol.*, **5**, 243–262, [https://doi.org/10.1175/1525-7541\(2004\)005<0243:TSSAMF>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2).
- Cressie, N., and C. K. Wikle, 2015: *Statistics for Spatio-Temporal Data*. John Wiley and Sons, 624 pp.
- Giorgetta, M. A., and Coauthors, 2013: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model. Earth Syst.*, **5**, 572–597, <https://doi.org/10.1002/jame.20038>.
- Gneiting, T., and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. *J. Amer. Stat. Assoc.*, **102**, 359–378, <https://doi.org/10.1198/016214506000001437>.
- Gräler, B., E. Pebesma, and G. Heuvelink, 2016: Spatio-temporal interpolation using gstat. *R J.*, **8**, 204–218, <https://doi.org/10.32614/RJ-2016-014>.
- Gudmundsson, L., 2016: Qmap: Statistical transformations for post-processing climate model output, version 1.0-4. R package, <https://CRAN.R-project.org/package=qmap>.
- , J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen, 2012: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations—A comparison of methods. *Hydrol. Earth Syst. Sci.*, **16**, 3383–3390, <https://doi.org/10.5194/hess-16-3383-2012>.
- Hall, A., 2014: Projecting regional change. *Science*, **346**, 1461–1462, <https://doi.org/10.1126/science.aaa0629>.
- Hanssen-Bauer, I., H. O. Hygen, H. Heiberg, E. Førland, and N. Berit, 2017: User needs for post processed climate data. Norwegian Centre for Climate Services NCCS Rep. 2, 22 pp., https://cms.met.no/site/2/klimaservicesenteret/rapporter-og-publikasjoner/_attachment/12099?_ts=15dd01f2d73.
- Hewitt, J., J. A. Hoeting, J. M. Done, and E. Towler, 2018: Remote effects spatial process models for modeling teleconnections. *Environmetrics*, **29**, e2523, <https://doi.org/10.1002/env.2523>.
- Huybers, P., K. A. McKinnon, A. Rhines, and M. Tingley, 2014: U.S. daily temperatures: The meaning of extremes in the context of nonnormality. *J. Climate*, **27**, 7368–7384, <https://doi.org/10.1175/JCLI-D-14-00216.1>.
- Hyndman, R., and Coauthors, 2018: Forecast: Forecasting functions for time series and linear models, version 8.4. R package, <http://pkg.robjhyndman.com/forecast>.
- IPCC, 2014: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects*. V. R. Barros et al., Eds., Cambridge University Press, 688 pp., http://www.ipcc.ch/pdf/assessment-report/ar5/wg2/WGIIAR5-PartB_FINAL.pdf.
- Jacob, D., and Coauthors, 2014: EURO-CORDEX: New high-resolution climate change projections for European impact research. *Reg. Environ. Change*, **14**, 563–578, <https://doi.org/10.1007/s10113-013-0499-2>.
- Jordan, A., F. Krüger, and S. Lerch, 2018: Evaluating probabilistic forecasts with scoringRules. 37 pp., <https://cran.r-project.org/web/packages/scoringRules/vignettes/article.pdf>.
- Kilsby, C., and Coauthors, 2007: A daily weather generator for use in climate change studies. *Environ. Modell. Software*, **22**, 1705–1719, <https://doi.org/10.1016/j.envsoft.2007.02.005>.
- Kotlarski, S., and Coauthors, 2014: Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geosci. Model Dev.*, **7**, 1297–1333, <https://doi.org/10.5194/gmd-7-1297-2014>.
- Lindgren, F., H. Rue, and J. Lindström, 2011: An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. Roy. Stat. Soc.*, **73B**, 423–498, <https://doi.org/10.1111/j.1467-9868.2011.00777.x>.
- Lussana, C., T. Saloranta, T. Skaugen, J. Magnusson, O. E. Tveito, and J. Andersen, 2018a: seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day. *Earth Syst. Sci. Data*, **10**, 235–249, <https://doi.org/10.5194/essd-10-235-2018>.
- , O. Tveito, and F. Uboldi, 2018b: Three-dimensional spatial interpolation of 2 m temperature over Norway. *Quart. J. Roy. Meteor. Soc.*, **144**, 344–364, <https://doi.org/10.1002/qj.3208>.
- Maraun, D., and Coauthors, 2010: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.*, **48**, RG3003, <https://doi.org/10.1029/2009RG000314>.
- , and Coauthors, 2017: Towards process-informed bias correction of climate change simulations. *Nat. Climate Change*, **7**, 764–773, <https://doi.org/10.1038/nclimate3418>.
- , and M. Widmann, 2018: *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press, 347 pp.
- Mohr, M., 2009: Comparison of versions 1.1 and 1.0 of gridded temperature and precipitation data for Norway. Norwegian Meteorological Institute Note 19/2009, 46 pp.
- Nelsen, R. B., 2007: *An Introduction to Copulas*. Springer Science & Business Media, 272 pp.
- Pebesma, E. J., 2004: Multivariable geostatistics in S: The gstat package. *Comput. Geosci.*, **30**, 683–691, <https://doi.org/10.1016/j.cageo.2004.03.012>.
- , 2012: spacetime: Spatio-temporal data in R. *J. Stat. Software*, **51**, 1–30, <https://doi.org/10.18637/jss.v051.i07>.
- Piani, C., and J. O. Haerter, 2012: Two dimensional bias correction of temperature and precipitation copulas in climate

- models. *Geophys. Res. Lett.*, **39**, L20401, <https://doi.org/10.1029/2012GL053839>.
- , G. Weedon, M. Best, S. Gomes, P. Viterbo, S. Hagemann, and J. Haerter, 2010: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *J. Hydrol.*, **395**, 199–215, <https://doi.org/10.1016/j.jhydrol.2010.10.024>.
- Posthuma Partners, 2018: Lmvar: Linear regression with non-constant variances, version 1.5.0. R package, <https://CRAN.R-project.org/package=lmvar>.
- R Core Team, 2018: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, <https://www.R-project.org/>.
- Räisänen, J., and O. Räty, 2013: Projections of daily mean temperature variability in the future: Cross-validation tests with ENSEMBLES regional climate simulations. *Climate Dyn.*, **41**, 1553–1568, <https://doi.org/10.1007/s00382-012-1515-9>.
- Rockel, B., A. Will, and A. Hense, 2008: The regional climate model COSMO-CLM (CCLM). *Meteor. Z.*, **17**, 347–348, <https://doi.org/10.1127/0941-2948/2008/0309>.
- Rummukainen, M., 2010: State-of-the-art with regional climate models. *Wiley Interdiscip. Rev.: Climate Change*, **1**, 82–96, <https://doi.org/10.1002/wcc.8>.
- Semenov, M. A., and E. M. Barrow, 1997: Use of a stochastic weather generator in the development of climate change scenarios. *Climatic Change*, **35**, 397–414, <https://doi.org/10.1023/A:1005342632279>.
- Thorarinsdottir, T. L., T. Gneiting, and N. Gissibl, 2013: Using proper divergence functions to evaluate climate models. *SIAM-ASA J. Uncertainty Quantif.*, **1**, 522–534, <https://doi.org/10.1137/130907550>.
- Venables, W. N., and B. D. Ripley, 2002: *Modern Applied Statistics with S*. 4th ed. Springer, 495 pp.
- Voldoire, A., and Coauthors, 2013: The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dyn.*, **40**, 2091–2121, <https://doi.org/10.1007/s00382-011-1259-y>.
- Volosciuk, C., D. Maraun, M. Vrac, and M. Widmann, 2017: A combined statistical bias correction and stochastic downscaling method for precipitation. *Hydrol. Earth Syst. Sci.*, **21**, 1693–1719, <https://doi.org/10.5194/hess-21-1693-2017>.
- Vrac, M., 2018: Multivariate bias adjustment of high-dimensional climate simulations: The rank resampling for distributions and dependences (R^2D^2) bias correction. *Hydrol. Earth Syst. Sci.*, **22**, 3175–3196, <https://doi.org/10.5194/hess-22-3175-2018>.
- , and P. Friederichs, 2015: Multivariate—intervariable, spatial, and temporal—bias correction. *J. Climate*, **28**, 218–237, <https://doi.org/10.1175/JCLI-D-14-00059.1>.
- , P. Drobinski, A. Merlo, M. Herrmann, C. Lavaysse, L. Li, and S. Somot, 2012: Dynamical and statistical downscaling of the French Mediterranean climate: Uncertainty assessment. *Nat. Hazard Earth Syst.*, **12**, 2769–2784, <https://doi.org/10.5194/nhess-12-2769-2012>.
- Wallis, K. F., 2014: The two-piece normal, binormal, or double Gaussian distribution: Its origin and rediscoveries. *Stat. Sci.*, **29**, 106–112, <https://doi.org/10.1214/13-STS417>.
- Wilks, D. S., 1999: Multisite downscaling of daily precipitation with a stochastic weather generator. *Climate Res.*, **11**, 125–136, <https://doi.org/10.3354/cr011125>.
- , 2009: A gridded multisite weather generator and synchronization to observed weather data. *Water Resour. Res.*, **45**, W10419, <https://doi.org/10.1029/2009WR007902>.
- , 2010: Use of stochastic weather generators for precipitation downscaling. *Wiley Interdiscip. Rev.: Climate Change*, **1**, 898–907, <https://doi.org/10.1002/wcc.85>.
- , 2012: Stochastic weather generators for climate-change downscaling, part II: Multivariable and spatially coherent multisite downscaling. *Wiley Interdiscip. Rev.: Climate Change*, **3**, 267–278, <https://doi.org/10.1002/wcc.167>.
- Wong, G., D. Maraun, M. Vrac, M. Widmann, J. M. Eden, and T. Kent, 2014: Stochastic model output statistics for bias correcting and downscaling precipitation including extremes. *J. Climate*, **27**, 6940–6959, <https://doi.org/10.1175/JCLI-D-13-00604.1>.
- Wood, A. W., L. R. Leung, V. Sridhar, and D. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, **62**, 189–216, <https://doi.org/10.1023/B:CLIM.0000013685.99609.9e>.
- Xu, C.-Y., 1999: Climate change and hydrologic models: A review of existing gaps and recent research developments. *Water Resour. Manage.*, **13**, 369–382, <https://doi.org/10.1023/A:1008190900459>.