High-resolution downscaling and bias-correction of temperature and precipitation: advances in statistical methods

A Dissertation Presented

by

Maike Holthuijzen

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
Specializing in Complex Systems and Data Science

May, 2023

Defense Date: April 11, 2022
Dissertation Examination Committee:

Brian Beckage, Ph.D., Advisor
Chris Danforth, Ph.D., Chairperson
Patrick J. Clemins, Ph.D.
Donna Rizzo, Ph.D.
Dave Higdon, Ph.D.
Cynthia J. Forehand, Ph.D., Dean of the Graduate College

Abstract

High-resolution, bias-corrected climate data is necessary for climate impact studies and modeling efforts at local scales. General circulation models (GCMs) provide important information about historical and future larger-scale climate trends, but their spatial resolution is too coarse to investigate localized effects of climate processes. Additionally, raw GCM output is characterized by some degree of bias. Two post-processing procedures known as downscaling and bias-correction are typically applied to raw climate model output prior to its use in further modeling applications. Downscaling is the process in which data at a coarse spatial scale is transformed to a fine spatial scale. Bias-correction refers to a collection of methods in which climate model output is adjusted such that its statistical properties (e.g. mean, variance, and potentially higher moments) resemble those of observations in a common climatological period. Bias-correction is a challenge, due to relatively short calibration and long future time periods and potential spatial misalignment issues between gridded climate model output and observed data. Issues that warrant further research are 1) spatially-coherent bias-correction, 2), processing of extremes, 3) temporally-coherent bias-correction, and 4) balancing the bias-correction of future model output with the preservation of the climate change signal. Performing spatially-coherent bias-correction is particularly difficult, as model and observed data must be present in the same location where bias-correction is applied. Depending on the type of observed data used, this may not be the case. Extremes are challenging to represent accurately during bias-correction, because extreme values in both observed and model data are highly variable, limited, and there is greater uncertainty regarding their correction. Finally, very few bias-correction methods explicitly correct temporal dependence structures of model output. However, it is important that the temporal dependence of model data resembles that of observed data, as climate variability is closely linked to temporal dependence. In this body of work, I developed methodological workflows to generate high-resolution climate data products in which 1) bias-correction is carried out in a spatially-coherent manner, and 2) precipitation extremes are accurately represented. I also created a new, two-step bias-correction approach in which the temporal dependence and distributional properties of model output are corrected. This method allows for sensible bias-correction in both historical and future time periods and minimizes distortion to the future climate change signal.

Citation Page

Material from this dissertation has been published in the following form:

Holthuijzen, Maike F., Beckage, B., Clemins P. J., Higdon, D., Winter, J. M.. "Constructing High-Resolution, Bias-Corrected Climate Products: A Comparison of Methods." *Journal of Applied Meteorology and Climatology* 60.4 (2021): 455-475.

Holthuijzen, Maike F., Beckage, B., Clemins P. J., Higdon, D., Winter, J. M.. "Robust bias-correction of precipitation extremes using a novel hybrid empirical quantile-mapping method." *Journal of Theoretical and Applied Climatology* (accepted).

Holthuijzen, Maike F., Clemins P. J., Higdon, D., Brian, B.., "Calibrating output from numerical climate models featuring a process convolution approach for correcting temporal dependence." *Environmetrics* (in preparation).

Acknowledgements

I would like to thank the members of my committee for their advice, suggestions, constructive criticism, and moral support. Dr. Brian Beckage provided inspiration for my research and helped me refine my research ideas. I am grateful to Dr. Patrick Clemins for his thoughtful counsel, IT support, and his willingness to answer all my computer-related questions. Dr. Donna Rizzo was always there to help me when things did not go as planned; I appreciate her encouragement and support. I wish to thank Dr. Chris Danforth and Dr. Peter Dodds for all their work in establishing the UVM Complex Systems and Data Science PhD program. Dr. Dave Higdon (Virginia Tech) challenged my statistical and analytical competence and helped me develop skills for tackling challenging modeling problems. He also introduced me to the rich and varied world of Gaussian process modeling, for which I am most grateful. I also extend my gratitude to Dr. Jonathan M. Winter (Dartmouth college) for his thoughtful comments, criticism, and assistance in preparation of manuscripts. Dr. Winter, Dr. Huanping Huang and Charlotte Cockburn provided the data used in this research. Susan Durham (Utah State University) provided me with suggestions and recommendations for various statistical challenges I encountered while carrying out this research. I had the pleasure of working with three bright, hardworking EPSCoR interns (Connor Zwonick, Ann-Marie Matheny, and Nate Fogg), and I am thankful for their contributions to parts of this research. I would like to thank Dr. Jill Lundell for her helpful comments on an earlier version of Chapter 4. I am also grateful for Dr. Cindy Bruyere's expertise on climate models. Finally, I am thankful for Dr. Maggie Eppstein for being another strong, inspiring woman in my life. Dr. Eppstein was always happy to offer advice and guidance when I needed it.

I am deeply grateful to the people who indirectly helped me through their friendship and generosity. I am thankful to Dr. Jill Preston for believing in me when I had my doubts, and for our many Sunday runs together. I appreciate all of my cycling friends, especially Roxy Bombardier and Kate Barash-Engel, who introduced me to the world of gravel cycling and kept me company on many biking adventures. I am thankful for the supportive community at my crossfit gym and I especially appreciate the owners Sam Werbel and Molly Purvis for their enthusiasm and positive outlook. My coach, Suzan Ballmer, helped me become a better triathlete, person, and researcher through her encouragement, support, and weekly training schedules. Last but not least, my sister Wieteke Holthuijzen and my parents, Toni Holthuijzen and Lenora Oosterhuis, were always there to offer their love and steadfast support when I was frustrated or disheartened; they were my greatest fans, even though my research may remain somewhat of a mystery to them. Funding for the first 4 years of my PhD program was supported by the National Science Foundation under VT EPSCoR Grant No. NSF OIA 1556770. I am grateful to the Department of Computer Science for funding the final year of my tenure at the University of Vermont.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context and background

The research presented in the following chapters was initiated and financially supported by the Vermont Established Program to Stimulate Competitive Research (VT EPSCoR), a National Science Foundation program that brings together researchers, policymakers, stakeholders, members of the private sector and other collaborators to solve problems related to the sustainability of natural, social, and economic resources of the Lake Champlain Basin. Lake Champlain Basin is a topographically diverse region of approximately 13,500km^2 that encompasses Lake Champlain and its associated watersheds.

The effects of climate change on the ecological, social, and economic systems of Lake Champlain Basin are of great concern. This region, as well as the Northeastern United States in general, is experiencing significant warming and increased precipitation. These trends, along with greater climatic variability, extreme precipitation events as well as droughts, are expected to continue in the future [1]. One objective of

VT EPSCoR is to provide pertinent scientific research to assist policymakers, stakeholders and the broader public in developing strategies to mitigate and/or adapt to climate change. In this context, the availability of high quality climate products at local scales is critical.

The research in this dissertation contributes novel statistical methods to construct accurate climate data products at fine spatiotemporal resolutions. The statistical methods and workflows I developed improve the accuracy and reduce uncertainty in downscaled, bias-corrected climate data. I applied these methods to precipitation and temperature simulations from a regional climate model and constructed high-resolution, daily climate products over historical and future time periods. The proposed methodologies can be applied to other downscaling and bias-correction applications.

1.2 Literature review

As the impacts of climate change become more severe and widespread, scientists, practitioners, and policy makers require accurate, fine-scale climate data for local climate impact studies. Ecological, hydrological, agricultural, and economic studies, which often rely on models driven by climate data, are typically conducted at fine ($\leq$ 1km) spatial resolutions [2]–[4]. Though general circulation models (GCMs) are indispensable tools for modeling past and future climate, their coarse spatial resolution prohibits them from capturing complex, fine-scale topography, orography, and climate processes [5], [6]. In addition, GCMs simplify climate through parameterization schemes, resulting in the inadequate representation of some climate processes

[7]. Consequently, output from GCMs is characterized by a non-trivial degree of bias [7]–[9]. Typically, post-processing steps such as downscaling and bias-correction are applied to climate model output prior to its use in modeling applications.

1.2.1 Overview of downscaling methods

In the downscaling process, climate model output is transformed from a coarse to finer resolution. Downscaling is especially important in regions characterized by topographically varied terrain [10], [11], as elevation has a large influence on climate variables such as temperature and precipitation [12]. The two main types of downscaling are dynamical and statistical downscaling. In dynamical downscaling, a regional climate model (RCM) is forced by GCM or reanalysis data. An RCM simulates climate processes at a finer resolution than forcing data by incorporating fine-scale landscape and atmospheric processes [5], [6], [13], [14]. RCMs are computationally intensive, although they typically require less processing power than GCMs [15], [16]. Localized climate processes, including extremes [17], are generally better reproduced in RCMs compared to GCMs [18], although seasonal biases in extremes may remain [19]. Despite the sophistication and the ability of RCMs to model climate over regional scales, they 1) are unable to fully capture the fine-scale effects of complex topography [20], 2) simplify climatic processes through model parameterizations 3) cannot account for all the natural variability of climatic processes, and 4) under- or over-estimate complex atmospheric feedbacks [7].

Statistical downscaling involves establishing statistical relationships between coarse-scale and fine-scale climate variables [14]. Statistical downscaling is computationally efficient and can be applied to a variety of climate variables [21], [22]. In contrast

to dynamical downscaling, a substantial amount of observational data is necessary to derive statistical relationships necessary for statistical downscaling [14], which may not always be available. Frequently-used approaches for statistical downscaling are regression-based methods [5] and principal components analysis [23], [24], weather classification schemes, and weather generators [14]. Recently, machine learning methods such as artificial neural networks [25], deep learning [26], and random forests [27] have been used for downscaling both temperature and precipitation variables.

Linear regression (LR) models establish linear relationships between climate variables and predictors. LR is simple, easy to implement, efficient, and therefore widely used [5]. The main shortcomings of LR is the requirement of a Gaussian distribution for the response variable and a linear relationship between response and predictor variables [28]. LR is not suitable for downscaling precipitation, as precipitation data are generally highly right-skewed, and relationships of precipitation and predictors such as elevation are generally non-linear [12], [28]. LR is best suited for downscaling temperature climate variables [29].

In weather classification, atmospheric "states" are identified in long-term time series of observational data that correspond to similar states of future projections from a GCM [30]. The chosen historical states are then simulated under future conditions. An advantage of weather classification is that it does not require climate variables to be normally distributed, so it can be applied to precipitation data [28]. However, weather classification requires observational weather data of at least 30 years and is more computationally demanding than regression methods [28].

Weather generators are statistical models that simulate local, daily climate variables based on monthly climate time series from nearby weather stations [28]. These

models first replicate precipitation and then simulate other atmospheric variables such temperature, humidity, and wind conditionally on precipitation. The underlying mechanism in weather generators are Markov processes, where the value of the climate variable of interest on day t is only dependent on the value on day $t-1$ [30]. Weather generators are especially sensitive to missing values and errors in observational data [28].

Simple statistical interpolation methods, such as topographical downscaling can also be an effective downscaling approach. Topographic downscaling combines inverse distance weighting (IDW) and lapse rates to adjust for fine-scale elevation [31]. In IDW interpolation, observed values close to prediction locations are assumed to be more influential in the prediction compared to observed values far from prediction locations. A limitation of topographic downscaling is the requirement of accurate, fine-scale elevation data.

More recently, machine learning methods including deep neural networks [26] [32], random forests [33], and support and relevance vector machines [34] have been used for downscaling temperature and precipitation variables. Relevance vector machines [35], artificial neural networks [36], [37], random forests [38] and deep learning neural networks have been shown to be effective in downscaling precipitation [32]. [37] and [35] found that support and vector machines and artificial neural networks resulted in better representation of the mean, rather than upper or lower quantiles of precipitation variables. [32] reported that adding more layers to a deep neural network resulted in near perfect representation of temperature and precipitation variables at a fine resolution. However, [34] noted that machine learning methods tended to underestimate the standard deviation and several other studies reported that machine

learning methods did not perform as well as regression-based methods. [34] and [37] found that linear regression-based downscaling methods generally outperformed the machine learning methods of deep learning neural networks and support and relevance vector machines. The major drawback to deep learning methods is their computational cost. As computing power continues to increase, machine learning methods may become more attractive for downscaling.

The accurate representation of extreme events in downscaled data continues to be a challenge [39]. Furthermore, there is a need for robust, statistical downscaling methods that can quantify the uncertainty of downscaled climate data [6], [40], [41].

1.2.2 Review of bias-correction methods

While downscaling can be achieved by relatively simple interpolation methods, bias-correction is far more difficult, due to several factors. First, climate model output is typically characterized by a short calibration period and long future period and observed data may not be available over the entire study area. In addition, uncertainty of the bias-correction method is difficult to quantify. Finally, there is often a spatial misalignment issue between model output, which is represented as averages over grid cells, and observations, which may be point-based.

The main goal of bias-correction is to modify statistical properties of climate model output such that they resemble those of observed data. Bias-correction adjusts the mean, variance, and potentially higher moments of climatological variables [8], [9]. Generally, the most widely-used bias-correction methods can be classified into three categories: 1) linear scaling [42], [43]; 2) nonlinear scaling [44]; and 3) quantile-mapping, including distribution mapping [45] and empirical (distribution-

free) quantile mapping [46]–[48]. Methods within the three categories differ in their ability to correct higher-order moments of simulated climatological variables. For bias-correcting temperature variables, linear scaling and empirical quantile mapping (EQM) are often used [22], [49]–[52]. EQM, a nonparametric and highly flexible method, can correct the mean, variance, and higher moments of temperature and precipitation variables [22], [53]. Linear scaling is a simple method in which differences between observed monthly means and those of simulated data are added to simulated data [42]. Despite its simplicity, linear scaling has been shown to be effective for bias-correcting temperature variables [42], [54].

In linear scaling (LS), model output of meteorological variables is corrected based on differences between raw model and observed data [42]. It is closely related to the "delta change" and "factor change" methods for bias correction [46]. In the delta change method, an observed time series is added to a climate change signal (CCS), typically by month. In the expressions below, T and P refer to temperature and precipitation time series, respectively; subscripts Obs, Mod, and $Corr$ refer to observed, raw model, and corrected model values, and subscripts f and h denote data in the future period and historical period of model output and single values of time series T and P will be denoted with a subscript i. Variables with a bar (e.g. $\bar{T}$) denote the mean of that variable.

Temperature is corrected in an additive manner (generally by month):

$$T_{Corr,i} = T_{Mod,i} + \bar{T}_{Obs} - \bar{T}_{Mod}.$$

Raw model precipitation is corrected with a multiplicative ratio by month to avoid

negative values:

$$P_{Corr,i} = P_{Mod,i} \cdot \frac{\bar{P}_{Obs}}{\bar{P}_{Mod}}.$$

After the application of LS, the mean of corrected model output will correspond perfectly to those of observed data [42]. LS requires only a modest amount of observed data [46] and has been used used to correct GCM simulations for daily precipitation [42], maximum temperature, and precipitation runoff [43]. LS is limited in that it can only correct the mean of climate variables and does not account for non-stationarity in future climate model projections [22]. Simple scaling methods may also cause large deviations in model errors compared to more sophisticated methods [55].

The power transform method (PT) [56], a non-linear scaling method, corrects the mean and standard deviation and has only been applied to precipitation variables [44], [47]. Daily precipitation values are corrected according to the following equation:

$$P_{corr,i} = aP^{b}_{Mod,i},$$

where $P_{Mod,i}$ is a raw model precipitation value, $P_{corr,i}$ is corrected model precipitation value, and a and b are constants. The scaling exponent, b is first calculated using an iterative method such that the coefficient of variation for model and observed data match on a monthly basis [8]. Then, the scaling constant a is calculated such that corrected precipitation values have a mean equivalent to that of observed data [8]. PT was used to bias-correct precipitation simulations across all distributional moments in Great Britain [8] as well as the Tarim River Basin, China [22]. However, [46] found that although PT corrects precipitation quantiles, the probability of dry days and precipitation intensity may not be adequately represented. Thus, PT may be less

suitable for model data that exhibits large biases in wet-day frequency [44], [46].

Distribution mapping (DM) is a parametric approach within the quantile-mapping family of bias-correction methods. In DM, simulated and observed climate variables are assumed to follow parametric distributions [8]. Bias-correction is accomplished via the use of a transfer function (TF), which minimizes the difference between the cumulative density functions (CDFs) of model climate variables and observed data. Wet-day precipitation is typically modeled with a gamma probability distribution function (PDF) [45], [57]:

$$f(P|\theta,k) = \frac{e^{-P/\theta}P^{k-1}}{\Gamma(k)\theta^k},$$

where $f(P|\theta,k)$ is the Gamma PDF, P is daily precipitation, k and θ are the shape and scale parameters, respectively, and Γ is the Gamma function [58]. Typically, Gamma PDFs are fitted by month, but the performance of DM can be enhanced by constructing one TF for each day of the year, where distribution parameters are estimated over 3-month [59], [60], 61-day [53], or 45-day sliding windows [47]. The DM correction for precipitation is expressed as:

$$P_{Corr,i} = F_P^{-1}(F_P(P_{Mod,i}|\theta_{Mod},k_{Mod})|\theta_{Obs},k_{Obs}), \tag{1.1}$$

where F_P and F_P^{-1} denote the Gamma and inverse Gamma CDFs, respectively, and θ and k are estimated shape and rate parameters of Gamma CDFs for both raw model and observed distributions for a particular month (or sliding window time period). DM for precipitation variables is not limited to modeling with the Gamma distribution. Mixture distributions such as the Bernoulli-Gamma, Bernoulli-Weibull,

Bernoulli-Log-normal, and Bernoulli-exponential have been used for bias-correcting precipitation [61]. While the Gamma distribution and Gamma mixture models are generally adequate for representing mean daily precipitation, Kappa and Pearson Type III distributions [62]–[64], as well as Gamma-Generalized Pareto Distribution mixture models [65]–[68] and Gamma-Gumbel mixture models [69] perform better in capturing extreme tails of wet-day precipitation probabilities compared to the Gamma distribution.

Temperature variables are generally modeled with the Gaussian distribution [46], [70]:

$$f(T|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{((T-\mu)^2}{2\sigma^2}},$$

where $f(T|\mu,\sigma)$ is the Gaussian PDF with mean μ and standard deviation σ. The correction for temperature variables can be expressed as:

$$T_{corr,i} = F_T^{-1}(F_T(T_{Mod,i}|\mu_{Mod},\sigma_{Mod})|\mu_{Obs},\sigma_{Obs}), \tag{1.2}$$

where F_T and F_T^{-1} are the Gaussian and inverse Gaussian CDFs, respectively, and μ and σ are the fitted mean and standard deviation parameters for Gaussian distributions fitted to model and observed data. DM is a widely-used, efficient method and is advantageous because higher order distributional moments can be corrected. DM has limited use in situations where climate variables exhibit poor fits to known distributions [71].

Empirical quantile mapping (EQM) is simular to DM except that the TF is composed of empirical CDFs. EQM can be considered a non-parametric version of DM

[53]. It is a popular method, as all distributional moments of a modeled climate variable can be adjusted to match those of observed data [22], and it can be applied to both precipitation and temperature variables [22], [53]. It generally outperforms simpler bias-correction methods [8], [52], [61], [72], [73], as well as its parametric counterpart, DM [61], [65], [67]. The transfer function for correcting model climate variables can be represented by the empirical cumulative distribution function (ecdf) and its inverse (ecdf^{-1}):

$$X_{Corr,i} = \text{ecdf}_{Obs}^{-1}(\text{ecdf}_{Mod}(X_{Mod,i})),$$

where $X_{corr,i}$ is the corrected value of a meteorological variable, ecdf_{obs}^{-1} is the inverse ecdf of observed data, and ecdf_{Mod} is the ecdf of raw model data, and $X_{Mod,i}$ is the value of raw model data. Figure 1.1 shows how a TF is used to correct model values.

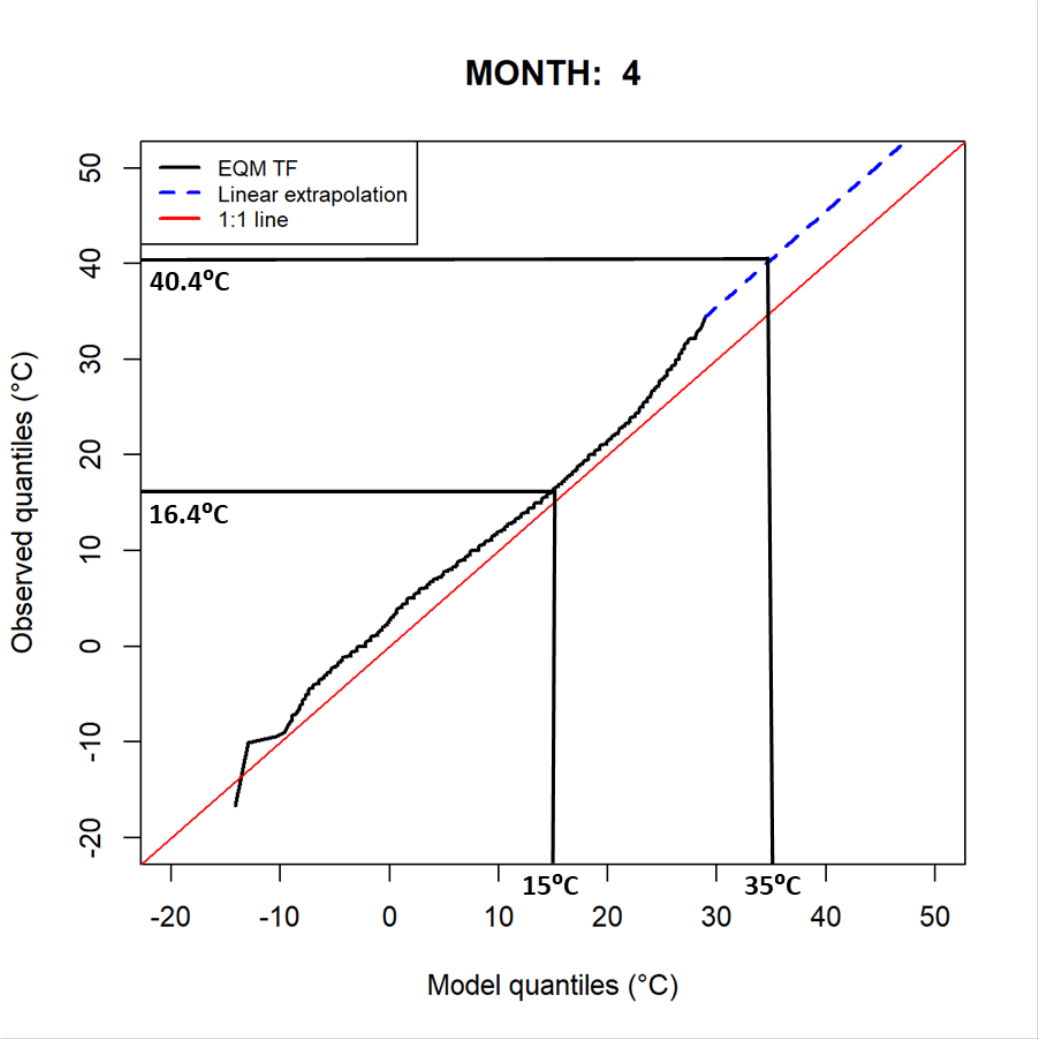

Figure 1.1: EQM TF of daily temperature data for the month of April over the period 1976-2005. Model data was generated by a regional climate model, and observed data was obtained from climate stations within the Global Historical Climate Network [74]. The corrected value is obtained by evaluating the TF (black line) at model quantiles. For example, the corrected value for 15° C is 16.4° C, and the corrected value for 35° C is 40.4° C. Note that the corrected value for 35° C is obtained via a linear extrapolation, as values beyond the extrapolation threshold are beyond the ranges of data used to construct the TF. In this figure, the linear extrapolation by [75] is used, but other extrapolations exist.

.

1.2.3 Challenges in bias-correction

Temporally and spatially coherent bias-correction

Gridded, observational climate products (e.g. Livneh, [76]; Daymet, [77]; and PRISM, [78]) are often used for bias-correction because of their extensive spatial and temporal coverage. However, the interpolation algorithms used to create gridded climate products can introduce bias [79] and additional uncertainty when used for bias-correcting climate model output [80]. Though the spatial resolution of some gridded datasets (e.g. Livneh and Daymet) is quite high (1km), interpolation can result in some degree of smoothing of sharp topographical gradients and regional atmospheric processes that may have a large influence on extremes [81]. A valuable alternative to gridded observational data products are long-term, curated station data, such as data from the Global Historical Climate Network [74]. Station data represent direct climatological measurements and are available globally [82], [83]. The use of station data, rather than gridded observational products, removes some uncertainty during bias-correction. Station data are frequently utilized to validate the accuracy of bias-corrected climate model output but can also be effectual for bias-correcting output from climate models. For instance, [84] downscaled monthly temperature and precipitation simulations from an RCM to climate stations and bias-corrected the simulated climate variables with station data, resulting in an appreciable improvement in the accuracy of a hydrological model. [85] showed that incorporating station data in a geostatistical downscaling and bias-correction approach resulted in full-coverage, high-resolution monthly temperature and precipitation data that better captured the complex topographical features of the study area.

Despite the advantages of station data, its use in constructing full-coverage, bias-corrected, downscaled climate data, particularly at high spatial and temporal resolutions, is limited. The density and spatial distribution of climate stations are often irregular, especially in mountainous and high-elevation regions [78]. While the TFs can be derived from irregularly spaced station data, the resulting TFs must be spatially transferred so that model data in regions void of climate stations can be adequately corrected. For instance, [86] used an interpolation scheme to match stations to downscaled model grid points and then applied EQM to bias-correct the data. However, the task of spatially transferring TFs has generally received little attention in the literature. In Chapter 2 I develop six downscaling and bias-correction workflows that allow for spatially-coherent bias-correction based on the use of station data and evaluate their effectiveness in bias-correcting daily maximum temperature simulations generated by an RCM. I quantified performance of the six methods with the root-mean-square-error (RMSE) and Perkins skill score (PSS) and used two ANOVA models to analyze how performance varied among methods. I validated the six methods using two calibration periods of observed data (1980-1989 and 1980-2014) and two testing sets of RCM data (1990-2014 and 1980-2014).

It is important that temporal variability is realistically represented in bias-corrected climate model output, as the modeling of extreme events in modeling applications is influenced by temporal variability [87]. Additionally, factors such as soil moisture, sunlight, and heat that are important to human health are influenced by variability in a variety of meteorological variables as well as extreme climate events [88]. The correction of temporal dependence as part of a bias-correction method has received less attention in the literature. [89] and [90] developed a multivariate bias-correction

method that resulted in the correction of spatial and temporal correlation structures. In nested bias-correction (NBC) [91]–[93], temporal biases at multiple, pre-defined time scales are corrected. For instance, [91] used NBC to correct the mean, standard deviation, and lag-1 autocorrelation of daily GCM precipitation simulations at annual and monthly scales. In an extension to NBC, [91] performed the correction at daily, monthly, seasonal, annual, and tri-annual timescales. Though the nested approaches in [91] and [94] resulted in better representation of temporal structures of GCM simulations compared to EQM, these approaches assume a temporal dependence structure (e.g. lag-1 linear autoregressive model) and stationarity of model errors, and require the modeler to choose the temporal scale(s) in which the correction should be reflected. Bias-correction of temporal dependence has also been done via signal processing techniques. [95] used the Fast Fourier Transform to decompose observed and model time series; temporal dependence was then corrected via EQM in the frequency domain. However, this approach assumes stationarity, and may limit its use for bias-correcting future model simulations. In a trend-preserving approach, [96] decomposed observed and model time series using the discrete wavelet transform (DWT) and corrected biases in mean, standard deviation, and temporal variability in the frequency domain. However, neither of these studies considered non-annual variability or variability at fine (sub-monthly) temporal frequencies. In Chapter 4, I propose a novel trend-preserving method for bias-correction of temperature variables that features a process convolution approach for correcting temporal dependence.

Bias-correction of precipitation extremes

Because extreme climate events have disproportionately large impacts on human and natural systems, it is important that climate products include plausible scenarios of extreme climate events as well as information on the frequency, intensity and future trends of extreme events [97], [98]. The ability of GCMs and RCMs to reproduce extreme tails of climate variables is limited [44]. Therefore, bias-correction methods for extremes become increasingly important.

Bias-correction of extremes is a challenging task, due to the scarcity of values in extreme tails of model and observed data and uncertainty regarding correction of extreme tails. Bias-correction of precipitation extremes is especially difficult, due to the high variability and skewed distributional properties of precipitation data.

A disadvantage of QM methods, and EQM in particular, is its propensity to overfit on calibration data, especially at precipitation extremes where data is scarce and highly variable [8], [45], [99]–[101]. In EQM, TFs are interpolated using linear interpolation, splines, or other smoothing techniques [102]. Highly flexible methods such as EQM can generate TFs that correct model data nearly perfectly (overfitting) but may not generalize to out-of-sample or future model data, which can lead to instability of the TF at higher quantiles [99], [103], [104]. When applied to future projections, EQM has been shown to significantly distort future climate change signals [7], [105] and exaggerate or deflate extreme trends, introducing additional uncertainty into bias-corrected data [47], [106].

Hybrid EQM approaches that combine parametric and non-parametric modeling can reduce the degree of overfitting of the TF at extreme tails [106]. In such hybrid approaches, bias-correction below a specified threshold is achieved via an empirical

TF (EQM), while bias-correction above the threshold is with DM, based on heavy-tailed distributions, such as the exponential, Generalized Pareto, Weibull, or Gumbel distributions [65]–[68], [101], [106]–[108]. Hybrid EQM methods combine the flexbility of EQM for correcting lower to middle quantiles with the robustness of parametric distributions for correcting upper quantiles. The use of extreme or heavy-tailed distributions for modeling extremes can improve bias-correction of tail quantiles [66], [101], [106]–[109], but the risk of overfitting the TF at distributional tails still exists, as poor fits to heavy-tailed distributions can introduce outliers [69], [110]. In addition, selection of the threshold is difficult, as the amount of data beyond the threshold must be sufficiently large to allow for distribution fitting and must approximate a known heavy-tailed distribution [66], [111]. There is a need for a hybrid EQM method in which bias-correction of extremes can be performed without the risk of overfitting and the introduction of outliers. In Chapter 3, I propose such a hybrid approach, EQM-LIN, that combines the efficacy of EQM for correcting lower quantiles, with a robust linear correction for upper quantiles.

Stationarity of model errors

One of the major assumptions of many bias-correction methods is that of stationary model errors over time [55]. In climate statistics, stationarity usually refers to weak or second-order stationarity (a less restrictive definition than strong stationarity) in model *errors* and thus does not refer to the obvious seasonal patterns most climate variables exhibit [105]. In a weakly or second-order stationary process, the mean and variance are constant, and the covariance between values can be described by a covariance function which only depends on the separation distance between the values

[112]. The assumption of stationary model errors over time implies that the correction established during the calibration period can be applied outside the calibration period as well [113]. Clearly, the stationarity assumption warrants some scrutiny, and the validity of this assumption has been debated in the literature [7], [47], [105], [113]–[115]. Though there is consensus that future model projections should be bias-corrected with respect to observations [7], there is some debate whether the general future climate change signal (CCS) should be preserved [7], [47], [113], [116] or not [103], [115]. Highly flexible bias-correction methods such as EQM are susceptible to overfitting on (historical) calibration data and have been shown to significantly misrepresent the CCS of future climate model output [7], [47], [105] [18]. The distortion of the CCS can be particularly severe for higher distributional quantiles, which could ultimately result in over- or under-estimation of extreme floods, droughts, and heat waves [47], [106], [116]. The application of a stationary bias-correction method to non-stationary data will also add additional uncertainty into bias-corrected data [7], [114]. So-called trend-preserving bias-correction methods attempt to leverage differences between historical model simulations and future model projections such that the bias-correction adequately resolves model bias while preserving the mean CCS [47]. Effective trend-preserving methods represent variations on EQM and include popular methods such as equidistant quantile matching (EDQM) [116], quantile delta mapping (QDM) [47], and scaled distribution mapping (SDM) [113].

Generally, in trend-preserving variants of EQM, differences between the distributions of climate variables in historical (e.g. calibration) and future time periods are taken into account. These methods consist of two main steps: (1) calculating absolute or relative changes between quantiles of model data in the calibration and

future time periods, and 2) bias-correcting future model projections via additive or multiplicative operations.

In EDQM, future model projected data are corrected in an additive manner, where future quantiles, normalized relative to model simulations during the calibration period, are used to adjust the degree to which future data are corrected [116]. For precipitation, the correction is achieved via 1.3:

$$P_{corr,i} = P_{Mod,i} + ecdf_{Obs}^{-1}(F_f(P_{Mod,i})) - F_{Mod,h}^{-1}(F_{Mod,h}(P_{Mod,i})). \tag{1.3}$$

Like in EQM, TFs in EDQM are generally constructed by month. [116] demonstrated that EDQM significantly decreased model bias relative to observed data, especially at low and high quantiles. [117] showed that EDQM better reproduced rainfall intensity and return intervals of extreme events at climate stations in Canada compared to non-trend-preserving quantile mapping-based methods. Finally, [47] proposed a method fundamentally equivalent to EDQM suitable for correcting precipitation, and showed that, compared to EQM, EDQM resulted in realistic corrections of future daily precipitation projections especially for upper quantiles.

SDM is a multi-step approach, that, unlike EDQM, scales the distribution of observed data according to projected future changes in the frequency and intensity of rain days, as well as the likelihood of individual rainfall events, although it can also be applied to temperature variables [113]. [113] found that constructing TFs using parametric distributions resulted in more stable results compared to empirical CDFs. [113] reported that SDM outperformed EDQM, EQM, and QDM in preserving the future CCS of model data. In practice, SDM is somewhat limited as it is computationally intensive. In Chapter 4, I propose a novel trend-preserving method for

bias-correction of temperature variables that features a process convolution approach for correcting temporal dependence. I found that the proposed method adjusts temporal dependence of model data such that it resembles that of observed data, and distributional biases of model data were corrected to a greater degree compared to EQM.

References

[1] L. C. B. Program, *Recreation resources*, Accessed: 2019-09-06, 2019. [Online]. Available: https://www.lcbp.org/culture-recreation/recreation-resources/.

[2] D. N. Karger, O. Conrad, J. Böhner, *et al.*, "Climatologies at high resolution for the earth's land surface areas," *Scientific data*, vol. 4, no. 1, pp. 1–20, 2017.

[3] A. Hamann and T. Wang, "Potential effects of climate change on ecosystem and tree species distribution in british columbia," *Ecology*, vol. 87, no. 11, pp. 2773–2786, 2006.

[4] A. Zia, A. Bomblies, A. W. Schroth, *et al.*, "Coupled impacts of climate and land use change across a river–lake continuum: Insights from an integrated assessment model of lake champlain's missisquoi basin, 2000–2040," *Environmental Research Letters*, vol. 11, no. 11, p. 114 026, 2016.

[5] M. Ekström, M. R. Grose, and P. H. Whetton, "An appraisal of downscaling methods used in climate change research," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 6, no. 3, pp. 301–319, 2015.

[6] L. R. Leung, L. O. Mearns, F. Giorgi, and R. L. Wilby, "Regional climate research: Needs and opportunities," *Bulletin of the American Meteorological Society*, vol. 84, no. 1, pp. 89–95, 2003.

[7] D. Maraun, T. G. Shepherd, M. Widmann, *et al.*, "Towards process-informed bias correction of climate change simulations," *Nature Climate Change*, vol. 7, no. 11, p. 764, 2017.

[8] T. Lafon, S. Dadson, G. Buys, and C. Prudhomme, "Bias correction of daily precipitation simulated by a regional climate model: A comparison of methods," *International Journal of Climatology*, vol. 33, no. 6, pp. 1367–1381, 2013.

[9] A. J. Cannon, C. Piani, and S. Sippel, "Bias correction of climate model output for impact models," in *Climate Extremes and Their Implications for Impact and Risk Assessment*, Elsevier, 2020, pp. 77–104.

[10] I. Hanssen-Bauer, C. Achberger, R. Benestad, D. Chen, and E. Førland, "Statistical downscaling of climate scenarios over Scandinavia," *Climate Research*, vol. 29, no. 3, pp. 255–268, 2005.

[11] Z. A. Holden, J. T. Abatzoglou, C. H. Luce, and L. S. Baggett, "Empirical downscaling of daily minimum air temperature at very fine resolutions in complex terrain," *Agricultural and Forest Meteorology*, vol. 151, no. 8, pp. 1066–1073, 2011.

[12] G. E. Liston and K. Elder, "A meteorological distribution system for high-resolution terrestrial modeling (micromet)," *Journal of Hydrometeorology*, vol. 7, no. 2, pp. 217–234, 2006.

[13] P. Caldwell, H.-N. S. Chin, D. C. Bader, and G. Bala, "Evaluation of a WRF dynamical downscaling simulation over California," *Climatic change*, vol. 95, no. 3-4, pp. 499–521, 2009.

[14] R. L. Wilby, S. Charles, E. Zorita, B. Timbal, P. Whetton, and L. Mearns, "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 TGCIA*, vol. 27, 2004.

[15] F. Feser, B. Rockel, H. von Storch, J. Winterfeldt, and M. Zahn, "Regional climate models add value to global model data: A review and selected examples," *Bulletin of the American Meteorological Society*, vol. 92, no. 9, pp. 1181–1192, 2011.

[16] F. Giorgi, C. Jones, G. R. Asrar, *et al.*, "Addressing climate information needs at the regional level: The cordex framework," *World Meteorological Organization Journal*, vol. 58, no. 3, p. 175, 2009.

[17] X. Gao, J. S. Pal, and F. Giorgi, "Projected changes in mean and extreme precipitation over the mediterranean region from a high resolution double nested rcm simulation," *Geophysical Research Letters*, vol. 33, no. 3, 2006.

[18] D. Maraun, "Bias correcting climate change simulations-a critical review," *Current Climate Change Reports*, vol. 2, no. 4, pp. 211–220, 2016.

[19] H. Fowler and M. Ekström, "Multi-model ensemble estimates of climate change impacts on uk seasonal precipitation extremes," *International Journal of Climatology: A Journal of the Royal Meteorological Society*, vol. 29, no. 3, pp. 385–416, 2009.

[20] R. Seager and G. A. Vecchi, "Greenhouse warming and the 21st century hydroclimate of southwestern North America," *Proceedings of the National Academy of Sciences*, vol. 107, no. 50, pp. 21 277–21 282, 2010.

[21] L. Mearns, F. Giorgi, P. Whetton, D. Pabon, M. Hulme, and M. Lal, "Guidelines for use of climate scenarios developed from regional climate model experiments," *Data Distribution Centre of the Intergovernmental Panel on Climate Change*, 2003.

[22] G. Fang, J. Yang, Y. Chen, and C. Zammit, "Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in china," *Hydrology and Earth System Sciences*, vol. 19, no. 6, pp. 2547–2559, 2015.

[23] R. Huth, "Statistical downscaling in central europe: Evaluation of methods and potential predictors," *Climate Res.*, vol. 13, no. 2, pp. 91–101, 1999.

[24] H. Kettle and R. Thompson, "Statistical downscaling in european mountains: Verification of reconstructed air temperature," *Climate Res.*, vol. 26, no. 2, pp. 97–112, 2004.

[25] J. T. Schoof and S. Pryor, "Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks," *International Journal of Climate*, vol. 21, no. 7, pp. 773–790, 2001.

[26] T. Vandal, E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, "Deepsd: Generating high resolution climate change projections through single image super-resolution," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2017, pp. 1663–1672.

[27] C. Hutengs and M. Vohland, "Downscaling land surface temperatures at regional scales with random forest regression," *Remote Sensing of Environment*, vol. 178, pp. 127–141, 2016.

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

[29] D. Jeong, A. St-Hilaire, T. Ouarda, and P. Gachon, "Comparison of transfer functions in statistical downscaling models for daily temperature and precipitation over canada," *Stochastic environmental research and risk assessment*, vol. 26, no. 5, pp. 633–653, 2012.

[30] R. L. Wilby and T. Wigley, "Downscaling general circulation model output: A review of methods and limitations," *Progress in physical geography*, vol. 21, no. 4, pp. 530–548, 1997.

[31] J. M. Winter, B. Beckage, G. Bucini, R. M. Horton, and P. J. Clemins, "Development and evaluation of high-resolution climate simulations over the mountainous northeastern united states," *Journal of Hydrometeorology*, vol. 17, no. 3, pp. 881–896, 2016.

[32] F. Wang, D. Tian, L. Lowe, L. Kalin, and J. Lehrter, "Deep learning for daily precipitation and temperature downscaling," *Water Resources Research*, vol. 57, no. 4, e2020WR029308, 2021.

[33] B. Pang, J. Yue, G. Zhao, and Z. Xu, "Statistical downscaling of temperature with the random forest model," *Advances in Meteorology*, vol. 2017, 2017.

[34] T. Vandal, E. Kodra, and A. R. Ganguly, "Intercomparison of machine learning methods for statistical downscaling: The case of daily and extreme precipitation," *Theoretical and Applied Climatology*, vol. 137, no. 1, pp. 557–570, 2019.

[35] D. Sachindra, K. Ahmed, M. M. Rashid, S. Shahid, and B. Perera, "Statistical downscaling of precipitation using machine learning techniques," *Atmospheric research*, vol. 212, pp. 240–258, 2018.

[36] X. Li, Z. Li, W. Huang, and P. Zhou, "Performance of statistical and machine learning ensembles for daily temperature downscaling," *Theoretical and Applied Climatology*, vol. 140, no. 1, pp. 571–588, 2020.

[37] J. Baño-Medina, R. Manzanas, and J. M. Gutiérrez, "Configuration and intercomparison of deep learning neural models for statistical downscaling," *Geoscientific Model Development*, vol. 13, no. 4, pp. 2109–2124, 2020.

[38] T. Bong, Y.-H. Son, S.-H. Yoo, and S.-W. Hwang, "Nonparametric quantile mapping using the response surface method–bias correction of daily precipitation," *Journal of Water and Climate Change*, vol. 9, no. 3, pp. 525–539, 2018.

[39] S. Pour, S. Harun, and S. Shahid, "Genetic programming for the downscaling of extreme rainfall events on the east coast of peninsular malaysia," *Atmosphere*, vol. 5, no. 4, pp. 914–936, 2014.

[40] R. M. B. Harris, M. R. Grose, G. Lee, N. L. Bindoff, L. L. Porfirio, and P. Fox-Hughes, "Climate projections for ecologists," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 5, no. 5, pp. 621–637, 2014.

[41] I. Takayabu, H. Kanamaru, K. Dairaku, R. Benestad, H. von Storch, and J. H. Christensen, "Reconsidering the quality and utility of downscaling," *Journal of the Meteorological Society of Japan. Ser. II*, vol. 94, pp. 31–45, 2016.

[42] G. Lenderink, A. Buishand, and W. v. Deursen, "Estimates of future discharges of the river rhine using two scenario methodologies: Direct versus delta approach," *Hydrology and Earth System Sciences*, vol. 11, no. 3, pp. 1145–1159, 2007.

[43] L. E. Hay, R. L. Wilby, and G. H. Leavesley, "A comparison of delta change and downscaled gcm scenarios for three mountainous basins in the united states 1," *JAWRA Journal of the American Water Resources Association*, vol. 36, no. 2, pp. 387–397, 2000.

[44] R. Leander and T. A. Buishand, "Resampling of regional climate model output for the simulation of extreme river flows," *Journal of Hydrology*, vol. 332, no. 3-4, pp. 487–496, 2007.

[45] C. Piani, J. Haerter, and E. Coppola, "Statistical bias correction for daily precipitation in regional climate models over europe," *Theoretical and Applied Climatology*, vol. 99, no. 1-2, pp. 187–192, 2010.

[46] C. Teutschbein and J. Seibert, "Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods," *Journal of Hydrology*, vol. 456, pp. 12–29, 2012.

[47] A. J. Cannon, S. R. Sobie, and T. Q. Murdock, "Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?" *Journal of Climate*, vol. 28, no. 17, pp. 6938–6959, 2015.

[48] A. W. Wood, E. P. Maurer, A. Kumar, and D. P. Lettenmaier, "Long-range experimental hydrologic forecasting for the eastern united states," *Journal of Geophysical Research: Atmospheres*, vol. 107, no. D20, ACL–6, 2002.

[49] E. P. Maurer and P. B. Duffy, "Uncertainty in projections of streamflow changes due to climate change in california," *Geophysical Research Letters*, vol. 32, no. 3, 2005.

[50] K. Hayhoe, C. Wake, B. Anderson, *et al.*, "Regional climate change projections for the northeast usa," *Mitigation and Adaptation Strategies for Global Change*, vol. 13, no. 5-6, pp. 425–436, 2008.

[51] A. W. Wood, L. R. Leung, V. Sridhar, and D. Lettenmaier, "Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs," *Climatic Change*, vol. 62, no. 1-3, pp. 189–216, 2004.

[52] J. C. Bennett, M. R. Grose, S. P. Corney, *et al.*, "Performance of an empirical bias-correction of a high-resolution climate dataset," *International Journal of Climatology*, vol. 34, no. 7, pp. 2189–2204, 2014.

[53] M. Jakob Themeßl, A. Gobiet, and A. Leuprecht, "Empirical-statistical downscaling and error correction of daily precipitation from regional climate models," *International Journal of Climatology*, vol. 31, no. 10, pp. 1530–1544, 2011.

[54] M. Shrestha, S. C. Acharya, and P. K. Shrestha, "Bias correction of climate models for hydrological modelling–are simple methods still useful?" *Meteorological Applications*, vol. 24, no. 3, pp. 531–539, 2017.

[55] D. R. Roberts, W. H. Wood, and S. J. Marshall, "Assessments of downscaled climate data with a high-resolution weather station network reveal consistent but predictable bias," *International Journal of Climatology*, vol. 39, no. 6, pp. 3091–3103, 2019.

[56] M. Shabalova, W. Van Deursen, and T. Buishand, "Assessing future discharge of the river rhine using regional climate model integrations and a hydrological model," *Climate research*, vol. 23, no. 3, pp. 233–246, 2003.

[57] R. W. Katz, "Extreme value theory for precipitation: Sensitivity analysis for climate change," *Advances in Water Resources*, vol. 23, no. 2, pp. 133–139, 1999.

[58] D. S. Wilks, *Statistical methods in the atmospheric sciences*. Academic press, 2011, vol. 100.

[59] P. Smitha, B. Narasimhan, K. Sudheer, and H. Annamalai, "An improved bias correction method of daily rainfall data using a sliding window technique for climate change impact assessment," *Journal of Hydrology*, vol. 556, pp. 100–118, 2018.

[60] O. Räty, J. Räisänen, and J. S. Ylhäisi, "Evaluation of delta change and bias correction methods for future daily precipitation: Intermodel cross-validation using ensembles simulations," *Climate dynamics*, vol. 42, no. 9-10, pp. 2287–2303, 2014.

[61] L. Gudmundsson, J. Bremnes, J. Haugen, and T. Engen-Skaugen, "Downscaling rcm precipitation to the station scale using statistical transformations–a comparison of methods," *Hydrology and Earth System Sciences*, vol. 16, no. 9, pp. 3383–3390, 2012.

[62] X. Qin and C. Dai, "Comparison of different quantile delta mapping schemes in frequency analysis of precipitation extremes over mainland southeast asia under climate change," *Journal of Hydrology*, p. 127 421, 2022.

[63] J.-H. Heo, H. Ahn, J.-Y. Shin, T. R. Kjeldsen, and C. Jeong, "Probability distributions for a quantile mapping technique for a bias correction of precipitation data: A case study to precipitation data under climate change," *Water*, vol. 11, no. 7, p. 1475, 2019.

[64] L. Ye, L. S. Hanson, P. Ding, D. Wang, and R. M. Vogel, "The probability distribution of daily precipitation at the point and catchment scales in the united states," *Hydrology and Earth System Sciences*, vol. 22, no. 12, pp. 6519–6531, 2018.

[65] P. Reiter, O. Gutjahr, L. Schefczyk, G. Heinemann, and M. Casper, "Does applying quantile mapping to subsamples improve the bias correction of daily precipitation?" *International Journal of Climatology*, vol. 38, no. 4, pp. 1623–1633, 2018.

[66] O. Gutjahr and G. Heinemann, "Comparing precipitation bias correction methods for high-resolution regional climate simulations using cosmo-clm," *Theoretical and applied climatology*, vol. 114, no. 3, pp. 511–529, 2013.

[67] M. A. Ivanov and S. Kotlarski, "Assessing distribution-based climate model bias correction methods over an alpine domain: Added value and limitations," *International Journal of Climatology*, vol. 37, no. 5, pp. 2633–2653, 2017.

[68] C. Volosciuk, D. Maraun, M. Vrac, and M. Widmann, "A combined statistical bias correction and stochastic downscaling method for precipitation," *Hydrology and Earth System Sciences*, vol. 21, no. 3, pp. 1693–1719, 2017.

[69] J.-Y. Shin, T. Lee, T. Park, and S. Kim, "Bias correction of rcm outputs using mixture distributions under multiple extreme weather influences," *Theoretical and Applied Climatology*, vol. 137, no. 1, pp. 201–216, 2019.

[70] L. E. Hay and M. Clark, "Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western united states," *Journal of Hydrology*, vol. 282, no. 1-4, pp. 56–75, 2003.

[71] S. McGinnis, D. Nychka, and L. O. Mearns, "A new distribution mapping technique for climate model bias correction," in *Machine learning and data mining approaches to climate science*, Springer, 2015, pp. 91–99.

[72] V. M. Garibay, M. W. Gitau, N. Kiggundu, D. Moriasi, and F. Mishili, "Evaluation of reanalysis precipitation data and potential bias correction methods for use in data-scarce areas," *Water Resources Management*, pp. 1–16, 2021.

[73] C. Miao, L. Su, Q. Sun, and Q. Duan, "A nonstationary bias-correction technique to remove bias in gcm simulations," *Journal of Geophysical Research: Atmospheres*, vol. 121, no. 10, pp. 5718–5735, 2016.

[74] N. Oceanic and A. Administration, *Climate data online search*, Accessed: 2017-09-30, 2018. [Online]. Available: https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND.

[75] J. Boé, L. Terray, F. Habets, and E. Martin, "Statistical and dynamical downscaling of the seine basin climate for hydro-meteorological studies," *International Journal of Climatology: A Journal of the Royal Meteorological Society*, vol. 27, no. 12, pp. 1643–1655, 2007.

[76] B. Livneh, T. J. Bohn, D. W. Pierce, *et al.*, "A spatially comprehensive, hydrometeorological data set for mexico, the us, and southern canada 1950–2013," *Scientific data*, vol. 2, no. 1, pp. 1–12, 2015.

[77] P. E. Thornton, M. M. Thornton, B. W. Mayer, *et al.*, "Daymet: Daily surface weather on a 1 km grid for north america, 1980-2008," *Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center for Biogeochemical Dynamics (DAAC)*, 2012.

[78] C. Daly, G. Taylor, W. Gibson, T. Parzybok, G. Johnson, and P. Pasteris, "High-quality spatial climate data sets for the united states and beyond," *Transactions of the ASAE*, vol. 43, no. 6, p. 1957, 2000.

[79] R. Behnke, S. Vavrus, A. Allstadt, T. Albright, W. E. Thogmartin, and V. C. Radeloff, "Evaluation of downscaled, gridded climate data for the conterminous united states," *Ecological applications*, vol. 26, no. 5, pp. 1338–1351, 2016.

[80] D. Walton and A. Hall, "An assessment of high-resolution gridded temperature datasets over california," *Journal of Climate*, vol. 31, no. 10, pp. 3789–3810, 2018.

[81] A. M. Wootten, K. W. Dixon, D. J. Adams-Smith, and R. A. McPherson, "Statistically downscaled precipitation sensitivity to gridded observation data and downscaling technique," *International Journal of Climatology*, 2020.

[82] T. C. Peterson and R. S. Vose, "An overview of the global historical climatology network temperature database," *Bulletin of the American Meteorological Society*, vol. 78, no. 12, pp. 2837–2850, 1997.

[83] I. Durre, M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, "Comprehensive automated quality assurance of daily surface observations," *Journal of Applied Meteorology and Climatology*, vol. 49, no. 8, pp. 1615–1633, 2010.

[84] J. F. Mejia, J. Huntington, B. Hatchett, D. Koracin, and R. G. Niswonger, "Linking global climate models to an integrated hydrologic model: Using an individual station downscaling approach," *J. Contemp. Water Res. Educ.*, vol. 147, no. 1, pp. 17–27, 2012.

[85] L. Poggio and A. Gimona, "Downscaling and correction of regional climate models outputs with a hybrid geostatistical approach," *Spat. Stat.*, vol. 14, pp. 4–21, 2015.

[86] J. Rajczak, S. Kotlarski, N. Salzmann, and C. Schaer, "Robust climate scenarios for sites with sparse observations: A two-step bias correction approach," *International Journal of Climatology*, vol. 36, no. 3, pp. 1226–1243, 2016.

[87] F. Lehner, I. Nadeem, and H. Formayer, "Evaluating quantile-based bias adjustment methods for climate change scenarios," *Hydrology and Earth System Sciences Discussions*, pp. 1–26, 2021.

[88] A. K. Mishra, V. P. Singh, and S. K. Jain, "Impact of global warming and climate change on social development," *Journal of Comparative Social Welfare*, vol. 26, no. 2-3, pp. 239–260, 2010.

[89] M. Clark, S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, "The schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields," *Journal of Hydrometeorology*, vol. 5, no. 1, pp. 243–262, 2004.

[90] M. Vrac and P. Friederichs, "Multivariate—intervariable, spatial, and temporal—bias correction," *Journal of Climate*, vol. 28, no. 1, pp. 218–237, 2015.

[91] F. Johnson and A. Sharma, "A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations," *Water Resources Research*, vol. 48, no. 1, 2012.

[92] F. Johnson and A. Sharma, "Assessing future droughts in australia-a nesting model to correct for long-term persistence in general circulation model precipitation simulations," in *8th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation*, Modell., Simul. Soc. of Aust., and NZ, Cairns Queensland, Australia, 2009, pp. 3935–3941.

[93] F. Johnson, S. Westra, A. Sharma, and A. J. Pitman, "An assessment of gcm skill in simulating persistence across multiple time scales," *Journal of Climate*, vol. 24, no. 14, pp. 3609–3623, 2011.

[94] R. Mehrotra and A. Sharma, "An improved standardization procedure to remove systematic low frequency variability biases in gcm simulations," *Water Resources Research*, vol. 48, no. 12, 2012.

[95] H. Nguyen, R. Mehrotra, and A. Sharma, "Correcting for systematic biases in gcm simulations in the frequency domain," *Journal of Hydrology*, vol. 538, pp. 117–126, 2016.

[96] C. Kusumastuti, Z. Jiang, R. Mehrotra, and A. Sharma, "A signal processing approach to correct systematic bias in trend and variability in climate model simulations," *Geophysical Research Letters*, vol. 48, no. 13, e2021GL092953, 2021.

[97] J. Haerter, S. Hagemann, C. Moseley, and C. Piani, "Climate model bias correction and the role of timescales," *Hydrology and Earth System Sciences*, vol. 15, no. 3, pp. 1065–1079, 2011.

[98] H. Fowler, M. Ekström, S. Blenkinsop, and A. Smith, "Estimating change in extreme european precipitation using a multimodel ensemble," *Journal of Geophysical Research: Atmospheres*, vol. 112, no. D18, 2007.

[99] M. G. Grillakis, A. G. Koutroulis, and I. K. Tsanis, "Multisegment statistical bias correction of daily gcm precipitation output," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 8, pp. 3150–3162, 2013.

[100] M. F. Holthuijzen, B. Beckage, P. J. Clemins, D. Higdon, and J. M. Winter, "Constructing high-resolution, bias-corrected climate products: A comparison of methods," *Journal of Applied Meteorology and Climatology*, vol. 60, no. 4, pp. 455–475, 2021.

[101] A. Mamalakis, A. Langousis, R. Deidda, and M. Marrocu, "A parametric approach for simultaneous bias correction and high-resolution downscaling of climate model rainfall," *Water Resources Research*, vol. 53, no. 3, pp. 2149–2170, 2017.

[102] L. Gudmundsson, *Qmap: Statistical transformations for post-processing climate model output*, R package version 1.0-4, 2016.

[103] A. Gobiet, M. Suklitsch, and G. Heinrich, "The effect of empirical-statistical correction of intensity-dependent model errors on the temperature climate change signal," *Hydrology and Earth System Sciences*, vol. 19, no. 10, pp. 4055–4066, 2015.

[104] J. Hnilica, M. Hanel, and V. Puš, "Multisite bias correction of precipitation data from regional climate models," *International Journal of Climatology*, vol. 37, no. 6, pp. 2934–2946, 2017.

[105] M. G. Grillakis, A. G. Koutroulis, I. N. Daliakopoulos, and I. K. Tsanis, "A method to preserve trends in quantile mapping bias correction of climate modeled temperature," *Earth System Dynamics*, vol. 8, no. 3, p. 889, 2017.

[106] S. Tani and A. Gobiet, "Quantile mapping for improving precipitation extremes from regional climate models," *Journal of Agrometeorology*, vol. 21, no. 4, pp. 434–443, 2019.

[107] W. Yang, J. Andréasson, L. Phil Graham, J. Olsson, J. Rosberg, and F. Wetterhall, "Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies," *Hydrology Research*, vol. 41, no. 3-4, pp. 211–229, 2010.

[108] E. M. Laflamme, E. Linder, and Y. Pan, "Statistical downscaling of regional climate model output to achieve projections of precipitation extremes," *Weather and climate extremes*, vol. 12, pp. 15–23, 2016.

[109] D.-I. Kim, H.-H. Kwon, and D. Han, "Exploring the long-term reanalysis of precipitation and the contribution of bias correction to the reduction of uncertainty over south korea: A composite gamma-pareto distribution approach to the bias correction," *Hydrology and Earth System Sciences Discussions*, pp. 1–53, 2018.

[110] M. Luo, T. Liu, F. Meng, *et al.*, "Comparing bias correction methods used in downscaling precipitation and temperature from regional climate models: A case study from the kaidu river basin in western china," *Water*, vol. 10, no. 8, p. 1046, 2018.

[111] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels, *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.

[112] O. Schabenberger and C. A. Gotway, *Statistical methods for spatial data analysis*. Chapman and Hall/CRC, 2017.

[113] M. B. Switanek, P. A. Troch, C. L. Castro, *et al.*, "Scaled distribution mapping: A bias correction method that preserves raw climate model projected changes," *Hydrology and Earth System Sciences*, vol. 21, no. 6, pp. 2649–2666, 2017.

[114] S. Hagemann, C. Chen, J. O. Haerter, J. Heinke, D. Gerten, and C. Piani, "Impact of a statistical bias correction on the projected hydrological changes obtained from three gcms and two hydrology models," *Journal of Hydrometeorology*, vol. 12, no. 4, pp. 556–578, 2011.

[115] M. A. Ivanov, J. Luterbacher, and S. Kotlarski, "Climate model biases and modification of the climate change signal by intensity-dependent bias correction," *Journal of Climate*, vol. 31, no. 16, pp. 6591–6610, 2018.

[116] H. Li, J. Sheffield, and E. F. Wood, "Bias correction of monthly precipitation and temperature fields from intergovernmental panel on climate change ar4 models using equidistant quantile matching," *Journal of Geophysical Research: Atmospheres*, vol. 115, no. D10, 2010.

[117] R. K. Srivastav, A. Schardong, and S. P. Simonovic, "Equidistance quantile matching method for updating idfcurves under climate change," *Water resources management*, vol. 28, no. 9, pp. 2539–2562, 2014.

Chapter 2

Constructing high-resolution, bias-corrected climate products: a comparison of methods

2.1 Abstract

High-resolution, bias-corrected climate data is necessary for climate impact studies at local scales. Gridded historical data is convenient for bias-correction but may contain biases resulting from interpolation. Long-term, quality-controlled station data represent true climatological measurements, but as the distribution of climate stations is irregular, station data are challenging to incorporate into downscaling and bias-correction approaches. Here, we compared six novel methods for constructing full-coverage, high-resolution, bias-corrected climate products using daily maximum temperature simulations from a regional climate model (RCM). Only station data

were used for bias-correction. We quantified performance of the six methods with the root-mean-square-error (RMSE) and Perkins skill score (PSS) and used two ANOVA models to analyze how performance varied among methods. We validated the six methods using two calibration periods of observed data (1980-1989 and 1980-2014) and two testing sets of RCM data (1990-2014 and 1980-2014). RMSE for all methods varied throughout the year and was larger in cold months, while PSS was more consistent. Quantile-mapping bias-correction techniques substantially improved PSS, while simple linear transfer functions performed best in improving RMSE. For the 1980-1989 calibration period, simple quantile-mapping techniques outperformed empirical quantile mapping (EQM) in improving PSS. When calibration and testing time periods were equivalent, EQM resulted in the largest improvements in PSS. No one method produced substantial improvements in both RMSE and PSS. Our results indicate that simple quantile-mapping techniques are less prone to overfitting than EQM and are suitable for processing future climate model output, while EQM is ideal for bias-correcting historical climate model output.

2.2 Introduction

High-resolution ($\leq$1km) gridded climate products with both fine spatial and temporal resolutions are crucial to assessing the effects of a changing climate on social and ecological systems at local scales [11], [118], [119]. Such products are important for climate impact assessments [4], agricultural modeling [120], and ecological studies [11], [121]. General circulation models (GCMs) provide useful information about larger-scale climate, but their spatial resolution (100 - 450km) is too coarse to gain

insight into localized responses to climate change [5], [8]. In addition, GCMs simplify climate processes through parameterization schemes, resulting in the unrealistic representation of some climate processes [7]. Consequently, output from GCMs is characterized by a non-trivial degree of bias [7]–[9]. Typically, post-processing steps such as downscaling and bias-correction are applied to climate model output prior to its use in applications or other downstream models.

In the downscaling process, output generated by climate models is transformed from a coarse to finer resolution. The two main types of downscaling are *dynamical* and *statistical*. In *dynamical downscaling*, a regional climate model (RCM) is forced by GCM or reanalysis data. An RCM simulates climate processes at a finer resolution than forcing data by incorporating fine-scale landscape and atmospheric processes [5], [6], [13], [14]. RCMs are computationally intensive, although they typically require less processing power than GCMs [15], [16]. *Statistical downscaling*, in contrast, involves establishing statistical relationships between coarse-scale and fine-scale climate variables, often leveraging local, observed phenomena or attributes [14]. Statistical downscaling is computationally efficient and can be applied to both precipitation and temperature [21], [22]. In contrast to dynamical downscaling, a substantial amount of observational data is necessary to derive statistical relationships necessary for statistical downscaling [14]. In addition, statistical downscaling can result in a reduction in the physical coherence of climate simulations [122]. Approaches for statistical downscaling include regression-based methods [5], principal components analysis [23], [24], weather classification schemes, and weather generators [14]. Recently, machine learning methods such as artificial neural networks [25], deep learning [26], and random forests [27] have been used for downscaling both temperature and precipitation vari-

ables. Downscaling is especially important for accurate representation of temperature in regions characterized by topographically varied terrain [10], [11].

High-resolution climate data can also be generated by applying statistical downscaling to RCM output [123]. While this combination of dynamical and statistical downscaling is complex, it is an effective workflow for generating high-resolution climate data simulations as it combines physical and statistical relationships [31], [124], [125].

Bias-correction is another post-processing procedure that can correct the mean, variance, and higher moments of climatological variables [8], [9]. Generally, bias-correction methods can be classified into four categories: 1) linear scaling [42], [43]; 2) nonlinear scaling [44]; 3) distribution mapping [45]; and 4) empirical (distribution-free) quantile mapping [46]–[48]. The techniques differ in their ability to correct higher-order moments of simulated climatological variables. For bias-correcting temperature variables, linear scaling and empirical quantile mapping (EQM) are often used [22], [49]–[52]. EQM, a sophisticated technique, can correct the mean, variance, and higher moments of temperature and precipitation variables [22], [53]. Linear scaling is a simple technique in which the difference between monthly mean observed and simulated data is added to simulated data. Despite its simplicity, it is effective for bias-correcting temperature variables [42], [54]. Most bias-correction methods assume stationarity of model errors over time [55], and sufficient observational data is necessary to derive robust transfer functions.

Gridded, observational climate products [e.g. Livneh, [76]; Daymet, [77]; and PRISM, [78]] are often used for bias-correction due to their extensive spatial and temporal coverage. However, the interpolation algorithms used to create gridded cli-

mate products can introduce bias [79] and additional uncertainty when used for bias-correcting climate model output [80]. In particular, [79] found that in the United States, gridded observational products (including Livneh, Daymet, and PRISM) generally exhibited a negative bias for maximum daily temperature and that biases were exacerbated in topographically complex regions. Similarly, [126] found that in the Northeastern US, PRISM data products [78] demonstrated a cold bias for mean monthly temperature that increased at higher elevations.

A valuable alternative to gridded observational data products are long-term, curated station data, such as data from the Global Historical Climate Network [74]. Station data represent direct climatological measurements and are available globally [82], [83]. The use of station data, rather than gridded observational products, removes uncertainty during bias-correction. Station data are often used to validate the accuracy of bias-corrected climate model output but can also be effectual for bias-correcting output from climate models. Methods that account for the spatial autocorrelation of climate variables can improve the accuracy of gridded products created from sparsely distributed station data [112]. For instance, [84] downscaled monthly temperature and precipitation simulations from an RCM to climate stations and bias-corrected the simulated climate variables with station data, resulting in appreciable improvement in the accuracy of a hydrological model. [85] showed that incorporating station data in a geostatistical downscaling and bias-correction approach resulted in full-coverage, high-resolution monthly temperature and precipitation data that better captured the complex topographical features of their study area. Recently, [127] constructed 1km gridded datasets of monthly temperature over a region in China using a sophisticated geostatistical model, resulting in reduced uncertainty in the resulting

datasets.

Despite the advantages of station data, its use in constructing full-coverage, bias-corrected, downscaled climate data, especially at high spatial and temporal resolutions, is limited. The density and spatial distribution of climate stations are often irregular, especially in mountainous and high-elevation regions [78]. Another challenge is that for constructing full-coverage climate datasets, it is not sufficient to bias-correct only at station locations, as bias-correction must be applied at locations where stations are not present. There is a need for methods in which station data is leveraged to create full coverage, high-resolution bias-corrected climate data.

In this study, we leverage station data to develop and compare the performance of six downscaling and bias-correction methods for constructing high-resolution (1km), daily gridded temperature climate products. All of the six methods are specifically developed to address the challenge of creating full-coverage climate products using only station data. The methods incorporate well-established interpolation and bias-correction techniques, but the workflows of the methods are novel and unique. We apply the methods to daily RCM simulations of 2-meter maximum air temperature (TMAX) over a region in the northeastern United States. The relationship between elevation and temperature (lapse rate) is important to incorporate during downscaling, so we include fine-scale elevation during the downscaling process. However, in doing so, the adjustment of temperature due to elevation is difficult to disentangle with the adjustment due to bias-correction. For this reason, all six methods were implemented with and without the incorporation of fine-scale elevation. We validate the methods using two calibration time periods and apply a spatial cross-validation prior to calculating performance metrics to ensure that the ability of the methods to

bias-correct in a spatially coherent manner is accounted for.

This paper aims to address the following questions:

1. How do the different bias-correction and interpolation techniques used in the six methods affect performance, as measured by the root-mean square error (RMSE) and Perkins skill score (PSS)?

2. Does performance among methods vary by month, and is performance among methods improved when elevation lapse rates are used during downscaling?

3. Is any one method particularly well-suited for high-resolution downscaling and bias-correction with respect to both RMSE and PSS?

The article is organized as follows: in section 2, we describe the study area, station and RCM data, and downscaling and bias-correction methods. In section 2, we also provide specific justifications for each of the six methods and describe validation of the methods. In section 3, we present our results, and in section 4 we discuss our results and provide conclusions and recommendations.

2.3 Methods

2.3.1 Study area and data

The study area, the Lake Champlain Basin, consists of parts of Vermont, New Hampshire, eastern New York and southern Quebec, Canada (Figure 2.1). The region is topographically varied; the Green Mountains, Adirondack Mountains, and White Mountains span portions of Vermont, New York, and New Hampshire, respectively

[31]. Elevations in the study area range from 30 to 1500 m above mean sea level (MSL).

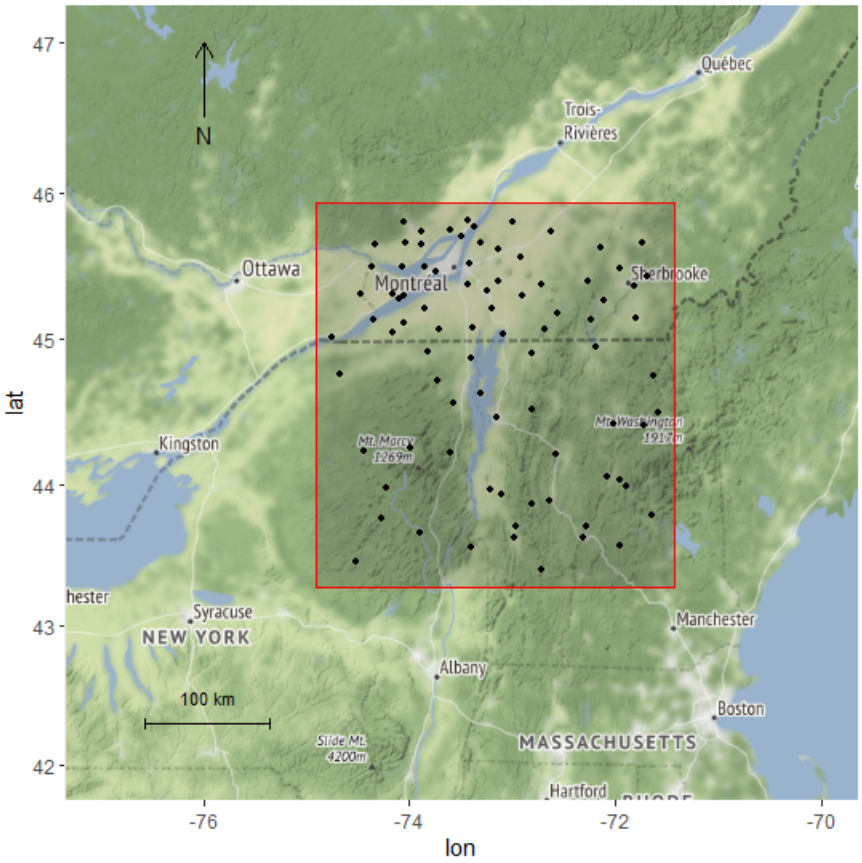

Figure 2.1: GHCND stations (n = 78) (black points) within the study area (red box).

Daily historical TMAX simulations over 1980-2014 were generated by the Advanced Weather and Research Forecasting model (WRF) version 3.9.1 [128]. WRF is widely used as both a regional climate model and numerical weather prediction system [128]. Initial and lateral boundary conditions were obtained from ERA-Interim, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA-interim has an approximate spatial resolution of 80 km [129] and was downscaled to 4 km using three one-way nests (36 km, 12 km, 4km) [130]. Only output

from the inner, 4km resolution domain was used in this study. Specific physics settings for WRF are shown in Supplementary Material. A total of 4387 WRF grid cells covered the study area.

Historical daily weather station data was obtained from the Global Historical Climate Network (GHCND) (https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND). GHCND data records are adjusted to account for changes in instrumentation and other anomalies [74], [82]. We retained only those stations with at least 70% complete records over the historical time period 1980-2014 (73 stations). In this study, WRF simulations were downscaled to a 1km grid; elevation estimates at each 1km grid cell were derived from a 30m digital elevation model (DEM) [131]. Elevation values were interpolated to the 1km grid using inverse distance weighting (IDW). The 1-km resolution was chosen based on spatial resolution requirements for local climate impact assessments [31], [132].

2.3.2 Description of downscaling and bias-correction methods

The six downscaling and bias-correction methods described in this paper can be divided into two groups: those that employ empirical quantile mapping (EQM) for bias-correction and those that employ linear transfer regression functions for bias-correction. Within the two groups, methods differ mainly with respect to interpolation techniques (IDW, kriging) and procedures to transfer bias-correction to locations void of stations.

Elevation has a major effect on climatological variables such as maximum temper-

ature [31], [133]. Therefore, during downscaling, it is important to account for lapse rates, especially in topographically rich regions, such as the Lake Champlain Basin [31]. However, we found that when elevation was incorporated (using lapse rates) during downscaling, it became difficult to disentangle the effects of downscaling with those of bias-correction. Therefore, all methods were implemented with and without the use of lapse rates or elevation covariate (depending on the interpolation method). When elevation was not accounted for, neither lapse rates nor the inclusion of an elevation covariate were included during interpolation of WRF data. In this study, we will regard steps involving the interpolation of WRF to GHCND station locations or the fine-scale grid as downscaling.

Empirical quantile mapping-based methods: EQM_krig, EQM_IDW, and EQM_grid

One way station data can be leveraged for bias-correcting WRF simulations in locations where stations are not present is to 1) interpolate WRF simulations to station locations, 2) bias-correct interpolated WRF simulations at station locations using empirical quantile mapping (EQM), and 3) interpolate bias-corrected WRF simulations at station locations to the fine-scale grid. This general workflow is implemented in EQM_krig and EQM_IDW (Figure 2.2; Table 2.1). As the suffixes suggest, the interpolation methods for EQM_krig and EQM_IDW were kriging and IDW, respectively. Both kriging, a geostatistical procedure, and IDW, a deterministic one, are common interpolation methods for downscaling [85], [134]–[137].

Table 2.1: Summary of six bias-correction and downscaling methods.

Method	Bias-correction technique	Details
EQM_krig	empirical quantile mapping	Bias-correction occurs at GHCND station locations. Bias-corrected WRF is then interpolated to the fine-scale grid using kriging.
EQM_IDW	empirical quantile mapping	Bias-correction occurs at GHCND station locations. Bias-corrected WRF is then interpolated to the fine-scale grid using IDW.
EQM_grid	empirical quantile mapping	GHCND station and WRF data are interpolated to the fine-scale grid. Bias-correction occurs at the fine-scale grid.
LTQM_grid_V	linear transfer function: quantile mapping using rank-ordered regression	Linear transfer functions are constructed at GHCND station locations, and estimated parameters are kriged to the fine-scale grid. Kriged parameters are used for bias-correction at the fine-scale grid.
LTQM_grid_C	linear transfer function: quantile mapping using rank ordered regression	Linear transfer functions are constructed at GHCND station locations, and estimated parameters are kriged to the fine-scale grid. The medians of kriged parameters over the fine-scale grid is calculated and used for bias-correction (parameters are constant over the fine-scale grid)
LT_grid	linear transfer function: temporally-ordered regression	Linear transfer functions are constructed at GHCND station locations, and estimated parameters are kriged to the fine-scale grid. Kriged parameters are used for bias-correction at the fine-scale grid.

For both methods EQM_krig and EQM_IDW, daily WRF simulations were first interpolated to GHCND station locations. For EQM_IDW, interpolation was completed using IDW with and without topographic downscaling [31]. IDW with topographic downscaling combines IDW with lapse rates to adjust for fine-scale elevation and has been applied to high-resolution downscaling [31] (full details on topographic downscaling and IDW are given in Supplementary Material). For brevity, we will refer to IDW with topographic downscaling as topographic downscaling. Two parameters, the power, p, and number of nearest neighbor observations used in averaging, n, control the smoothness of IDW interpolation. Higher values of p and n result in

progressively smoother interpolated surfaces. Based on results from [31], who used a similar study area and data, as well as our own assessment, we chose values of 2 and 9 for p and n, respectively. Elevational lapse rates were calculated using historical GHCND TMAX data within the study region following methods in [31].

For EQM_krig, WRF simulations were interpolated with kriging. To account for fine-scale elevation, elevation (either at station locations or at the fine-scale grid) was included as a covariate in universal kriging models. In the case when fine-scale elevation was not accounted for, ordinary kriging was used. IDW and kriging were implemented with the `gstat` package [138] in R [139]. The prediction surface resulting from kriging depends on the location of observational data as well as the strength of spatial dependence among the data, which can be assessed with a variogram. Based on inspection of empirical variograms of daily WRF TMAX data, all kriging models were fit with the exponential covariance function. The effective range, partial sill, and nugget were set to 150km, 15, and 0.2, respectively (full kriging details are described in Supplementary Material). We compared the two interpolation techniques, kriging and IDW, because we wanted to determine whether a geostatistical (kriging) or deterministic (IDW) interpolation technique would significantly influence performance. Kriging methods often work better for interpolating sparsely distributed data [140], [141], such as the GHCND station data, but IDW is simple, computationally efficient, and generally better suited for interpolated densely gridded data [136]. However, any interpolation method that incorporates relationships between temperature data and topographic features such as elevation is likely to produce more realistic predictions of climate variables, especially in regions of varying topography [137].

Once WRF simulations were interpolated to GHCND station locations for all

days in the historical time period, WRF interpolations were bias-corrected at each GHCND station location using EQM (2.1). The EQM transfer function is expressed by the empirical cumulative distribution function (ecdf) and its inverse (ecdf^{-1}).

$$X_{corr,t} = \text{ecdf}^{-1}_{obs,m}(\text{ecdf}_{raw,m}(X_{raw,t})). \tag{2.1}$$

In (2.1), $X_{corr,t}$ is the corrected WRF TMAX value on day t, $\text{ecdf}^{-1}_{obs,m}$ is the inverse ecdf of GHCND station data for month m, and $\text{ecdf}_{raw,m}$ is the ecdf of interpolated WRF TMAX simulations at a GHCND station location for month m, and $X_{raw,t}$ is the interpolated, uncorrected WRF TMAX at a GHCND station location on day t. Thus, daily WRF simulations in a specific month were corrected with the corresponding monthly EQM transfer function. For example, a WRF simulated value of TMAX in January would be corrected with the EQM transfer function for January. EQM was implemented with the `qmap` package [102] in R. Finally, bias-corrected WRF simulations at GHCND station locations were interpolated to the fine-scale grid with the same method used to interpolate coarse-grid WRF simulations to GHCND station locations.

Despite the simplicity of EQM_krig and EQM_IDW, much of the original WRF data is not used, as ultimately only bias-corrected WRF simulations at GHCND station locations are interpolated to the fine-scale grid. Another approach to transferring information from stations to other locations for bias-correcting WRF data is to 1) interpolate both GHCND station and WRF data to the fine-scale grid and 2) bias-correct WRF interpolated data with interpolated station data on a grid-cell by grid-cell basis using EQM. The method EQM_grid (Figure 2.2; Table 2.1) has advantages over EQM_krig and EQM_IDW, since it preserves more spatial information

from WRF data (i.e. the *grid* suffix indicates that bias correction is applied at the fine-scale grid, rather than station level).

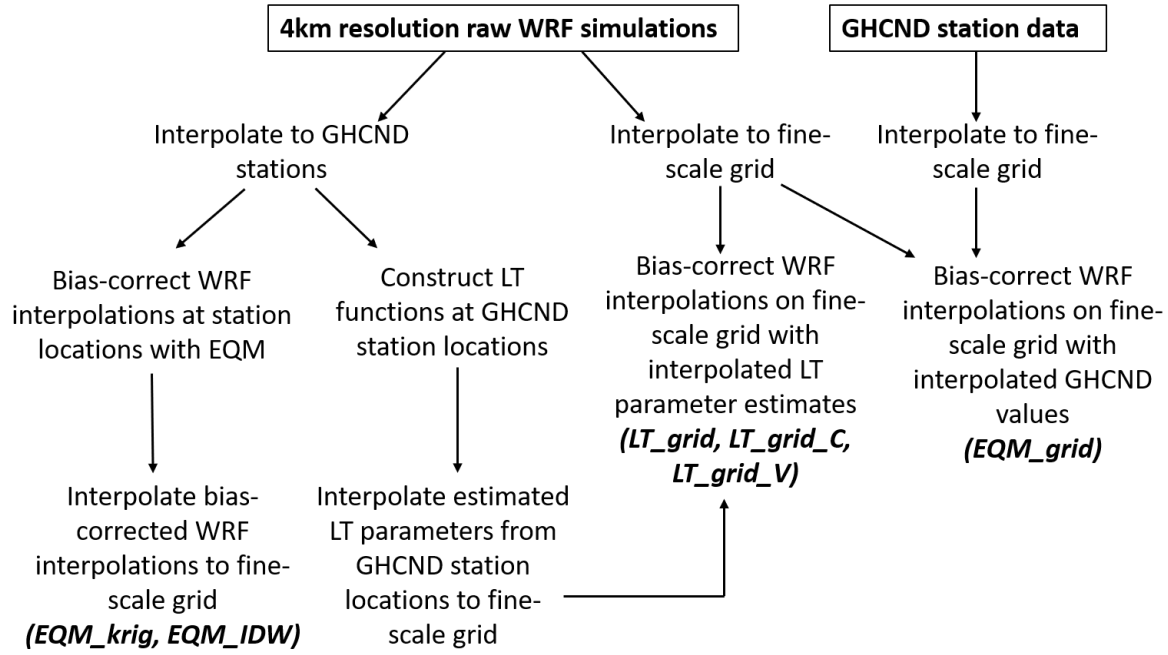

Figure 2.2: Workflows for the six bias-correction and downscaling methods described in this study. In EQM_IDW, EQM_krig, and EQM_grid, bias-correction was done with empirical quantile mapping (EQM). EQM_grid differs with respect to EQM_krig and EQM_IDW in that bias-correction was done at the grid *rather than station level. In LTQM_grid_V and LTQM_grid_C, linear transfer (LT) functions were constructed using rank-ordered WRF and GHCND station data. In LTQM_grid_V, interpolated LT parameters were used for bias-correction at the fine-scale grid level, so LT parameters were allowed to* vary *spatially (V = vary). In LTQM_grid_C, the median values of interpolated LT parameters at the fine-scale grid level were calculated and subsequently used for bias-correction, so LT parameters were* constant *over the fine-scale grid (C = constant). Interpolated parameters were also allowed to vary spatially over the fine-scale grid for method LT_grid, but LT functions were constructed using temporally-ordered, rather than rank-ordered, data.*

In method EQM_grid, WRF simulations and GHCND station data were interpolated to the fine-scale grid. WRF and GHCND data were interpolated with IDW and kriging, respectively. Kriging, rather than IDW, was used for GHCND station data, as it is generally better for interpolating sparsely distributed data [140]. The covari-

ance function and covariance parameters were identical to those used in EQM_krig. When elevation was accounted for, interpolation of WRF simulations was done via topographic downscaling. Interpolation of GHCND station data was done with universal kriging, which included an elevation covariate. Finally, after WRF simulations and GHCND station data were interpolated to the fine-scale grid, WRF interpolations were bias-corrected with kriged GHCND station data grid-cell by grid-cell using EQM (2.1).

Linear transfer function-based methods: quantile mapping and linear regression (LTQM_grid_C, LTQM_grid_V, LT_grid)

The linear transfer (LT) family of methods presents an alternative way to transfer information needed to bias-correct WRF simulations at any location on the fine-scale grid. In methods LT_grid, LTQM_grid_V, and LTQM_grid_C, bias-correction is done by applying LT functions derived from regression relationships between GHCND station data and WRF simulations (Figure 2.2; Table 2.1). In these methods, simple regression parameters (slopes and intercepts) are estimated at GHCND station locations and interpolated to locations on the fine-scale grid where bias-correction is to be performed. Thus, LT methods provide an alternative to the EQM methods (EQM_grid, EQM_krig, and EQM_IDW), as estimated parameters, rather than either bias-corrected data (EQM_krig, EQM_IDW) or GHCND station data (EQM_grid) are interpolated to the fine-scale grid and subsequently used to bias-correct WRF data at the grid level.

The main difference between both LTQM methods (LTQM_grid_V and LTQM_grid_C) and LT_grid is the ordering of the data used to construct the simple regres-

sions, which ultimately impacts the type of correction applied to WRF simulations. Two types of data ordering were considered: 1) temporally-ordered (calendar order) (LT_grid) and 2) rank-ordered (sorted from least to greatest) (LTQM_grid_V and LTQM_grid_C). In both cases 1) and 2), GHCND station data was expressed as a linear function of WRF data, and regression parameters (slope and intercept) were estimated via ordinary least squares (OLS). In the context of this study, resulting regression equations are applied to raw WRF data to complete the bias-correction. The intercept adjusts the mean, while the slope scales the variance. Thus, since the regression equation is linear in form, the transfer function is *linear*.

If OLS assumptions are met, then by definition, OLS estimates are BLUE (best linear unbiased estimators) [142], and the regression line is the only such line that minimizes the mean square error. It follows that for case 1), in which WRF and GHCND station data are temporally-ordered (LT_grid), the LT function is guaranteed to improve daily discrepancies between WRF and GHCND station data (RMSE). However, the approach is not guaranteed to improve distributional discrepancies to the same degree. For case 2), in which data are rank-ordered (LTQM_grid_V and LTQM_grid_C), the LT function acts as a simple type of quantile mapping and will thus improve distributional similarity (and PSS) between WRF and GHCND station data. However, RMSE is not guaranteed to improve. Since both LTQM_grid_C and LTQM_grid_V bias-correct via a simple quantile-mapping technique, the "QM" in LTQM_grid_V and LTQM_grid_C refers to "*Q*uantile *M*apping". The subtle difference between LTQM_grid_V and LTQM_grid_C will be discussed later.

While using rank-ordered data results in a simple form of quantile mapping, the quantile map between WRF and GHCND station data is modeled with a *linear* re-

gression line. EQM is more flexible, as first, quantiles of observed and station data are estimated, and then the quantile map is approximated via linear or spline interpolation [102]. It is important to note that if OLS assumptions (linearity, homoscedasticity of residual errors, and independence of observations) are not met, the OLS estimates are no longer BLUE.

The first step for methods LT_grid, LTQM_grid_V, and LTQM_grid_C was identical: daily WRF simulations were interpolated to the fine-scale grid using IDW (or topographic downscaling). Daily WRF simulations were also interpolated to GHCND station locations, where LT functions were formulated (2.2).

For all three methods (LTQM_grid_C, LTQM_grid_V, LT_grid), LT functions were constructed by regressing large-scale predictor variables (WRF data) on small-scale predictands (GHCND station data) at each GHCND station location. Separate LT functions were constructed for each month. The estimated regression parameters at each GHCND station location (slope and intercept coefficients) were kriged to the fine-scale grid, and interpolated WRF simulations on the fine-scale grid were bias-corrected with the corresponding kriged regression parameters grid-cell by grid-cell. Therefore, the term "grid" in all three methods refers to bias-correction taking place at the fine-scale grid center points, rather than station locations.

The LT function for LT_grid was a simple linear regression in which WRF interpolations at GHCND stations were predictor variables, and GHCND station data were the predictands (2.2). Data were sorted in temporal order. Twelve LT functions (one for each month) were constructed at each GHCND station location (2.2).

$$TMAX_{station,i,m} = \beta_{0,i,m} + \beta_{1,i,m} \times WRF_{IDW,i,m} \tag{2.2}$$

In (2.2), $TMAX_{station,i,m}$ is daily TMAX for GHCND station location i in month m, $\beta_{0,i,m}$ is the intercept for GHCND station location i in month m, $\beta_{1,i,m}$ is the slope for GHCND station location i in month m, and $WRF_{IDW,i,m}$ represents daily interpolated WRF values at GHCND station location i in month m. Monthly parameter estimates of slopes and intercepts at each GHCND station location were kriged to the fine-scale grid with ordinary Bayesian kriging.

Empirical variograms of estimated monthly slope and intercept parameters showed some degree of spatial autocorrelation, although the association was stronger in cold-season compared to warm-season months. The exponential covariance function was used for all Bayesian kriging models. Based on the inspection of empirical variograms, we used non-informative prior distributions for the intercept (β_0), partial sill (σ^2), and nugget (τ^2), and an informative prior for the effective range (ϕ).

$$\begin{aligned}
\beta_0 &\sim N(0, 100) \\
\phi &\sim Unif(\frac{3}{D_{max}}, \frac{3}{10}) \\
\sigma^2 &\sim InvGamma(2, 2) \\
\tau^2 &\sim InvGamma(2, 0.02).
\end{aligned}$$

D_{max} was the maximum distance between any two GHCND station locations (full details on Bayesian modeling are described in Supplementary Material, section 1). Bayesian kriging is preferable to non-Bayesian kriging when data is sparse, and there is some degree of uncertainty surrounding estimates of covariance function parameters [143]. Finally, interpolated WRF simulations on the fine-scale grid were bias-corrected grid-cell by grid-cell, using the corresponding kriged slope and intercept parameter

estimates (2.3):

$$TMAX^{*}_{i,m} = \tilde{\beta}_{0,i,m} + \tilde{\beta}_{1,i,m} \times WRF_{1km-interp,i,m}. \tag{2.3}$$

In (2.3), $TMAX^{*}_{i,m}$ is the bias-corrected, fine-scale WRF value for grid cell i in month m, $\tilde{\beta}_{0,i,m}$ is the kriged prediction for the intercept of grid cell i in month m, $\tilde{\beta}_{1,i,m}$ is the kriged slope parameter estimate at fine-scale grid cell i in month m, and $WRF_{1km-interp,i,m}$ is the interpolated WRF value at the center of fine-scale grid cell i in month m.

For methods LTQM_grid_V and LTQM_grid_C, LT functions were constructed using rank-ordered WRF and GHCND station data. In these LT functions, the nearest WRF grid-cell values to GHCND station locations were predictor variables and GHCND station data were predictands, similar to the approach of [144], who applied rank-ordered regression to bias-correct temperature and precipitation simulations. [144] found that modeling empirical quantiles of RCM and observed mean temperature data with a simple linear regression worked well if the quantile map between simulated and observed data was linear in form. Twelve LT functions were constructed at each GHCND station location (2.4).

$$TMAX_{i,m} = \beta_{0,i,m} + \beta_{1,i,m} \times WRF_{NN_{i,m}}. \tag{2.4}$$

In (2.4), $TMAX_{i,m}$ is daily TMAX at GHCND station location i in month k, $\beta_{0,i,m}$ is the intercept for GHCND station location i in month m , $WRF_{NN_{i,m}}$ are the one-nearest-neighbor grid cell WRF simulations relative to GHCND station location i in month m, and $\beta_{1,i,m}$ is the coefficient for station location i in month m.

There was one important difference between methods LTQM_grid_V and LTQM_grid_C. In method LTQM_grid_V, monthly estimates of intercepts and slopes were kriged to the fine-scale grid with ordinary Bayesian kriging using the same priors as in LT_grid. Then, the kriged slopes and intercepts were used to bias-correct interpolated WRF data on the fine-scale grid (2.3). In method LTQM_grid_C, however, the monthly medians of kriged slopes and intercepts over the fine-scale grid were used to bias-correct interpolated WRF data (2.3). In LTQM_grid_V, the kriged slopes and intercepts used to bias-correct WRF interpolations varied over the fine-scale grid (V for *vary*). In contrast to LTQM_grid_V, spatially constant (C for *constant*) slope and intercept values were used for bias-correction in LTQM_grid_C. We implemented variations in which estimated slopes and intercepts varied spatially (LTQM_grid_V) and in which they were spatially constant (LTQM_grid_C), because monthly kriged surfaces of estimated slopes and intercepts over the fine-scale grid were not always spatially smooth. A rougher parameter surface could potentially result spatially incoherent corrections in some locations. Using constant monthly medians of kriged slope and intercept estimates alleviates issues related to a rough kriging surface but sacrifices flexibility in that any spatial dependence among is no longer accounted for.

2.3.3 Performance measures and validation

Performance measures

Bias-corrected WRF simulations should exhibit day-to-day, as well as distributional, correspondence to GHCND station data. Thus, we chose root-mean-square prediction error (RMSE) and Perkins skill score (PSS) [145] as performance metrics which

1) quantify daily discrepancies and 2) distributional similarity between WRF and GHCND station data, respectively. PSS ranges between 0 and 1, where 1 indicates a perfect distributional overlap between simulated and observed data, and 0 indicates no distributional overlap [145]. PSS is calculated by summing minimum densities of overlapping bins of discrete histograms of simulated and observed data. PSS is not highly influenced by outliers, but it is sensitive to bin size [145]. However, larger daily discrepancies between simulated and observed data have a comparatively larger influence on RMSE, due to the squared term in its calculation. Both PSS and RMSE metrics are widely in the climate literature for validation [31], [34], [85], [127], [145]–[147].

To fairly assess the ability of the six methods to bias-correct WRF simulations at locations where stations are not present, we implemented a five-fold spatial cross-validation prior to calculating performance metrics. In each fold, 1) bias-correction was based on approximately 70% of GHCND stations and 2) bias-correction was applied to WRF interpolations at the remaining 30% of GHCND station locations.

Because all of the six methods had slightly different workflows, the five-fold spatial cross-validation was adjusted for each method to ensure that results were comparable.

For EQM_krig and EQM_IDW methods, the cross-validation was performed as follows for each of the $i = 1...k$, $k = 5$, folds: for fold i, bias-corrected WRF interpolations at GHCND station locations in fold $k \neq i$ were used as training data and were interpolated (via kriging or IDW) to GHCND station locations in fold i.

For EQM_grid, TMAX values at GHCND station locations in the $k \neq i$ fold were used as training data and were interpolated using ordinary kriging to station locations in fold i. Then, interpolated WRF data at GHCND station locations in the

i^{th} were bias-corrected using kriged GHCND station values. This was repeated for the $i = 1...k$, $k = 5$ folds.

For LT_grid, LTQM_grid_V, and LTQM_grid_C methods, LT functions (2.2 and 2.4) were constructed at GHCND station locations in folds $k \neq i$; Bayesian kriging was used to krig estimated LT parameters (slopes and intercepts) to GHCND station locations in fold i. Interpolated WRF values at GHCND station locations in the i^{th} fold were bias-corrected with kriged estimated LT parameters. This was repeated for the $i = 1...k$, $k = 5$ folds.

Like all cross-validation approaches, GHCND stations in each of the five folds were randomly selected prior to spatial cross-validation; thus, for each method, the stations in folds $k = 1 \ldots 5$ were the same to ensure that results would be comparable.

Spatially cross-validated, daily RMSE values were calculated by method and month using the following formula:

$$E_k(Y) = \sqrt{\frac{1}{n_k} \sum_{i \in k^{th} fold} (Y(s_j - \widehat{Y(s_j)^2})}$$

$$RMSE = \frac{1}{K} \sum_{k=1}^{K} E_k(Y),$$

where $Y(s_j)$ is the TMAX value at GHCND station s_j, $\widehat{Y(s_j)^2}$ is the bias-corrected WRF TMAX value at GHCND station location s_j, n_k is the number of observations in fold k and $K = 5$.

To calculate PSS, discrete probability density functions (PDFs) were constructed for bias-corrected WRF and GHCND station data using bin widths of 0.5°C as recommended by [145]. Spatially cross-validated PSS was calculated by method and

month using the following formula:

$$E_k(PSS) = \sum_{i}^{b_k} min(Z_i, Z_i^*)$$
$$PSS_m = \frac{1}{K} \sum_{k=1}^{K} E_k(PSS),$$

where Z_i is the normalized density of the PDF of GHCND station data in bin i, Z_i^* is the normalized density of the PDF of bias-corrected WRF data in bin i, and b_k is the number of bins used to construct the PDFs of GHCND station and bias-corrected WRF data in fold k, and $K = 5$.

2.3.4 Validation

We validated the six methods using two calibration time periods. Bias-correction was applied to 1) 1980-2014 WRF simulations using 1980-2014 GHCND station data and 2) 1990-2014 WRF simulations using 1980-1989 GHCND station data. The former approach helps evaluate performance of methods for processing historical simulations, while the latter approach assesses potential performance of methods for processing future projections. For clarity, we name these cases by referring to the subset of GHCND station data that are used for bias-correction (e.g. "1980-2014" and "1980-1989").

2.3.5 Analysis of performance metrics

Performance metrics of the six methods were analyzed with two linear analysis of variance (ANOVA) models (one for RMSE and one for PSS). Our observations show that raw WRF interpolations at GHCND station locations exhibit a distinct cold bias in winter and early spring compared to summer and early fall months (Figure 2.3), so we controlled for monthly variation in the two ANOVA models. We also controlled for whether or not elevation was accounted for in downscaling to help disentangle the effects of downscaling with those of bias-correction. Finally, we controlled for the calibration time period (1980-1989 or 1980-2014) used to bias-correct WRF simulations. We used linear ANOVA models to evaluate performance among methods, as they are easy to interpret and provide information on how PSS and RMSE differ among methods while controlling for variables. With the incorporation of interaction effects, linear ANOVA models can also help expand knowledge of more complex relationships among performance metrics, the six methods, and controlling variables (described below).

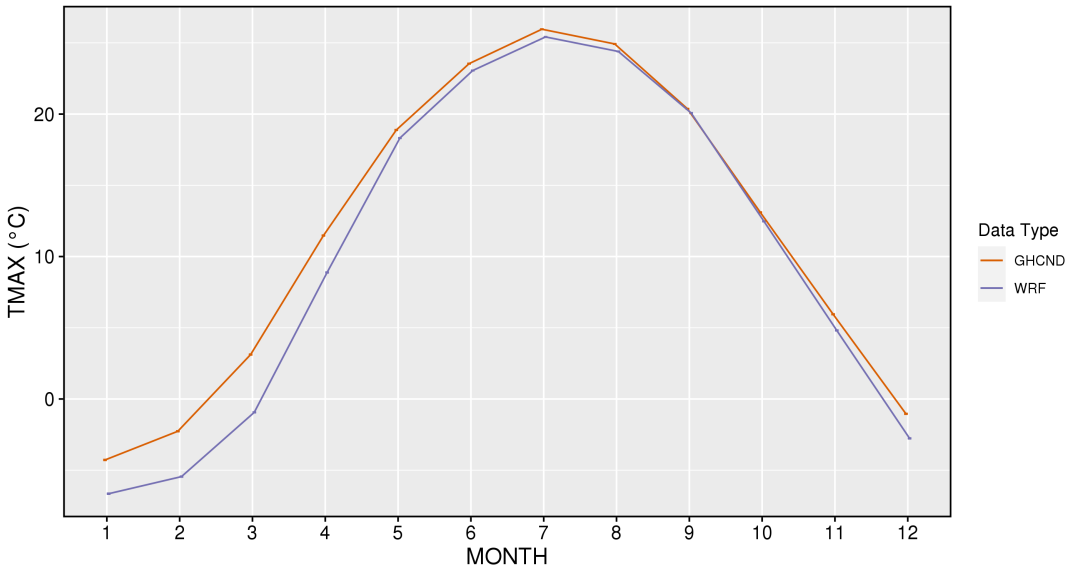

Figure 2.3: Monthly average TMAX (°C) of WRF interpolations at GHCND locations and GHCND station data from 1980-2014 showing a distinct cold bias in the WRF simulations for the winter and early spring months (month 1-5 and 11-12).

Prior to ANOVA model fitting, spatially cross-validated RMSE and PSS were averaged over the six methods and months. Full models for PSS and RMSE were fit with the following four fixed effects:

- *Method*: identifier for the downscaling and bias-correction method (EQM_krig, EQM_IDW, EQM_grid, LT_grid, LTQM_grid_V, and LTQM_grid_C)
- *Month*: month of the year (1-12)
- *Elevation*: binary variable to denote whether the effect of elevation was included with the use of elevational lapse rates ("YES") or not ("NO")
- *Bias_correction_years*: binary variable to denote if 1990-2014 WRF simulations were bias corrected with 1980-1989 GHCND station calibration dataset

("1980-1989") or whether 1980-2014 WRF simulations were bias-corrected with the 1980-2014 GHCND station calibration dataset ("1980-2014").

In addition, the initial full model fits included reasonable and informed two- and three-way interactions: *Month* × *Method*, *Elevation* × *Method* × *Bias_correction_years*, *Elevation* × *Method*, *Elevation* × *Bias_correction_years*, *Bias_correction_years* × *Method*, and *Method* × *Bias_correction_years* × *Elevation*. After full ANOVA models were fit, all variables with a p-value < 0.05 were eliminated, and both ANOVA models were fit again with remaining variables. After fitting final ANOVA models, pairwise comparisons, as well as estimated marginal means (necessary for interaction plots) were calculated with the R package `emmeans` [148]. Pairwise comparisons were carried out with the Bonferroni correction for multiple comparisons. We also calculated η^2 for all effects in the final models for PSS and RMSE. η^2 quantifies the proportion of variance associated with main effects and interactions in a linear model and is a useful indicator of effect size and strength of association in linear models [149], [150]. Values for η^2 range between 0 and 1, where higher values indicate greater variable importance. η^2 is calculated as the sum of squares of an independent variable ($SS_{between}$) divided by the total sum of squares (TSS) of the model:

$$\eta^2 = SS_{between}/TSS.$$

2.4 Results

2.4.1 Overall performance

Raw WRF interpolations at GHCND station locations exhibited a cold bias, which was most pronounced in months 11 and 12 and 1-4 (Figure 2.3). Generally, the average day-to-day correspondence between GHCND station data and bias-corrected WRF data, as measured by mean RMSE, varied little among methods, ranging between 3.1-3.5 (Figure 2.4 a). Distributional similarity between GHCND station data and bias-corrected WRF data, measured by PSS, ranged between 0.94-0.96 (Figure 2.4 b). All methods performed better than uncorrected WRF: RMSE of uncorrected WRF interpolations at GHCND station locations ranged between 3.6 and 3.9, while mean PSS ranged between 0.90 and 0.91 (Figures 2.4 a and b).

Performance metrics for all methods exhibited monthly variation: both mean monthly RMSE and PSS were worse in months 11, 12, and 1-4 compared to months 5-10 (Figures 2.5 a and b), although monthly variation was much more pronounced for RMSE than PSS. Overall, methods LT_grid and LTQM_grid_V performed best and worst, respectively, in terms of mean RMSE (Figure 2.4 a), while methods EQM_grid and LTQM_grid_V performed best and worst, respectively in terms of mean PSS (Figure 2.4 b).

Mean RMSE and PSS improved when bias-correction was based on 1980-2014 GHCND station data (and the correction was applied to 1980-2014 WRF data) compared to when bias-correction was based on 1980-1989 GHCND station data (and the correction was applied to 1990-2014 WRF simulations) (Figures 2.4 a and b, re-

spectively). Generally, when elevation was accounted for during downscaling (with lapse rates), mean RMSE decreased (Figure 2.4 a), but *Elevation* did not have an appreciable impact on mean PSS (Figure 2.4 b). There was no consistent relationship between low RMSE and high PSS. An example of a downscaled, bias-corrected data product for one particular day is shown in Figure 2.6 (only one example is shown, as downscaled, bias-corrected data for all methods were ocularly indistinguishable). Figure 2.6 clearly shows that the downscaled, bias-corrected data captures the fine-scale topographical variation of TMAX over the study region, and is a much more realistic compared to raw WRF.

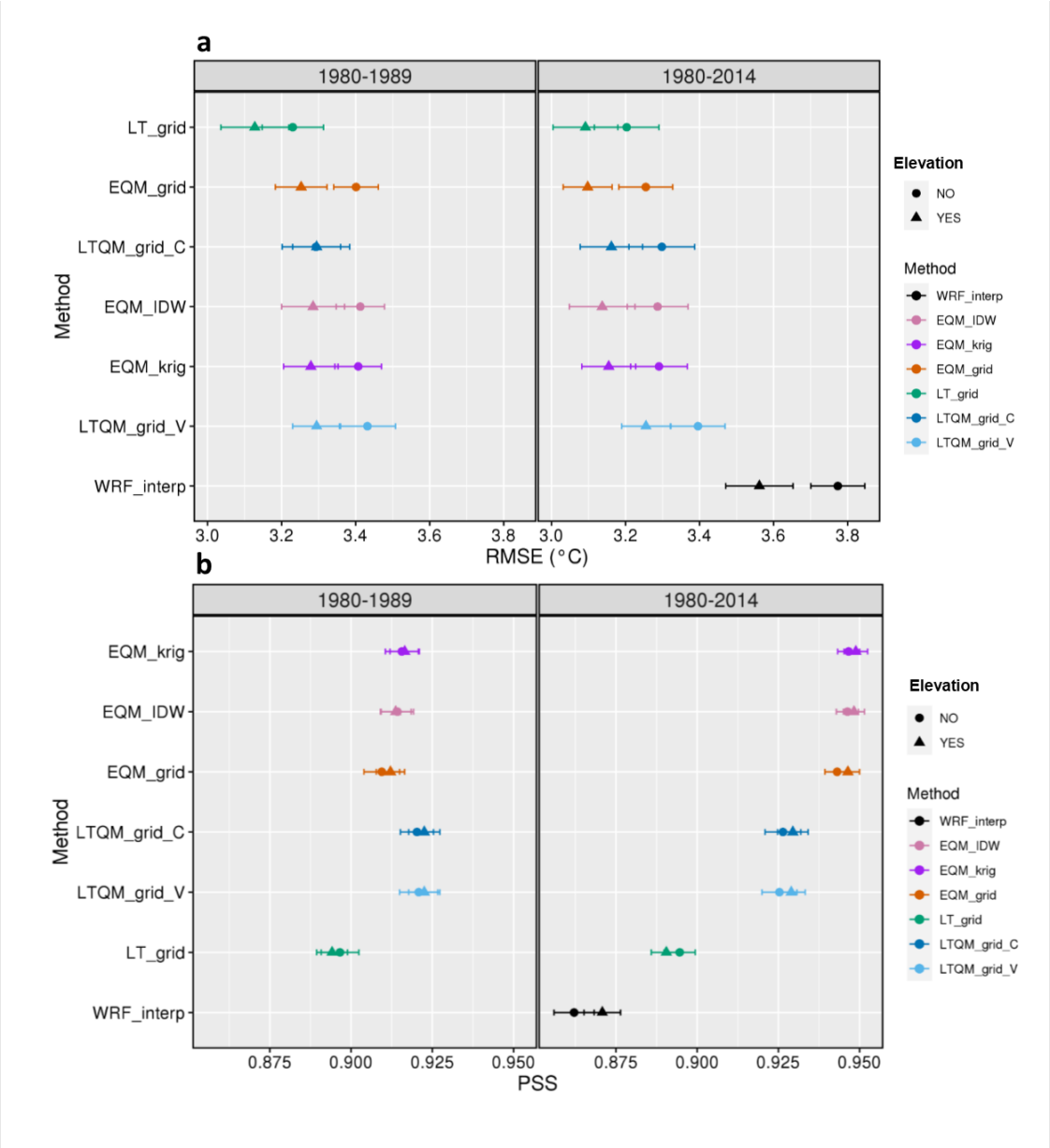

Figure 2.4: Mean RMSE (°C) (a) and PSS (b) by Method *and* Bias_correction_years, *where "1980-1989" and "1980-2014" refer to GHCND station datasets used to bias-correct 1990-2014 and 1980-2014 WRF simulations, respectively. Error bars represent standard errors over five spatial cross-validation folds. "WRF_interp" denotes the raw WRF simulations interpolated to station locations and are shown as a comparison to bias-corrected WRF data.*

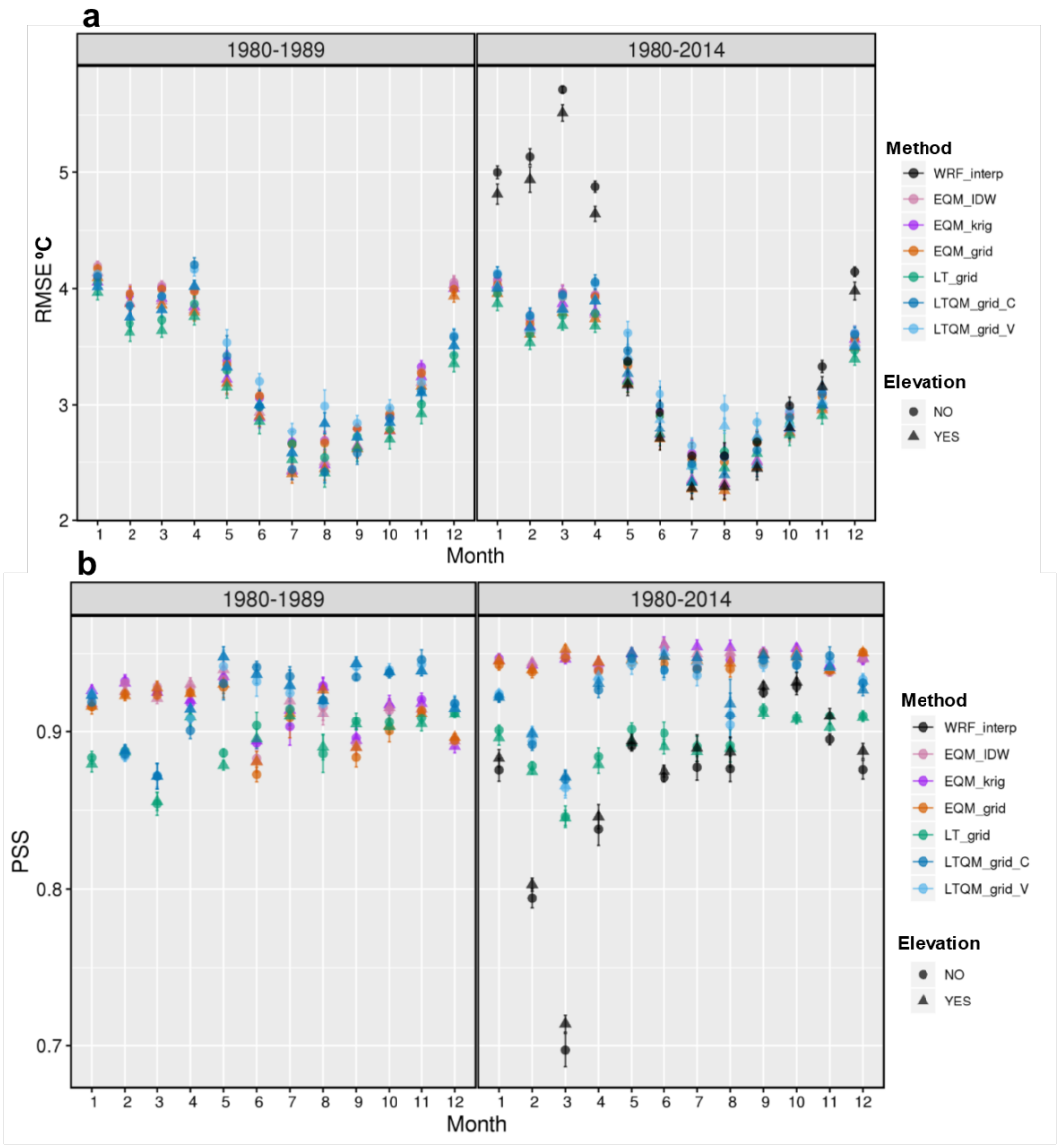

Figure 2.5: Mean RMSE (°C) (a) and PSS (b) by Method, Month *and* Bias_correction_-years, *where "1980-1989" and "1980-2014" refer to GHCND station datasets used to bias-correct 1990-2014 and 1980-2014 WRF simulations, respectively. Error bars represent standard errors over five spatial cross-validation folds. "WRF_interp" denotes raw WRF simulations interpolated to station locations and are shown to indicate relative improvement of all methods over raw WRF interpolated values.*

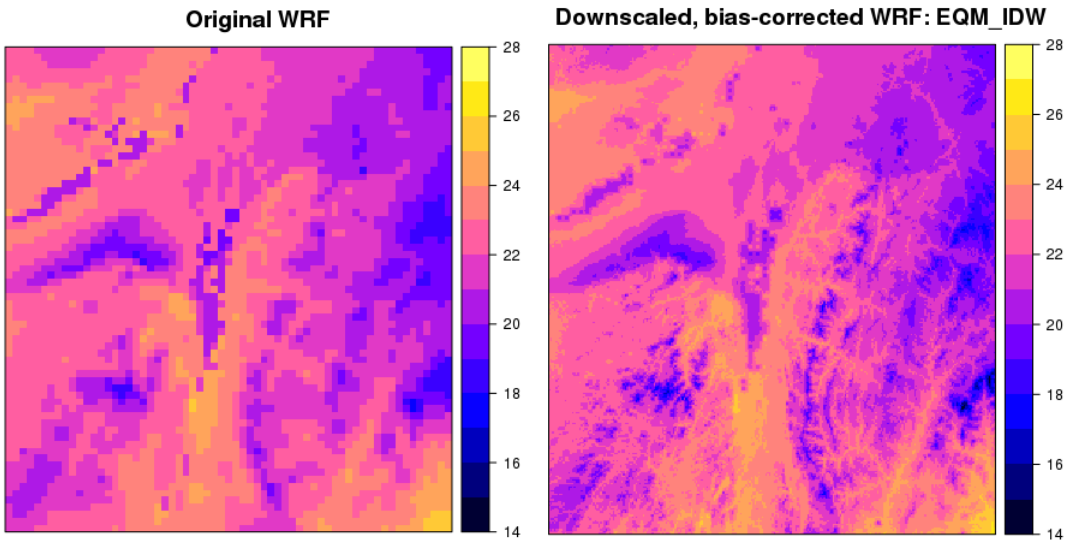

Figure 2.6: Original WRF simulations for TMAX (°C) and downscaled WRF TMAX (°C) using method EQM_IDW for August 5, 1982.

2.4.2 Statistical analysis of error metrics

The final ANOVA model for RMSE included the main effects *Month*, *Bias_correction_years*, *Elevation*, and *Method* as well as the interactions *Month* × *Method*, *Method* × *Bias_correction_years*, and *Method* × *Elevation*. P-values for all variables in the final RMSE ANOVA model were less than 10^{-5} (Table 2.4; See Appendix, Table 2.6 for the full ANOVA table). The final model for PSS included the main effects *Month*, *Method*, and *Bias_correction_years* and the interaction terms *Month* × *Method* and *Method* × *Bias_correction_years*. P-values for all variables in the final PSS ANOVA model were less than 10^{-5} (Table 2.5, see Appendix, Table 2.7 for the full model ANOVA). In contrast to results for RMSE, the effect of *Elevation* was not significant in the full model for PSS ($p = 0.857$; Appendix, Table 2.7).

Due to the significance of interaction as well as main effects, main effects are discussed in the context of interaction effects. Results for pairwise comparisons for interaction terms present in RMSE and PSS ANOVA models, as well as an alternative metric to PSS for quantifying distributional similarity, are shown in Supplementary Material.

Table 2.2: η^2 for the final RMSE ANOVA model

Predictor	η^2
Month	0.94
Month× Method	0.014
Elevation	0.013
Method	0.0092
Bias_correction_years	0.0063
Bias_correction_years × Method	0.0018
Elevation × Method	0.00069

Table 2.3: η^2 for the final PSS ANOVA model.

Predictor	η^2
Method	0.43
Month × Method	0.28
Bias_correction_years	0.14
Month	0.11
Bias_correction_years × Method	0.09

Month, *Method*, and *Month* × *Method*

RMSE η^2 for *Month* was 0.94, whereas η^2 for *Month* × *Method* and *Method* were 0.014 and 0.0092, respectively (Table 2.2). The large η^2 for *Month* indicates that *Month* was the most important variable in the model despite the statistical significance of the interaction *Month* × *Method*. Thus, RMSE varied substantially by month. Indeed, the monthly pattern of RMSE was consistent for all methods (Figure

2.7 a). The interaction plot for *Month* × *Method* shows that mean marginal RMSE of all methods was greater (3.2-4.2°C) in months 1-4, 11, and 12 compared to months 5-10 (2.5-3°C) (Figure 2.7 a). Overall, mean marginal RMSE of LT_grid was lower than that of all other methods and was significantly lower during months 2-6, 11, and 12.

PSS In contrast, for the PSS ANOVA model, the influence of *Method* ($\eta^2 = 0.43$) was greater than that of *Month*×*Method* ($\eta^2 = 0.28$) and *Month* ($\eta^2 = 0.11$) (Table 2.3). This means that PSS varied more among the six methods than among months (Figure 2.7 b). Mean marginal PSS for EQM_IDW, EQM_krig, and EQM_grid varied between 0.92 and 0.95, regardless of month (Figure 2.7 b). However, mean marginal PSS for LTQM_grid_C and LTQM_grid_V ranged between 0.88 and 0.90 in months 1-4 and then increased to between 0.94 and 0.96 in months 5-12 (Figure 2.7 b). Mean marginal PSS for LT_grid followed a similar pattern as LTQM_grid_V and LTQM_grid_C for months 1-4; however in months 5-10, mean marginal PSS for LT_grid was significantly lower than that of all other methods for every month, ranging between 0.90 and 0.91.

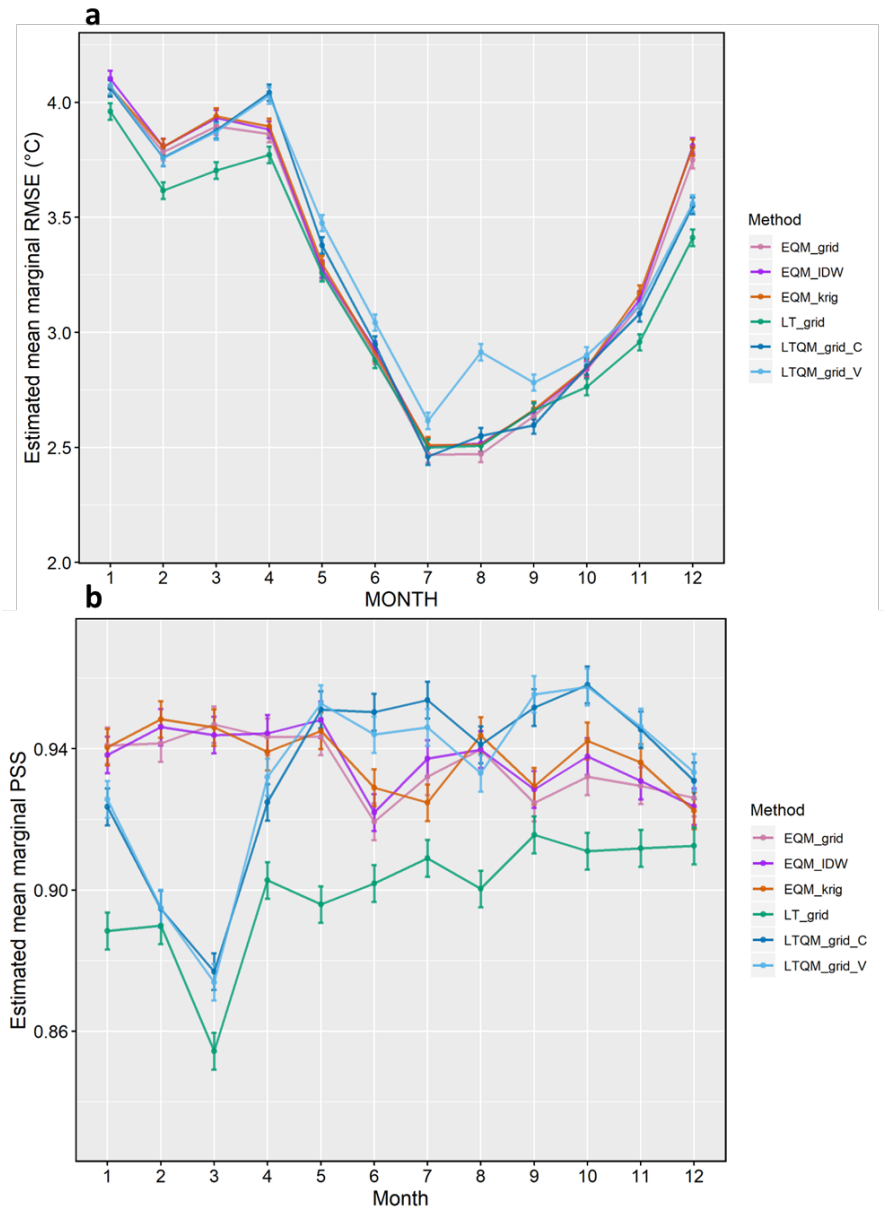

Figure 2.7: Interaction plots for Method × Month *showing estimated mean marginal RMSE (a) and PSS (b) for each month. Error bars represent 95% confidence intervals.*

Bias_correction_years and *Bias_correction_years* × *Method*

RMSE In the RMSE ANOVA model, η^2 values for *Bias_correction_years* × *Method* and *Bias_correction_years* (η^2 = 0.0018 and 0.0063, respectively) indicate that the main effect of *Bias_correction_years* was slightly more important than *Bias_correction_years* × *Method*. Overall, mean marginal RMSE slightly improved when bias-correction was based on 1980-2014 GHCND station data (3.15-3.25°C) compared to the 1980-1989 GHCND subset (3.18 and 3.57°C), although there were slight differences among methods (Figure 2.8 a). The improvement was more pronounced for the EQM-based methods than for the LT-based methods. Mean marginal RMSE of LT_grid was significantly lower (3.18°C) than that of all other methods (3.3-3.56°C) when bias-correction was based on the 1980-1989 GHCND dataset (Figure 2.8 a). When bias-correction was done with the 1980-2014 GHCND dataset, mean marginal RMSE of LT_grid was lowest overall, but it did not differ significantly from that of EQM_grid. Mean marginal RMSE of LTQM_grid_V was significantly greater than that of all other methods. Finally, it is important to note that η^2 values for *Bias_correction_years* and *Bias_correction_years* × *Method* were much smaller compared to that of η^2 of *Month*, which means that *Month* was relatively more important than *Bias_correction_years* and *Bias_correction_years* × *Method*.

PSS Mean PSS generally increased when the 1980-2014, as compared to the 1980-1989 GHCND dataset, was used for bias-correction, but the amount of increase varied among methods. Use of the 1980-2014 GHCND dataset for bias-correction set resulted in a consistent improvement in PSS for all EQM-based methods. The interaction *Bias_correction_years* × *Method* was particularly evident for methods

LTQM_grid_C and LTQM_grid_V. Mean marginal PSS for both LTQM_grid_C and LTQM_grid_V were significantly higher than that of all other methods when the 1980-1989 GHCND dataset was used for bias-correction (Figure 2.8 b). However, when the 1980-2014 GHCND dataset was used for bias-correction, mean marginal PSS of LTQM_grid_C and LTQM_grid_V fell significantly below that of EQM-based methods. In contrast to results for RMSE, LT_grid performed worst overall. Mean marginal PSS of LT_grid was significantly lower than that of all other methods, regardless of which GHCND dataset was used for bias-correction (Figure 2.8 b). The main effect *Bias_correction_years* and interaction *Bias_correction_years*×*Method* (η^2 = 0.09 and 0.14, respectively) were comparatively less influential in the model than the effects *Method* and *Month* × *Method* (η^2 = 0.43 and 0.28, respectively) (Table 2.3).

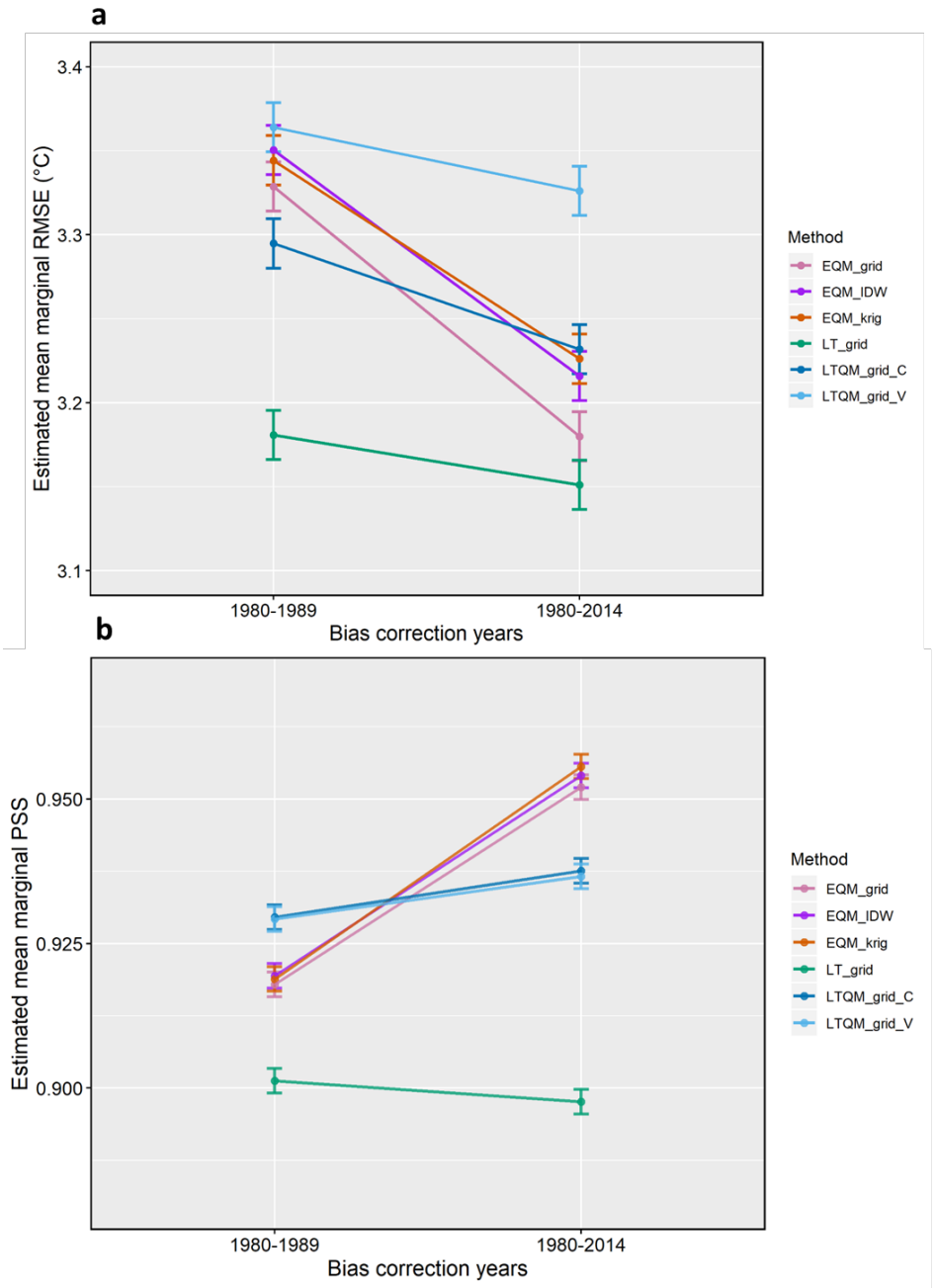

Figure 2.8: Interaction plots for Method × Bias_correction_years *("1980-1989" and "1980-2014" refer to GHCND station datasets used to bias-correct 1990-2014 and 1980-2014 WRF simulations, respectively). Plots show estimated mean marginal RMSE (a) and PSS (b) for the 1980-1989 and 1980-2014 calibration time periods. Error bars represent 95% confidence intervals.*

Elevation

RMSE Generally, RMSE improved when fine-scale elevation was accounted for during the downscaling process (Figure 2.9 a). η^2 for *Elevation* was nearly 19 times larger than that of *Elevation* $\times$ *Method* ($\eta^2 = 0.013$ and 0.00069, respectively; Table 2.2), indicating that the main effect of *Elevation* was more important in the RMSE ANOVA model than the interaction term. The relative importance of *Elevation* in the model was similar to the interaction *Month* $\times$ *Method* ($\eta^2 = 0.014$). Mean marginal RMSE of LT_grid was significantly less than, and mean marginal RMSE of LTQM_grid_V was significantly greater than that of all other methods, regardless of whether elevation was accounted for or not.

PSS The effect of *Elevation* was not significant in the full model for PSS (Table 2.7), and *Elevation* did not have any appreciable effect on PSS (Figure 2.9 b).

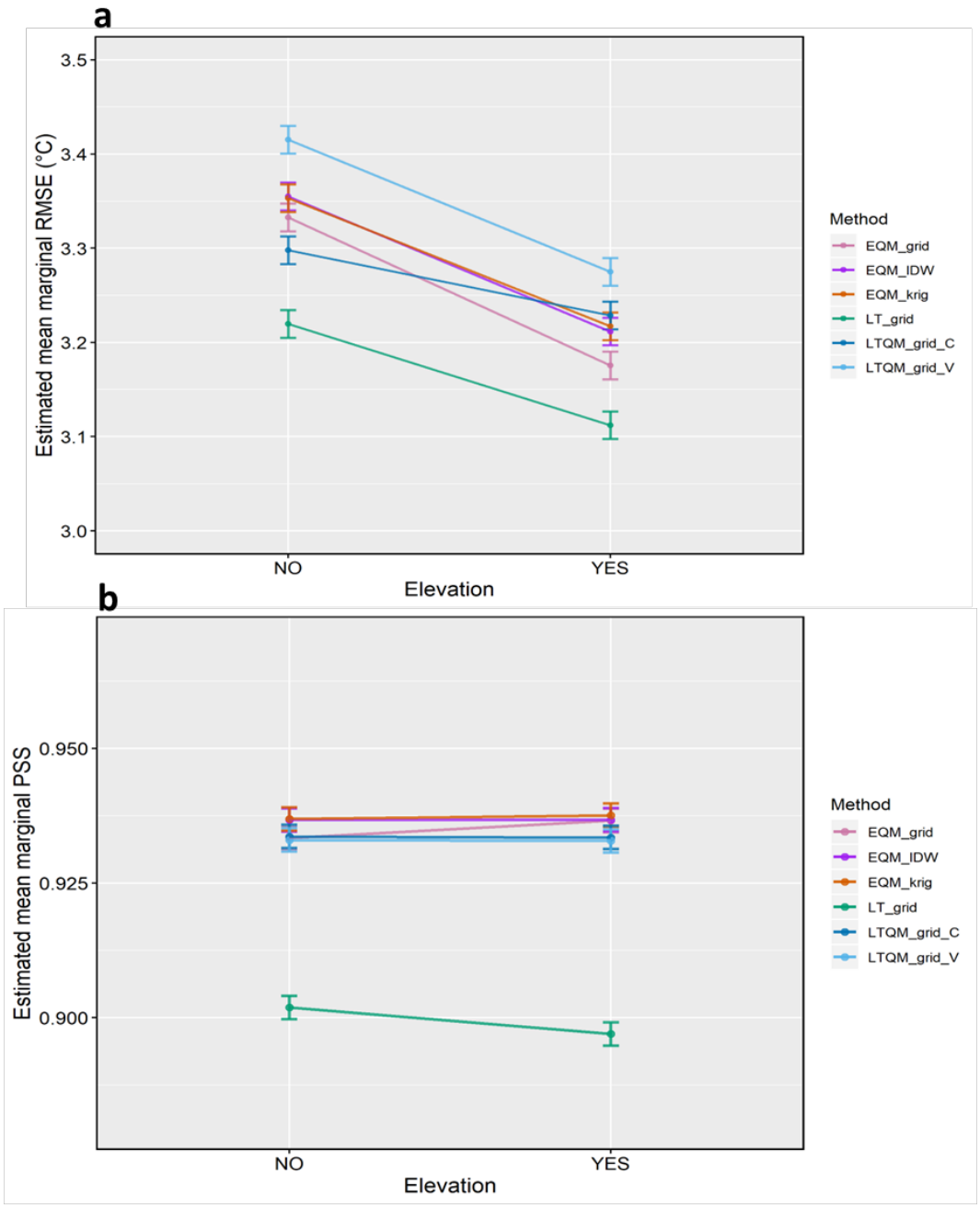

Figure 2.9: Interaction plots showing estimated mean marginal (a) RMSE and (b) PSS for Method × Elevation. *Error bars represent 95% confidence intervals.*

Table 2.4: Summary table for the final RMSE ANOVA model

	Degrees of freedom	Sum Squares	Mean Square	F value	Pr(>F)
Month	11	84.43	7.68	1485.08	0.0000
Bias_correction_years	1	0.57	0.57	109.68	0.0000
Elevation	1	1.14	1.14	219.73	0.0000
Method	5	0.82	0.16	31.74	0.0000
Month × Method	55	1.28	0.02	4.51	0.0000
Method × Bias_correction_years	5	0.16	0.03	6.18	0.0000
Method × Elevation	5	0.06	0.01	2.40	0.0388
Residuals	204	1.05	0.01		

Table 2.5: Summary table for the final PSS ANOVA model

	Degrees of freedom	Sum Squares	Mean Square	F value	Pr(>F)
Month	11	0.02	0.00	18.58	0.0000
Method	5	0.05	0.01	95.13	0.0000
Bias_correction_years	1	0.03	0.03	253.94	0.0000
Month × Method	55	0.06	0.00	9.48	0.0000
Method × Bias_correction_years	5	0.02	0.00	34.42	0.0000
Residuals	210	0.02	0.00		

2.5 Discussion

In this study, we developed six novel strategies for constructing high-resolution, bias-corrected gridded climate products of daily historical TMAX simulations from a regional climate model, where all bias-correction was based solely on station data. The six methods we present result in a substantial improvement over raw WRF simulations, are straightforward to implement, and can be applied to historical simulations as well as future projections of temperature variables. We found that no one method could concomitantly minimize RMSE and maximize PSS, which highlights the difficulty in correcting both the overall distributional discrepancies as well day-to-day discrepancies between simulated and observed data. Although performances were similar, the methods differ considerably in their ability to correct overall distributional discrepancies and day-to-day discrepancies between simulated and observed data. This is due mainly to the type of bias-correction, rather than the spatial

interpolation technique implemented in the methods. Distributional similarity, as measured by PSS, is achieved by matching the quantile-mapping techniques such as EQM. However, enhancing day-to-day correspondence, as measured by RMSE, is accomplished most effectively through a linear transfer (LT) function obtained by temporally-ordered linear regression between simulated and observed data. The most effective method thus depends on what is deemed most important in a particular application, day-to-day correspondence or distributional similarity, of simulated and observed data. Performance is further affected by seasonal bias of raw WRF simulations, the calibration period used for bias-correction, and the inclusion of elevation information during the downscaling step.

2.5.1 Monthly variation in performance

Our results show that performance of the six methods is affected by seasonal bias of WRF simulations. Raw WRF simulations for TMAX exhibited a distinct cold bias in winter and early spring and a slight warm bias in summer compared to GHCND station data. [130] compared TMAX simulations resulting from several parameterizations of WRF to Daymet gridded TMAX data and found that all parameterizations resulted in an annual cold bias during cold-season (months 11-5). The underestimation of mean temperature between months 11 and 5 is mainly due to the radiation scheme [130], and the best parameterization is reflected in the WRF data used for this study. [151] found that WRF run with a similar radiation scheme as used in this study overestimated the albedo of snow over the Tibetan Plateau, which is consistent with the winter cold bias in WRF simulations over the Lake Champlain Basin. As a result of the cold bias, performance in terms of both RMSE and to a lesser de-

gree, PSS declined in cold season months. Ideally, bias-correction techniques should improve both RMSE and PSS despite pronounced seasonal bias in raw WRF data.

PSS

Although raw WRF simulations exhibited considerable seasonal bias, EQM-based methods consistently improved PSS. However, for methods LT_grid_C and LT_grid_V, in which bias-correction was done with a simple quantile-mapping LT function, monthly PSS varied substantially throughout the year, and compared to PSS of EQM-based methods, was especially low in months 1-4. Rank-ordered regression (LTQM_grid_C and LTQM_grid_V) is a quantile-mapping technique similar in functionality to EQM. However, the resulting LT function (an estimated regression line) has limitations when the quantile-quantile map of simulated and observed data is nonlinear, as in our study. In cold season months, the quantile-quantile map of WRF and GHCND station data was nonlinear, especially in the tails, while during warmer months, it was relatively linear (Figure 2.10). The nonlinear quantile-quantile map of cold-season months contributed to the poor performance of LTQM_grid_C and LTQM_grid_V in improving PSS. This result supports findings by [144], who also found that a rank-ordered regression approach could not adequately correct seasonal biases of simulated data when quantile-quantile maps between observed and simulated data were nonlinear.

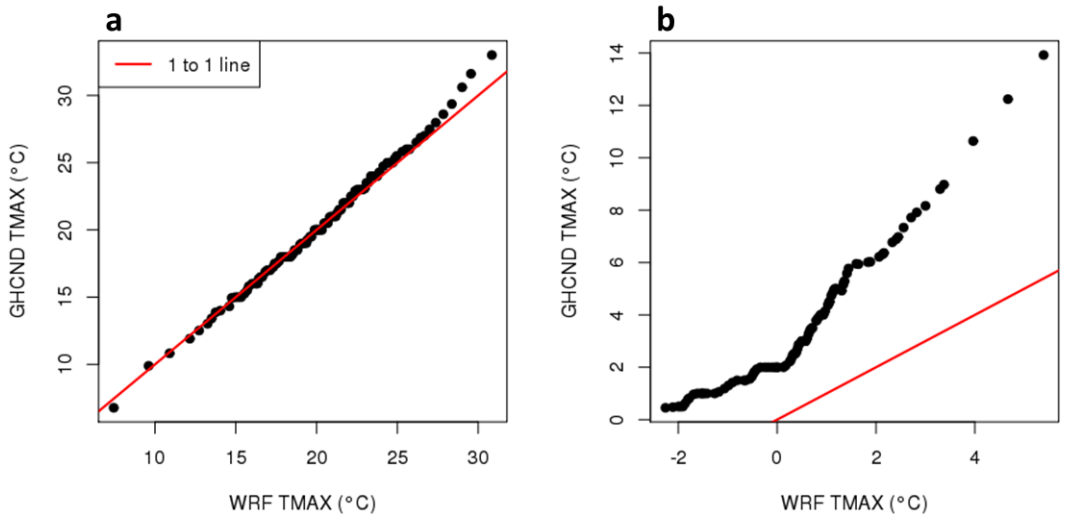

Figure 2.10: Quantile-quantile maps of GHCND station and WRF TMAX at a sample GHCND station location for (a) September and (b) February. One hundred estimated quantiles derived from 1980-2014 data are shown.

However, nonlinear quantile-quantile maps can be successfully approximated in most implementations of EQM through linear interpolation, splines, or other smoothing techniques [102]. Accordingly, we found that EQM-based methods substantially outperformed rank-ordered regression-based methods, LTQM_grid_C and LTQM_grid_V, in months ∼11-4. Ranked-ordered regression, in contrast to EQM, can only correct the first and second moments of a distribution, also contributing to the poor performance of LTQM_grid_C and LTQM_grid_V in months 11-4.

We did not find appreciable differences in the performance between the two simple quantile-mapping-based methods, LTQM_grid_C and LTQM_grid_V, despite the fact that LT parameter estimates varied over the fine-scale grid in LTQM_grid_V, whereas in LTQM_grid_C, LT parameter estimates were spatially constant. Although inspection of empirical variograms of estimated regression parameters sug-

gested some degree of spatial autorocorrelation, the amount of autocorrelation was likely not sufficient to produce a measurable difference in results. In our study, kriging LT parameter estimates over the fine-scale grid had no noticeable advantage over the use of spatially constant LT parameter estimates.

RMSE

The bias-correction techniques implemented in the six methods generally reduced RMSE compared to raw WRF simulations, but in contrast to PSS, bias-corrected WRF data RMSE from all methods exhibited a strong seasonal pattern. Quantile-mapping techniques (e.g. EQM and rank-ordered regression) perform well in correcting distributional discrepancies between simulated and observed data. However, the temporal ordering of data is not preserved in quantile-mapping techniques, which has a negative impact on the improvement of day-to-day discrepancies (RMSE). This has also been noted in other studies. For example, [152] found that quantile-mapping bias-correction techniques resulted in the greatest rank-ordered correlation (highest PSS) between simulated and observed data but did not appreciably improve RMSE.

Bias-correction techniques that preserve the temporal order of data, such as that implemented in LT_grid, outperformed quantile-mapping techniques in reducing day-to-day discrepancies. However, even after correction with a temporally ordered regression, seasonal patterns in RMSE persisted. In LT_grid, the transfer function was derived by fitting a (temporally-ordered) linear regression. Due to the strong cold bias in raw WRF data in cold-season months, the daily discrepancy between WRF and GHCND station data was, on average, larger during cold-season months compared to warm-season months. Therefore, regression lines fit in cold-season months were char-

acterized by a relatively large residual standard error. Consequently, the resulting LT function was not as effective in reducing RMSE, and seasonal patterns in RMSE persisted after bias-correction. If raw simulated data exhibit pronounced seasonal bias, substantial improvements in RMSE are difficult to achieve. Of all the bias-correction techniques evaluated in this study, temporally-ordered linear regression (LT grid), was most effective at reducing day-to-day discrepancies (RMSE) between simulated and observed data.

2.5.2 Implications for future and historical downscaling

To gain further insight into the suitability of methods for future or historical bias-correction, we used 1) 1980-1989 GHCND station data to bias-correct 1990-2014 WRF simulations and 2) 1980-2014 GHCND station data to bias-correct 1980-2014 WRF simulations. The 1980-2014 WRF simulations bias-corrected with 1980-2014 GHCND station data generally exhibited improved PSS and RMSE compared to 1990-2014 WRF simulations that were bias-corrected with the 1980-1989 GHCND subset.

With respect to distributional similarity (PSS), EQM outperformed rank-ordered regression when the 1980-2014 GHCND dataset was used for bias-correction, while the converse was true when the 1980-1989 GHCND dataset was used for bias-correction. The bias-variance trade-off, a well-known concept in statistical learning [153], can help to explain this result. Simple statistical models with few parameters, such as temporally- and rank-ordered linear regression, have high bias but low variance.

Highly flexible techniques, that require estimation of more parameters, such as EQM, have low bias but high variance [153]. Highly flexible models result in low training errors, as they can fit training data very well. However, they are also more prone to overfitting on training data, making them less able to generalize to new data [153]. The EQM transfer function will result in a nearly perfect quantile mapping if observational and simulated data of the same time period are used in its construction. When the quantile-mapping transfer function is subsequently used to bias-correct simulated data from the same time period, the correction will adjust simulated quantiles to very closely match those of observed quantiles. This explains why EQM performed very well when bias-correction was based on the 1980-2014 GHCND station dataset and the correction was also applied to 1980-2014 WRF data.

However, when bias-correction was based on the 1980-1989 GHCND dataset, and the correction was applied to 1990-2014 WRF data, EQM-based methods did not perform as well in terms of correcting distributional similarity (PSS). This indicates some degree of overfitting and lack of robustness of EQM for bias-correcting future projections. Our results agree with [45] and [144], who found that as the calibration time period increased, so did the risk of overfitting EQM transfer functions. Bias-correction via rank-ordered regression (implemented in LTQM_grid_V and LTQM_grid_C) resulted in a higher PSS compared to EQM when the 1980-1989 GHCND dataset was used to correct 1990-2014 WRF simulations. Rank-ordered regression is less prone to overfitting, which means the resulting transfer function is better able to generalize to bias-correcting future projections.

In the context of this study, relatively simple bias-correction techniques (rank-ordered regression) may be better suited to bias-correct future projections. Flexible

techniques such as EQM work well when the transfer function is applied to simulated data from the same time period that was used to construct it. Our results support studies in which simple methods such as linear scaling [54], rank-ordered regression [144], and multiple linear regression [154] performed as well as more sophisticated techniques for bias-correcting future temperature projections.

In terms of improving day-to-day correspondence, temporally ordered regression, LT_grid outperformed all other methods when the 1980-1989 calibration set was used for bias-correction. When bias-correction was based on the 1980-2014 calibration dataset, both LT_grid and EQM_grid performed significantly better than all other methods, regardless of which GHCND dataset was used for bias-correction.

2.5.3 Elevation

Elevation had comparatively little impact on performance of the methods compared to monthly variation. The distribution of elevation among GHCND stations in our study area was not particularly representative of the elevation range and topography study area. Elevations in the Lake Champlain Basin range from 30 to 1500m, whereas the majority of GHCND station elevations ranged from 30 and 400m, and only one station had an elevation greater than 1000m. This imbalance is unfortunate but not uncommon in climate studies [11], [31], [55], [78]. Including fine-scale elevation during downscaling is generally recommended, but likely due to the lack of higher elevation stations, its corrective effect on temperature was not as pronounced. [155] also found that when interpolating daily minimum and maximum temperatures from climate stations over a topographically complex region, interpolations were most accurate when station elevations accurately represented the topography of the study

region. Including fine-scale elevation data in the downscaling process did have minor positive influence on performance of the methods. It resulted in improved day-to-day discrepancies (RMSE), but it had no appreciable impact on distrubtional similarity (PSS). It could be that quantile-mapping bias-correction techniques had a greater positive influence on PSS than adjustments due to lapse rates.

2.6 Conclusion

The six high-resolution downscaling and bias-correction methods we presented in this study are straightforward and easy to implement, and depending on the method, result in substantially improved RMSE and PSS compared to uncorrected WRF simulations. Although we applied these methods to historical (1980-2014) daily maximum temperature simulations, methods could be applied to future climate projections and any type of temperature variable (minimum, maximum, or average). Although the ranges of performance metrics among methods were narrow, there were statistically significant differences in performance among methods, and we found that performance variation was mainly due to differences in bias-correction techniques. The selection of the most appropriate method for constructing high-resolution, bias-corrected temperature products using station data depends primarily on the intended use of the resulting data product.

We did not find that one method could simultaneously minimize RMSE and maximize PSS. Maximizing PSS is achieved by matching the quantiles of simulated and observed data (EQM), which, in turn increases distributional similarity between simulated and observed data. However, minimizing RMSE is achieved by decreasing the

discrepancy between daily modeled and observed data, which is done most effectively via a linear regression between simulated and observed data. Thus, bias-correction methods such as EQM improve PSS but not necessarily RMSE, whereas simple linear regression transfer functions improve RMSE but generally not PSS. While the objectives of minimizing RMSE and maximizing PSS are not mutually exclusive, they may be difficult to attain concomitantly in practice.

Based on our results, for processing historical simulations, the most effective method for improving day-to-day correspondence (RMSE) of simulated and observed data is LT_grid. In this method, simulations are bias-corrected at the fine-scale grid by transfer functions obtained from temporally-ordered regressions generated at station locations. In topographically varied regions performance is further enhanced by accounting for fine-scale elevation data in the downscaling process. Unlike rank-ordered regression, bias-correction with temporally ordered regression is more sensitive to seasonal bias in simulated data.

To achieve optimal distributional similarity (PSS) of simulated and observed data in historical downscaling, quantile mapping-based methods (EQM_IDW, EQM_krig and EQM_grid) are most effective. In EQM_IDW and EQM_krig, WRF simulations are bias-corrected at station locations and then interpolated to the fine-scale grid. In EQM_grid, bias-correction of interpolated WRF simulations occurs at the fine-scale grid, but this slight difference in workflow did not result in appreciable differences in performance compared to EQM_IDW and EQM_krig. The interpolation techniques (IDW, kriging) did not affect performance of EQM-based methods, which suggests that in creating full-coverage temperature products, deterministic interpolation techniques (IDW) perform as well as geostatistical techniques (kriging).

Moreover, performance of EQM-based methods did not benefit from the inclusion of fine-scale elevation data during downscaling. For historical downscaling, EQM-based methods are generally more resistant to seasonal bias in simulated data. However, EQM is susceptible to overfitting on calibration data and may not provide robust bias-correction for future projections.

Quantile-mapping methods in which LT functions are constructed through rank-ordered regression (LTQM_grid_V and LTQM_grid_C) are less prone to overfitting on calibration data. Therefore, such methods are better suited improving distributional similarity (PSS) of future temperature projections than EQM-based methods. In LTQM_grid_V and LTQM_grid_C, simulated data are bias-corrected at the fine-scale grid through transfer functions obtained from rank-ordered regressions generated at station locations. Similar to EQM-based methods, including fine-scale elevation data during downscaling does not significantly improve distributional similarity (PSS). We did not find differences in performance between LTQM_grid_V and LTQM_grid_C. As LTQM_grid_V accounts for spatial autcorrelation among LT parameter estimates at station locations, it would likely perform better if LT parameter estimates at station locations exhibit strong spatial autocorrelation.

2.7 Appendix

Table 2.6: A1. ANOVA table for full RMSE model

	Degrees of freedom	Sum Squares	Mean Square	F value	Pr(>F)
Month	11	84.43	7.68	1526.98	0.0000
Method	5	0.82	0.16	32.64	0.0000
Bias_correction_years	1	0.57	0.57	112.77	0.0000
Elevation	1	1.14	1.14	225.93	0.0000
Month × Method	55	1.28	0.02	4.63	0.0000
Method × Bias_correction_years	5	0.16	0.03	6.35	0.0000
Method × Elevation	5	0.06	0.01	2.46	0.0343
Bias_correction_years × Elevation	1	0.02	0.02	3.71	0.0555
Method × Bias_correction_years × Elevation	5	0.04	0.01	1.61	0.1593
Residuals	198	1.00	0.01		

Table 2.7: A2. ANOVA table for full PSS model

	Degrees of freedom	Sum Squares	Mean Square	F value	Pr(>F)
Month	11	0.02	0.00	17.88	0.0000
Method	5	0.05	0.01	91.55	0.0000
Bias_correction_years	1	0.03	0.03	244.40	0.0000
Elevation	1	0.00	0.00	0.03	0.8579
Month × Method	55	0.06	0.00	9.13	0.0000
Method × Bias_correction_years	5	0.02	0.00	33.13	0.0000
Method × Elevation	5	0.00	0.00	0.74	0.5910
Bias_correction_years × Elevation	1	0.00	0.00	0.24	0.6266
Method × Bias_correction_years × Elevation	5	0.00	0.00	0.02	0.9997
Residuals	198	0.02	0.00		

References

[4] A. Zia, A. Bomblies, A. W. Schroth, *et al.*, "Coupled impacts of climate and land use change across a river–lake continuum: Insights from an integrated assessment model of lake champlain's missisquoi basin, 2000–2040," *Environmental Research Letters*, vol. 11, no. 11, p. 114 026, 2016.

[5] M. Ekström, M. R. Grose, and P. H. Whetton, "An appraisal of downscaling methods used in climate change research," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 6, no. 3, pp. 301–319, 2015.

[6] L. R. Leung, L. O. Mearns, F. Giorgi, and R. L. Wilby, "Regional climate research: Needs and opportunities," *Bulletin of the American Meteorological Society*, vol. 84, no. 1, pp. 89–95, 2003.

[7] D. Maraun, T. G. Shepherd, M. Widmann, *et al.*, "Towards process-informed bias correction of climate change simulations," *Nature Climate Change*, vol. 7, no. 11, p. 764, 2017.

[8] T. Lafon, S. Dadson, G. Buys, and C. Prudhomme, "Bias correction of daily precipitation simulated by a regional climate model: A comparison of methods," *International Journal of Climatology*, vol. 33, no. 6, pp. 1367–1381, 2013.

[9] A. J. Cannon, C. Piani, and S. Sippel, "Bias correction of climate model output for impact models," in *Climate Extremes and Their Implications for Impact and Risk Assessment*, Elsevier, 2020, pp. 77–104.

[10] I. Hanssen-Bauer, C. Achberger, R. Benestad, D. Chen, and E. Førland, "Statistical downscaling of climate scenarios over Scandinavia," *Climate Research*, vol. 29, no. 3, pp. 255–268, 2005.

[11] Z. A. Holden, J. T. Abatzoglou, C. H. Luce, and L. S. Baggett, "Empirical downscaling of daily minimum air temperature at very fine resolutions in complex terrain," *Agricultural and Forest Meteorology*, vol. 151, no. 8, pp. 1066–1073, 2011.

[13] P. Caldwell, H.-N. S. Chin, D. C. Bader, and G. Bala, "Evaluation of a WRF dynamical downscaling simulation over California," *Climatic change*, vol. 95, no. 3-4, pp. 499–521, 2009.

[14] R. L. Wilby, S. Charles, E. Zorita, B. Timbal, P. Whetton, and L. Mearns, "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 TGCIA*, vol. 27, 2004.

[15] F. Feser, B. Rockel, H. von Storch, J. Winterfeldt, and M. Zahn, "Regional climate models add value to global model data: A review and selected examples," *Bulletin of the American Meteorological Society*, vol. 92, no. 9, pp. 1181–1192, 2011.

[16] F. Giorgi, C. Jones, G. R. Asrar, *et al.*, "Addressing climate information needs at the regional level: The cordex framework," *World Meteorological Organization Journal*, vol. 58, no. 3, p. 175, 2009.

[21] L. Mearns, F. Giorgi, P. Whetton, D. Pabon, M. Hulme, and M. Lal, "Guidelines for use of climate scenarios developed from regional climate model experiments," *Data Distribution Centre of the Intergovernmental Panel on Climate Change*, 2003.

[22] G. Fang, J. Yang, Y. Chen, and C. Zammit, "Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in china," *Hydrology and Earth System Sciences*, vol. 19, no. 6, pp. 2547–2559, 2015.

[23] R. Huth, "Statistical downscaling in central europe: Evaluation of methods and potential predictors," *Climate Res.*, vol. 13, no. 2, pp. 91–101, 1999.

[24] H. Kettle and R. Thompson, "Statistical downscaling in european mountains: Verification of reconstructed air temperature," *Climate Res.*, vol. 26, no. 2, pp. 97–112, 2004.

[25] J. T. Schoof and S. Pryor, "Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks," *International Journal of Climate*, vol. 21, no. 7, pp. 773–790, 2001.

[26] T. Vandal, E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, "Deepsd: Generating high resolution climate change projections through single image super-resolution," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2017, pp. 1663–1672.

[27] C. Hutengs and M. Vohland, "Downscaling land surface temperatures at regional scales with random forest regression," *Remote Sensing of Environment*, vol. 178, pp. 127–141, 2016.

[31] J. M. Winter, B. Beckage, G. Bucini, R. M. Horton, and P. J. Clemins, "Development and evaluation of high-resolution climate simulations over the mountainous northeastern united states," *Journal of Hydrometeorology*, vol. 17, no. 3, pp. 881–896, 2016.

[34] T. Vandal, E. Kodra, and A. R. Ganguly, "Intercomparison of machine learning methods for statistical downscaling: The case of daily and extreme precipitation," *Theoretical and Applied Climatology*, vol. 137, no. 1, pp. 557–570, 2019.

[42] G. Lenderink, A. Buishand, and W. v. Deursen, "Estimates of future discharges of the river rhine using two scenario methodologies: Direct versus delta approach," *Hydrology and Earth System Sciences*, vol. 11, no. 3, pp. 1145–1159, 2007.

[43] L. E. Hay, R. L. Wilby, and G. H. Leavesley, "A comparison of delta change and downscaled gcm scenarios for three mountainous basins in the united states 1," *JAWRA Journal of the American Water Resources Association*, vol. 36, no. 2, pp. 387–397, 2000.

[44] R. Leander and T. A. Buishand, "Resampling of regional climate model output for the simulation of extreme river flows," *Journal of Hydrology*, vol. 332, no. 3-4, pp. 487–496, 2007.

[45] C. Piani, J. Haerter, and E. Coppola, "Statistical bias correction for daily precipitation in regional climate models over europe," *Theoretical and Applied Climatology*, vol. 99, no. 1-2, pp. 187–192, 2010.

[46] C. Teutschbein and J. Seibert, "Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods," *Journal of Hydrology*, vol. 456, pp. 12–29, 2012.

[47] A. J. Cannon, S. R. Sobie, and T. Q. Murdock, "Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?" *Journal of Climate*, vol. 28, no. 17, pp. 6938–6959, 2015.

[48] A. W. Wood, E. P. Maurer, A. Kumar, and D. P. Lettenmaier, "Long-range experimental hydrologic forecasting for the eastern united states," *Journal of Geophysical Research: Atmospheres*, vol. 107, no. D20, ACL–6, 2002.

[49] E. P. Maurer and P. B. Duffy, "Uncertainty in projections of streamflow changes due to climate change in california," *Geophysical Research Letters*, vol. 32, no. 3, 2005.

[50] K. Hayhoe, C. Wake, B. Anderson, *et al.*, "Regional climate change projections for the northeast usa," *Mitigation and Adaptation Strategies for Global Change*, vol. 13, no. 5-6, pp. 425–436, 2008.

[51] A. W. Wood, L. R. Leung, V. Sridhar, and D. Lettenmaier, "Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs," *Climatic Change*, vol. 62, no. 1-3, pp. 189–216, 2004.

[52] J. C. Bennett, M. R. Grose, S. P. Corney, *et al.*, "Performance of an empirical bias-correction of a high-resolution climate dataset," *International Journal of Climatology*, vol. 34, no. 7, pp. 2189–2204, 2014.

[53] M. Jakob Themeßl, A. Gobiet, and A. Leuprecht, "Empirical-statistical downscaling and error correction of daily precipitation from regional climate models," *International Journal of Climatology*, vol. 31, no. 10, pp. 1530–1544, 2011.

[54] M. Shrestha, S. C. Acharya, and P. K. Shrestha, "Bias correction of climate models for hydrological modelling–are simple methods still useful?" *Meteorological Applications*, vol. 24, no. 3, pp. 531–539, 2017.

[55] D. R. Roberts, W. H. Wood, and S. J. Marshall, "Assessments of downscaled climate data with a high-resolution weather station network reveal consistent but predictable bias," *International Journal of Climatology*, vol. 39, no. 6, pp. 3091–3103, 2019.

[74] N. Oceanic and A. Administration, *Climate data online search*, Accessed: 2017-09-30, 2018. [Online]. Available: https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND.

[76] B. Livneh, T. J. Bohn, D. W. Pierce, *et al.*, "A spatially comprehensive, hydrometeorological data set for mexico, the us, and southern canada 1950–2013," *Scientific data*, vol. 2, no. 1, pp. 1–12, 2015.

[77] P. E. Thornton, M. M. Thornton, B. W. Mayer, *et al.*, "Daymet: Daily surface weather on a 1 km grid for north america, 1980-2008," *Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center for Biogeochemical Dynamics (DAAC)*, 2012.

[78] C. Daly, G. Taylor, W. Gibson, T. Parzybok, G. Johnson, and P. Pasteris, "High-quality spatial climate data sets for the united states and beyond," *Transactions of the ASAE*, vol. 43, no. 6, p. 1957, 2000.

[79] R. Behnke, S. Vavrus, A. Allstadt, T. Albright, W. E. Thogmartin, and V. C. Radeloff, "Evaluation of downscaled, gridded climate data for the conterminous united states," *Ecological applications*, vol. 26, no. 5, pp. 1338–1351, 2016.

[80] D. Walton and A. Hall, "An assessment of high-resolution gridded temperature datasets over california," *Journal of Climate*, vol. 31, no. 10, pp. 3789–3810, 2018.

[82] T. C. Peterson and R. S. Vose, "An overview of the global historical climatology network temperature database," *Bulletin of the American Meteorological Society*, vol. 78, no. 12, pp. 2837–2850, 1997.

[83] I. Durre, M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, "Comprehensive automated quality assurance of daily surface observations," *Journal of Applied Meteorology and Climatology*, vol. 49, no. 8, pp. 1615–1633, 2010.

[84] J. F. Mejia, J. Huntington, B. Hatchett, D. Koracin, and R. G. Niswonger, "Linking global climate models to an integrated hydrologic model: Using an individual station downscaling approach," *J. Contemp. Water Res. Educ.*, vol. 147, no. 1, pp. 17–27, 2012.

[85] L. Poggio and A. Gimona, "Downscaling and correction of regional climate models outputs with a hybrid geostatistical approach," *Spat. Stat.*, vol. 14, pp. 4–21, 2015.

[102] L. Gudmundsson, *Qmap: Statistical transformations for post-processing climate model output*, R package version 1.0-4, 2016.

[112] O. Schabenberger and C. A. Gotway, *Statistical methods for spatial data analysis*. Chapman and Hall/CRC, 2017.

[118] L. E. Flint and A. L. Flint, "Downscaling future climate scenarios to fine scales for hydrologic and ecological modeling and analysis," *Ecological Processes*, vol. 1, no. 1, p. 2, 2012.

[119] J. Franklin, F. W. Davis, M. Ikegami, *et al.*, "Modeling plant species distributions under future climates: How fine scale do climate projections need to be?" *Global change biology*, vol. 19, no. 2, pp. 473–483, 2013.

[120] J. W. Hansen, "Integrating seasonal climate prediction and agricultural models for insights into agricultural practice," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1463, pp. 2037–2047, 2005.

[121] J. D. Fridley, "Downscaling climate over complex terrain: High finescale (less than 1000 m) spatial variation of near-ground temperatures in a montane forested landscape (great smoky mountains)," *Journal of Applied Meteorology and Climatology*, vol. 48, no. 5, pp. 1033–1049, 2009.

[122] D. Maraun, F. Wetterhall, A. Ireson, *et al.*, "Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user," *Reviews of Geophysics*, vol. 48, no. 3, 2010.

[123] R. Haas and J. G. Pinto, "A combined statistical and dynamical approach for downscaling large-scale footprints of european windstorms," *Geophysical Research Letters*, vol. 39, no. 23, 2012.

[124] T. Engen-Skaugen, "Refinement of dynamically downscaled precipitation and temperature scenarios," *Climate Change*, vol. 84, no. 3-4, pp. 365–382, 2007.

[125] Z. Han, Y. Shi, J. Wu, Y. Xu, and B. Zhou, "Combined dynamical and statistical downscaling for high-resolution projections of multiple climate variables in the beijing–tianjin–hebei region of china," *Journal of Applied Meteorology and Climatology*, vol. 58, no. 11, pp. 2387–2403, 2019.

[126] D. A. Bishop and C. M. Beier, "Assessing uncertainty in high-resolution spatial climate data across the us northeast," *PloS one*, vol. 8, no. 8, e70260, 2013.

[127] C. Xu, J. Wang, and Q. Li, "A new method for temperature spatial interpolation based on sparse historical stations," *Journal of Climate*, vol. 31, no. 5, pp. 1757–1770, 2018.

[128] W. C. Skamarock, J. B. Klemp, J. Dudhia, *et al.*, "A description of the advanced research wrf model," 2019, Accessed: 2019-03-04. [Online]. Available: https://opensky.ucar.edu/islandora/object/opensky:2898.

[129] D. P. Dee, S. M. Uppala, A. Simmons, *et al.*, "The era-interim reanalysis: Configuration and performance of the data assimilation system," *Quarterly Journal of the royal meteorological society*, vol. 137, no. 656, pp. 553–597, 2011.

[130] H. Huang, J. M. Winter, E. C. Osterberg, *et al.*, "Simulating precipitation and temperature in the lake champlain basin using a regional climate model: Limitations and uncertainties," *Climate Dynamics*, vol. 54, no. 1-2, pp. 69–84, 2020.

[131] USGS, *The national map*, 2018. [Online]. Available: https://viewer.nationalmap.gov/basic/.

[132] T. Wang, A. Hamann, D. L. Spittlehouse, and T. Q. Murdock, "ClimateWNA—high-resolution spatial climate data for western North America," *Journal of Applied Meteorology and Climatology*, vol. 51, no. 1, pp. 16–29, 2012.

[133] R. G. Barry, *Mountain weather and climate*. Psychology Press, 1992.

[134] E. P. Boer, K. M. de Beurs, and A. D. Hartkamp, "Kriging and thin plate splines for mapping climate variables," *International Journal of Applied Earth Observation and Geoinformation*, vol. 3, no. 2, pp. 146–154, 2001.

[135] S. J. Jeffrey, J. O. Carter, K. B. Moodie, and A. R. Beswick, "Using spatial interpolation to construct a comprehensive archive of australian climate data," *Environmental Modeling and Software*, vol. 16, no. 4, pp. 309–330, 2001.

[136] C. K. Wikle, A. Zammit-Mangion, and N. Cressie, *Spatio-temporal Statistics with R*. CRC Press, 2019.

[137] C. Daly, "Guidelines for assessing the suitability of spatial climate data sets," *International Journal of Climatology*, vol. 26, no. 6, pp. 707–721, 2006.

[138] B. Gräler, E. Pebesma, and G. Heuvelink, "Spatio-temporal interpolation using gstat," *The R Journal*, vol. 8, pp. 204–218, 1 2016. [Online]. Available: https://journal.r-project.org/archive/2016/RJ-2016-014/index.html.

[139] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.R-project.org/.

[140] E. Varouchakis and D. Hristopulos, "Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins," *Environmental monitoring and assessment*, vol. 185, no. 1, pp. 1–19, 2013.

[141] N. Hofstra, M. Haylock, M. New, P. Jones, and C. Frei, "Comparison of six methods for the interpolation of daily, european climate data," *Journal of Geophysical Research: Atmospheres*, vol. 113, no. D21, 2008.

[142] G. A. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, 2012, vol. 329.

[143] J. Pilz and G. Spöck, "Why do we need and how should we implement bayesian kriging methods," *Stochastic Environmental Research and Risk Assessment*, vol. 22, no. 5, pp. 621–632, 2008.

[144] P. Berg, H. Feldmann, and H.-J. Panitz, "Bias correction of high resolution regional climate model data," *Journal of Hydrology*, vol. 448, pp. 80–92, 2012.

[145] S. Perkins, A. Pitman, N. Holbrook, and J. McAneney, "Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions," *Journal of climate*, vol. 20, no. 17, pp. 4356–4376, 2007.

[146] D. Keellings, "Evaluation of downscaled cmip5 model skill in simulating daily maximum temperature over the southeastern united states," *International Journal of Climatology*, vol. 36, no. 12, pp. 4172–4180, 2016.

[147] R. Bordoy and P. Burlando, "Bias correction of regional climate model simulations in a region of complex orography," *Journal of applied meteorology and climatology*, vol. 52, no. 1, pp. 82–101, 2013.

[148] R. Lenth, *Emmeans: Estimated marginal means, aka least-squares means*, R package version 1.4.4, 2020. [Online]. Available: https://CRAN.R-project.org/package=emmeans.

[149] T. R. Levine and C. R. Hullett, "Eta squared, partial eta squared, and misreporting of effect size in communication research," *Human Communication Research*, vol. 28, no. 4, pp. 612–625, 2002.

[150] K. E. Muller and B. L. Peterson, "Practical methods for computing power in testing the multivariate general linear hypothesis," *Computational Statistics and Data Analysis*, vol. 2, no. 2, pp. 143–158, 1984.

[151] X. Meng, S. Lyu, T. Zhang, *et al.*, "Simulated cold bias being improved by using modis time-varying albedo in the tibetan plateau in wrf model," *Environmental Research Letters*, vol. 13, no. 4, p. 044 028, 2018.

[152] C. Goodess, C. Anagnostopoulou, A. Bárdossy, *et al.*, "An intercomparison of statistical downscaling methods for europe and european regions–assessing their performance with respect to extreme temperature and precipitation events," *Climatic Change*, pp. 1–68, 2007.

[153] J. Friedman, T. Hastie, R. Tibshirani, *et al.*, *The elements of statistical learning*, 10. Springer series in statistics New York, 2001, vol. 1.

[154] H. J. Fowler, S. Blenkinsop, and C. Tebaldi, "Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling," *International journal of climatology*, vol. 27, no. 12, pp. 1547–1578, 2007.

[155] K. Stahl, R. Moore, J. Floyer, M. Asplin, and I. McKendry, "Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density," *Agricultural and Forest Meteorology*, vol. 139, no. 3-4, pp. 224–236, 2006.

Chapter 3

Robust bias-correction of precipitation extremes using a novel hybrid empirical quantile mapping method: advantages of a linear correction for extremes

3.1 Abstract

High-resolution, daily precipitation climate products that realistically represent extremes are critical for evaluating local-scale climate impacts. A popular bias-correction method, empirical quantile mapping (EQM), can generally correct distributional discrepancies between simulated climate variables and observed data but can be highly

sensitive to the choice of calibration period and is prone to overfitting. In this study, we propose a hybrid bias-correction method for precipitation, EQM-LIN, which combines the efficacy of EQM for correcting lower quantiles, with a robust linear correction for upper quantiles. We apply both EQM and EQM-LIN to historical daily precipitation data simulated by a regional climate model over a region in the northeastern United States. We validate our results using a five-fold cross-validation and quantify performance of EQM and EQM-LIN using skill score metrics and several climatological indices. As part of a high-resolution downscaling and bias-correction workflow, EQM-LIN significantly outperforms EQM in reducing mean, and especially extreme, daily distributional biases present in raw model output. EQM-LIN performed as good or better than EQM in terms of bias-correcting standard climatological indices (e.g., total annual rainfall, frequency of wet days, total annual extreme rainfall). In addition, our study shows that EQM-LIN is particularly resistant to overfitting at extreme tails and is much less sensitive to calibration data, both of which can reduce the uncertainty of bias-correction at extremes.

3.2 Introduction

Climate data is often necessary for social, ecological, and hydrological models and is routinely used in climate impact models and assessments. Model reliability is largely dependent on the quality and resolution of climate data products [11], [118], [119], [156]. The representation of extremes, in particular, can have a disproportionately large effect on such models [157]. Increases in the frequency, variability, and magnitude of extreme precipitation over the last several decades, especially in the north-

eastern United States, are well-documented [158], [159]. To study the future impacts of changing extremes at local scales, climate data products must represent extreme events accurately and be available at fine spatial and temporal resolutions [157]. General circulation models (GCMs) provide important information about historical and future larger-scale climate trends, but their resolution is too coarse to investigate localized effects of changes in extreme climate events [5], [8]. Additionally, raw GCM output is characterized by a non-trivial degree of bias [8], and the ability of GCMs to reproduce extreme tails of climate variables is limited [44]. Therefore, prior to its use in hydrological [54], [160], agricultural [161], or ecological models, GCM output is downscaled to a finer resolution and bias-corrected with respect to observed data [4]. These post-processing techniques result in climate data that is more realistic at finer spatial scales. Here, we propose a bias-correction method that more accurately captures precipitation extremes. We incorporate it into a high-resolution downscaling and bias-correction workflow for constructing daily, high- resolution data products for use in modeling efforts.

In the process of downscaling, model output is converted from a coarse to finer resolution. In dynamical downscaling, a regional climate model (RCM) is forced with a GCM, resulting in finer-scale output in which regional climate processes, topography, and orography are incorporated [15]. In statistical downscaling, statistical relationships between coarse-scale climate variables and local, observed data are established, and the effects of fine-scale predictors are integrated into downscaled data [122]. Dynamical downscaling is computationally intensive and can introduce additional biases [6], [13], but, localized climate processes, including extremes [17], are generally better reproduced than in GCMs [18]. However, RCMs do not perform well in capturing the

most extreme events [44], [162]. Statistical downscaling is efficient, can be applied to a variety of climate variables [21], and is especially effective in topographically complex terrain [10]. Climate data products with fine spatial resolutions, which are important for studying localized changes in extreme climate events, can be generated by combining statistical and dynamic downscaling, [163]. In this study, we combine statistical and dynamical downscaling to produce precipitation data products with a fine spatial resolution.

Downscaling is complemented by bias-correction, a procedure in which climate model output is adjusted such that its statistical properties (e.g. mean, variance, and potentially higher moments) resemble those of observations in a common climatological period [8], [9]. We note that the terms "downscaling" and "bias-correction" are sometimes used to refer to equivalent processes. However, in this study, downscaling only refers to the process in which coarse, gridded climate data is interpolated to a finer spatial resolution, and bias-correction refers specifically to applying transformations to climate model output such that distributional biases are reduced. Most bias-correction methods assume stationarity of model errors over time [55], which can be problematic for bias-correcting future climate model output over multi-decadal time spans [47], [154]. In addition, sufficient observational data is necessary to derive robust transfer functions [154]. Bias-correction methods for precipitation range from simple approaches such as the "delta change" or "delta factor" method [46] to more flexible and effective quantile-mapping based methods [46]–[48]. In quantile-mapping (QM) based methods, a transfer function (TF) maps quantiles of climate model output to those of observed data. QM methods can be parametric [45], non-parametric [8], or a combination of both [106]. Distribution mapping (DM) is a parametric

QM method in which known, parametric distributions are fit to observed and model data. The Gamma distribution is often used to model wet-day precipitation (e.g. [8], [61], [110]) but is generally not adequate for modeling extreme precipitation tails [63], [66]. Hybrid DM approaches in which the Gamma distribution is fit to lower quantiles and a heavy-tailed distribution is fit to tail quantiles can improve bias-correction of extreme precipitation [66], [164]. A non-parametric counterpart to DM, empirical quantile mapping (EQM), is a flexible method in which no distributional assumptions are made. In EQM, the TF represents a mapping from empirical model quantiles to observed quantiles and typically outperforms DM [53], [67]. EQM is effective in correcting precipitation variables [22], [53], [73], [165] and is attractive as a bias-correction method as it corrects the mean, standard deviation, and higher-order distributional moments [61].

A disadvantage of QM methods and EQM in particular, is their propensity to overfit on calibration data, especially at precipitation extremes where data is scarce and highly variable [8], [45], [99]–[101]. In EQM, TFs are interpolated using linear interpolation, splines, or other smoothing techniques [102]. Flexible methods such as EQM can result in TFs that can correct model data nearly perfectly (overfitting) but may not generalize to out-of-sample or future model data. Overfitting is problematic because it can lead to instability of the TF at higher quantiles [99], [103], [104]. When applied to future projections, EQM has been shown to significantly distort future climate change signals [7], [105] and exaggerate or deflate extreme trends, introducing additional uncertainty into bias-corrected data [47], [106]. Hybrid EQM approaches that combine parametric and non-parametric modeling can reduce the degree of overfitting of the TF at extreme tails [106]. In a hybrid approach, bias-correction below a

specified threshold is achieved via a non-parametric TF (EQM), while bias-correction above the threshold is with DM, based on an extreme distribution, such as the Generalized Pareto distribution [106]. Hybrid EQM methods combine the flexbility of EQM for correcting lower to middle quantiles with the robustness of parametric distributions for correcting upper quantiles. In particular, the use of extreme or heavy-tailed distributions for modeling extremes can improve bias-correction of tail quantiles [66], [101], [106]–[109]. However, the risk of overfitting the TF at distributional tails still exists, as poor fits to heavy-tailed distributions can introduce outliers [69], [110]. In addition, selection of the threshold is difficult, as the amount of data beyond the threshold must be sufficiently large to allow for distribution fitting and must approximate a known heavy-tailed distribution [66], [111]. There is a need for a hybrid EQM method in which bias-correction of extremes can be performed without the risk of overfitting and the introduction of outliers.

We propose and demonstrate a simple, hybrid EQM method for bias-correction that, when used in conjunction with downscaling, results in high-resolution (1km) daily precipitation data in which precipitation extremes are accurately represented. The proposed method, EQM-LIN, combines the effectiveness of EQM for correcting the bulk of the distribution with a robust, linear correction for extremes. As part of a high-resolution, downscaling and bias-correction workflow, we use EQM-LIN to bias-correct historical (1976-2005), daily precipitation data that were dynamically downscaled by a regional climate model (RCM). We also compare the effectiveness of EQM-LIN to EQM for bias-correction, with an emphasis on the ability of the two methods to accurately capture extremes. Because EQM-LIN is computationally cheap, easy to apply, and corrects both mean and extreme bias for precipitation

variables, it is an important methodological addition to the body of bias-correction literature.

3.3 Methods

3.3.1 Data

The study area, the Lake Champlain Basin, consists of parts of Vermont, New Hampshire, eastern New York, United States and southern Quebec, Canada (Figure 3.1). Eleven watersheds drain into Lake Champlain, and the Green Mountains, Adirondack Mountains, and White Mountains span portions of Vermont, New York, and New Hampshire, respectively [31]. The study area is approximately 82,657km^2. Elevations range from 30 to 1500 m above mean sea level (MSL). The study area is characterized by a subhumid continental climate with cold and snowy winters. At high elevations, mean annual precipitation can reach 1,000-1,520mm, while at low elevations, mean annual precipitation ranges between 750-900mm; locally intense precipitation in the form of thunderstorms is likely during summer months [166].

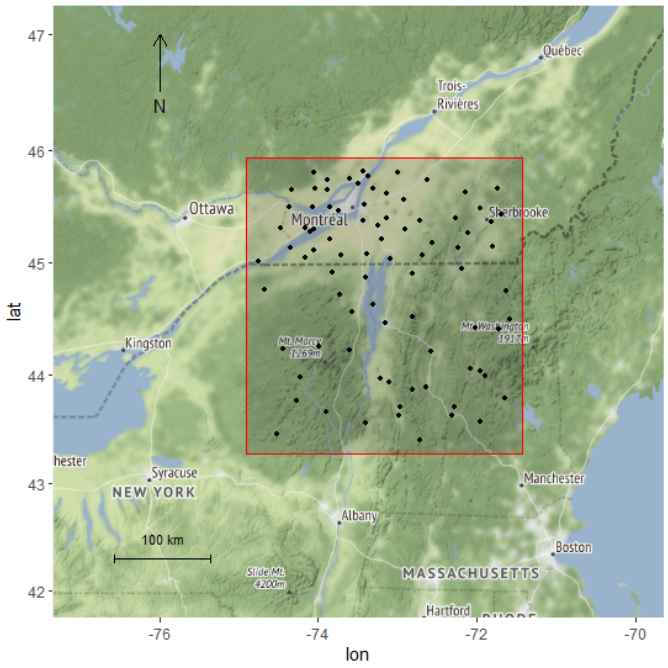

Figure 3.1: GHCND stations (black) within the study area (red). The study area is approximately 13,251km^2.

Simulated historical (1976-2005) precipitation (PRCP) data were generated by the Advanced Weather and Research Forecasting model (WRF) version 3.9.1, an RCM [128]. WRF output was generated at a daily temporal resolution. WRF is a widely used numerical weather prediction system for both research and applied forecasting purposes [128]. Historical simulations (1976-2005) were forced by bias-corrected Community Earth System Model 1 (CESM1), a GCM [167]. CESM1 historical simulations were dynamically downscaled with WRF to a 4km resolution using three one-way nests (36 km, 12 km, 4km). The 4km resolution WRF data were used for this study. Additional WRF model details are included in the Supplementary Materials, and a full description and evaluation of simulations can be found in [130].

Historical daily climate station data was obtained from the Global Historical Cli-

mate Network (GHCND) (https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND). GHCND data records are adjusted to account for changes in instrumentation and other anomalies [74], [82]. We retained only those stations with at least 70% complete records over the historical time period 1976-2005 (85 stations). We chose to use station data, rather than gridded data products (e.g. Livneh, [76]; Daymet, [77]; and PRISM, [78]), because interpolation algorithms used to create gridded climate products can introduce bias [79] and additional uncertainty when used for bias-correcting climate model output [80]. Gridded products can misrepresent extreme tails [168], and [169] showed that Daymet, Livneh, and PRISM precipitation products varied widely in their representation of wet-day occurrences, length of wet and dry periods, and precipitation intensity in the South-Central United States. Station data represent direct climatological measurements and are available throughout the Northeast [82], [83]. We acknowledge that there is a spatial misalignment between gridded model data and point-based GHCND station data. In the study region, elevation has the most significant impact on precipitation. The WRF model accounts for elevation at a 4km spatial resolution, which is adequate to capture the main effects of elevation within the study region. In addition,the effect of fine-scale (1km) elevation is incorporated via topographical downscaling [31], adding further value to model data. There are numerous studies in the bias-correction literature that likewise treat point-based station and downscaled model data as equivalent (e.g [63], [66], [86]).

In the proposed workflow, historical WRF simulations (model output) were downscaled to a 1km grid using topographic downscaling, a variation of inverse distance weighting (IDW) that incorporates elevational lapse rates [31]. Elevation estimates at

each 1km grid cell were derived by interpolating elevation values from a 30m digital elevation model (DEM) [131] via IDW. The 1-km grid cell size was chosen based on resolution requirements for climate impacts modeling efforts over the Lake Champlain Basin [31], [132].

Prior to bias-correction, historical model data were interpolated to GHCND station locations via topographical downscaling for the purpose of constructing TFs. Downscaled raw WRF model data exhibited a wet bias that was most pronounced during summer months (Figure 3.2). This type of seasonal bias in WRF model simulations has also been found in other studies in the northeastern United States, e.g. [130]. To generate high-resolution, bias-corrected data products, bias-correction was applied to model data downscaled to the 1km grid. All performance metrics were calculated using model data topographically downscaled to station locations and GHCND station data.

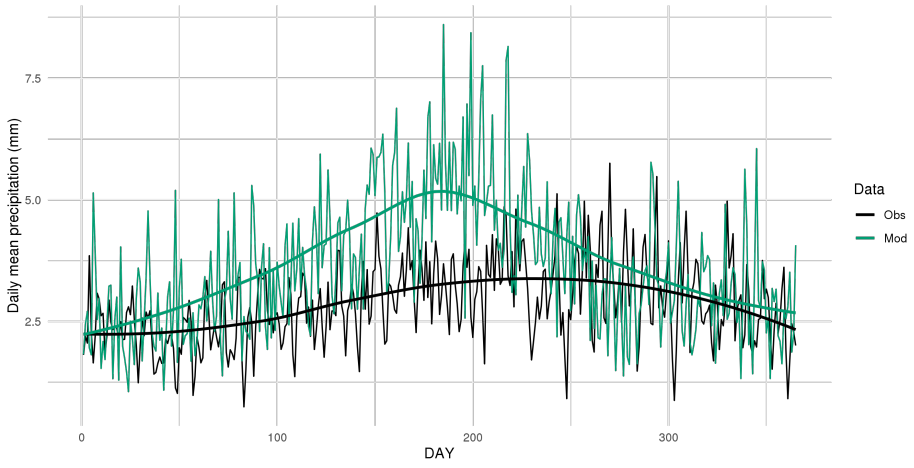

Figure 3.2: Mean daily precipitation (mm/day) for raw model (Mod) topographically downscaled to GHCND station locations and GHCND station data (Obs) with loess smoothers (smooth solid lines) overlaid. Daily means are calculated over the 85 GHCND station locations for years 1976-2005.

3.3.2 Bias-correction methods

The proposed approach, empirical quantile mapping with a linear correction for extremes, EQM-LIN, was compared to empirical quantile mapping (EQM), which is one of the most frequently-used and effective methods for bias-correction. In addition, we compared EQM-LIN to DM with the Gamma distribution (DM-GAMMA), a hybrid EQM approach in which lower quantiles were corrected using EQM, and upper quantiles were fit to Generalized Pareto Distributions (GPDs) (EQM-GPD) [106], as well as a trend-preserving method, quantile delta mapping (QDM) [47]. The results are presented in the Supplementary Material but not evaluated in the main manuscript, since none of the additional methods performed as well as or significantly better than EQM or EQM-LIN.

For both bias-correction methods EQM-LIN and EQM, TFs were constructed by spatially pooling GHCND station and model data downscaled to station locations. The same TF was applied to all model values, regardless of spatial location. We chose to spatially pool data because 1) much of the spatial variation in the data is due to elevation, which is accounted for during the downscaling procedure, 2) additional interpolation necessary to construct separate TFs based on spatial location would have added uncertainty to bias-corrected data. Spatially explicit bias-correction in general can be a difficult task and involves estimating the TF at every location at which bias-corrected data is desired [100], which is contrary to our desire to develop a bias-correction approach that is simple, efficient, and easily implemented.

For both bias-correction methods, twelve TFs were constructed, one for each month of the year [45], [53]. Daily raw model values interpolated to the 1km grid

were corrected with the corresponding monthly TF. Because GHCND station gauges are accurate to 0.1mm [74], we defined wet-day precipitation days as days in which daily precipitation was greater than or equal to 0.1mm. Prior to construction of TFs and bias-correction, values of daily $MOD_{interp,station}$ and model interpolations to the 1km grid below 0.1mm were set to 0. All analyses were conducted in R Statistical Language [139].

Empirical quantile mapping: EQM The TF used in EQM is expressed by the empirical cumulative distribution function (ecdf) and its inverse (ecdf^{-1}). Monthly TFs are of the form:

$$x_{corr,t} = \text{ecdf}_{obs}^{-1}(\text{ecdf}_{mod}(x_{mod,t})), \tag{3.1}$$

where, $x_{corr,t}$ is the corrected model precipitation value on day t, ecdf_{obs}^{-1} is the inverse ecdf of observed data, ecdf_{mod} is the ecdf of model data, and $x_{mod,t}$ is the raw model precipitation value on day t. Monthly TFs were constructed using 10,000 estimated quantiles, and interpolation of the TF was accomplished with monotone Hermite splines using the `qmap` package [102] in R. Values exceeding the range of the TF were corrected using the method of constant extrapolation [75]. The approximate shape of the TF can be examined by plotting estimated quantiles of model and observed data against one another to form a "quantile-quantile-" or "qq-" map (Figure 3.3). The shape of the quantile-quantile map can provide insight into the type and magnitude of model bias. For instance, if the TF falls below (rises above) the 1:1 line, model quantiles are too high (low) relative to observed quantiles.

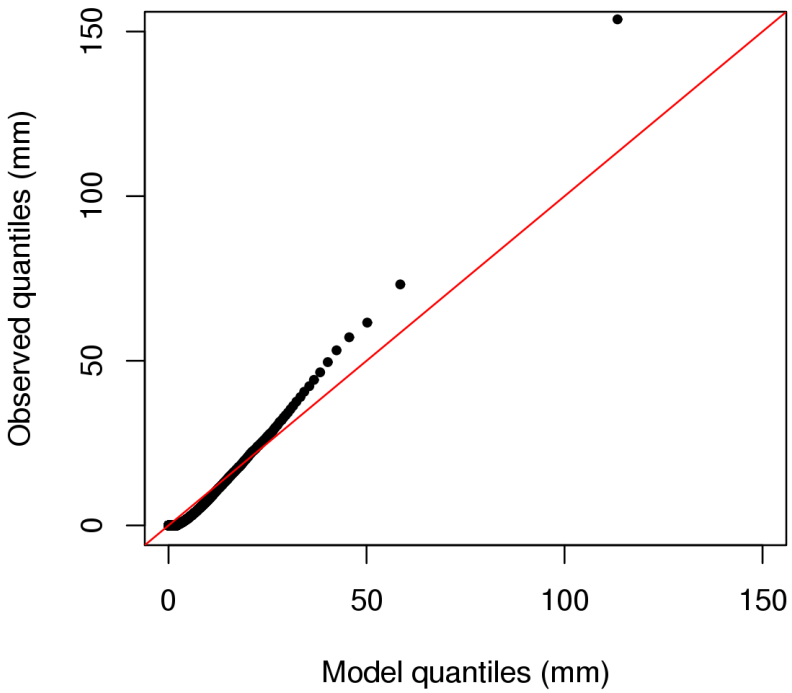

Figure 3.3: A quantile-quantile map for August constructed with 10,000 quantiles of model and observed data during the calibration period. The red solid line denotes the 1:1 line. Here, raw model data exhibits a low bias, especially at upper quantiles, as the qq-map lies above the 1:1 line.

Empirical quantile mapping with a linear correction for extremes: EQM-LIN

In EQM-LIN, the majority of model data are bias-corrected via EQM using (3.1), while model data beyond a specified threshold are adjusted with a constant correction via a linear TF (3.2). All bias-correction by EQM was done with the `qmap` package [102] in R, and custom code was used to construct the linear TF. The following steps describe the EQM-LIN procedure:

1. Calibration data is divided into two datasets in which model data is less than

(*CAL-LOW*) and greater than a specified threshold (*CAL-HIGH*). The threshold, T is a function of the inverse ecdf of model data and is expressed as $T = ecdf_{mod}^{-1}(\tau_{LIN})$, where $0 < \tau_{LIN} < 1$. Thus, T is a precipitation value in mm that indicates where both model and observed datasets are divided. The procedures for estimating T and τ_{LIN} are thoroughly outlined in the Appendix.

2. Next, the intercept for the linear TF, δ is obtained (details are discussed in the Appendix). The intercept represents the constant correction that will be applied to extreme model values (all model values in *CAL-HIGH*). The linear TF is expressed as $x_{corr,t} = \delta + x_{mod,t}$ and is applied to model values in *CAL-HIGH* (eq. 3.2). Model values in *CAL-LOW* are corrected via EQM. The TF for EQM-LIN is expressed as:

$$x_{corr,t} = \begin{cases} \text{ecdf}_{obs}^{-1}(\text{ecdf}_{mod}(x_{mod,t})), & x_{mod,t} < T \\ x_{mod,t} + \delta, & x_{mod,t} \geq T, \end{cases} \tag{3.2}$$

where $x_{corr,t}$ and $x_{mod,t}$ are as defined in (eq. 3.1). Thus, the linear portion of the TF ($x_{corr,t} = \delta + x_{mod,t}$) always has a slope of 1 and intercept δ.

In this study, we only consider linear TFs with a slope of 1 and intercept of δ. Optimizing the slope as well as the threshold would increase the overall complexity of EQM-LIN and could introduce the potential for overfitting on out-of-sample data.

We chose τ_{LIN} to be 0.79, based on a grid search over a range of values in a five fold cross-validation approach (details are discussed in the Appendix). We chose the value of τ_{LIN} that resulted in the minimization of the mean absolute error of observed and model ecdfs above the 95th percentile (MAE95), [170] (section 3.4).

MAE95 quantifies the distributional similarity between observed and model data at extremes. Since the focus of this study was on accurately representing distributional extremes, we chose the minimization of MAE95 rather than another metric. However, we found that minimization of MAE95 resulted in improvements in all performance metrics and indices.

The shape of the EQM-LIN TF is identical to that of EQM below T, while above the threshold the TF is linear. Figure 3.4 shows a quantile-quantile map for model and observed data for the month of August and the associated EQM and EQM-LIN TFs.

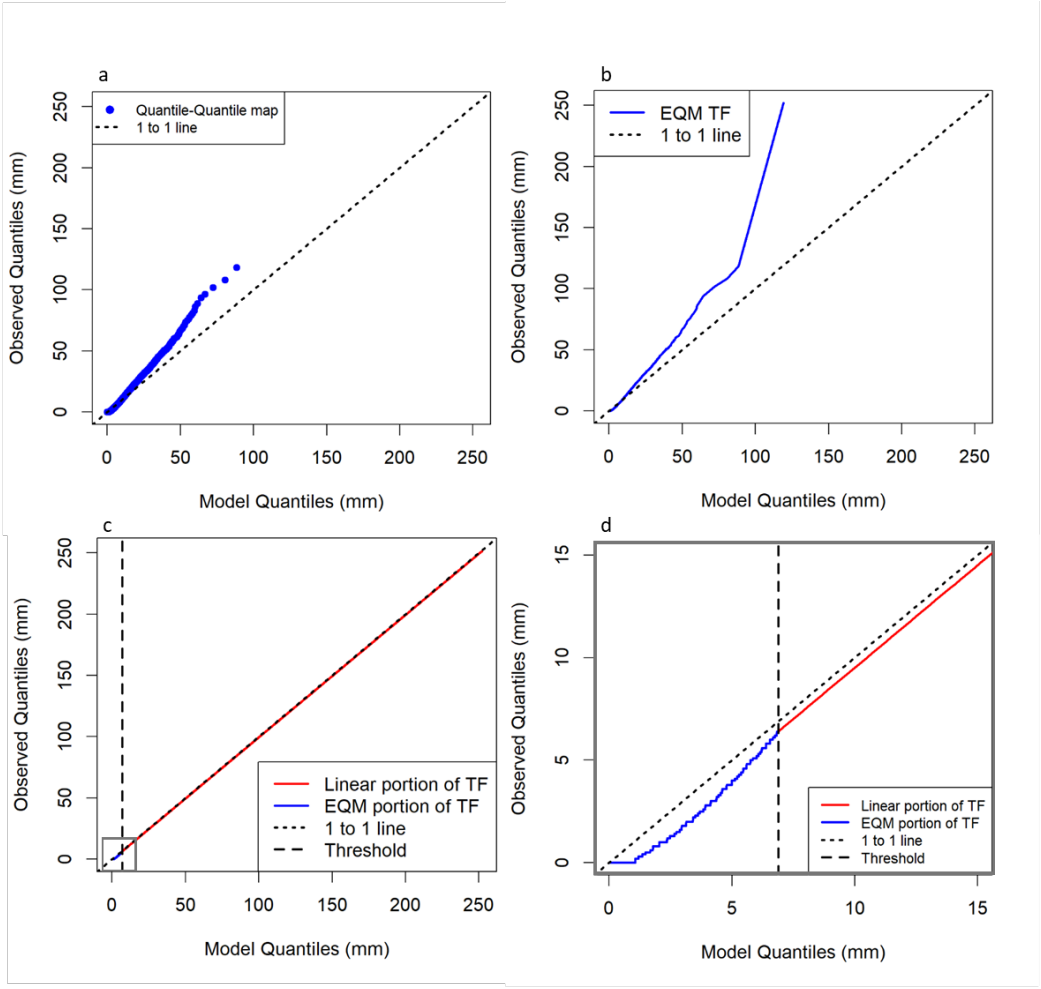

Figure 3.4: The quantile-quantile map and corresponding EQM and EQM-LIN TFs for daily observed and model data during the month of August over the calibration period 1976-2005. a) quantile-quantile map, constructed using 10,000 quantiles evenly spaced between 0 and 1; b) EQM TF ; c) EQM-LIN TF, with the blue line denoting the non-parametric (EQM) portion of the TF and the red line indicating the linear portion; d) enlarged section of EQM-LIN TF in c) (grey box) to illustrate the transition from EQM portion to the linear portion of the TF. In c) and d), the threshold (dashed line), indicates the 79th quantile of model data (6.88mm).

3.4 Validation

Performance evaluation of EQM and EQM-LIN was accomplished with a five-fold cross-validation procedure using observed and model data during the calibration period (1976-2005). Cross-validation is commonly used to evaluate the efficacy of bias-correction methods, as out-of-sample data can be considered proxies for future projections [53], [61], [106]. Test datasets always consisted of consecutive years. For example, if training data consisted of years 1976-2000, test data would contain years 2001-2005.

We chose performance metrics and indices that quantified 1) model skill and 2) the effectiveness of bias-correction methods in capturing overall climatology with an emphasis on extreme tails. All performance metrics were calculated using model data topographically downscaled to GHCND station locations and GHCND station data. Model skill, distributional similarity between model and observed data, was quantified with the mean absolute error (MAE). We chose MAE, rather than other skill metrics, such as the Perkins Skill Score [145], because it is more sensitive to outliers. Since TFs for EQM and EQM-LIN were constructed on a monthly basis, MAE metrics were evaluated by month. MAE was calculated between distributions of daily observed and raw model data as well as between distributions of daily observed and bias-corrected data at GHCND station locations for a given month [61]. MAE95 was used to quantify model skill at extreme tails. MAE95 is computed similarly to MAE, but only the upper 5% of daily observed and model distributions are used [170]. The number of quantiles estimated in the calculation of MAE95 was equal to the maximum number of 95th quantile values in observed or model distributions.

Generally, the number of values greater than the 95th quantile in each data type (raw model, observed, and corrected model) did not differ appreciably. MAE and MAE95 values were calculated by month for each of the five cross-validated data folds for raw and bias-corrected data, and results are reported as the average metric over the five folds. MAE and MAE95 quantify distributional error between model and observed data; lower values are indicative of better model skill, with an ideal mean absolute error of 0 (no error).

We used a subset of ETCCDI indices [171] to assess how well bias-corrected data captured overall climate characteristics of observed data. ETCCDI indices are standard indices that allow for the comparison of results over varying time periods, geographical regions, and source data, and are recommended by the World Research Climate Program (WRCP) [172]. ETTCDI indices were computed annually with spatially pooled data. Prior to calculating ETCCDI indices, downscaled raw model, bias-corrected model, and station data were averaged over the 85 station locations for each day in the 30-year calibration period (10950 days). Thirty annual values of each ETCCDI index were calculated for observed, raw model, and bias-corrected model data. The choice of indices was based on the preference of stakeholders.

'D' indices (D90, D95, and D99) are defined as the annual number of days in which mean daily precipitation exceeded the 90th, 95th, or 99th quantiles. 'S' metrics (S90, S95, and S99) are defined as the annual sum of mean daily precipitation (mm) for days in which mean daily precipitation exceeded the 90th, 95th, or 99th quantiles. TotalP is the annual sum of mean daily precipitation (mm) on wet days (days for which mean daily precipitation 0.1 mm), WetDays is the annual count of wet days, and the simple precipitation index (SPI) is calculated as TotalP/WetDays (mm/day).

SPI is a measure of precipitation intensity. The nine indices characterize the extreme tails, as well as general characteristics, of the 30-year climatology of precipitation. An overview of MAE metrics and ETCCDI indices is given in Table 3.1.

Performance evaluated by ETCCDI indices or MAE metrics cannot be directly compared, since each provides assessments on different temporal scales. MAE metrics quantify distributional errors of the entire distribution of daily model data compared to observed data. ETCCDI indices quantify how well model data capture 30-year climatology at a temporally coarser (annual) scale using spatially averaged data. In combination, both evaluation metrics give insight in the overall adequacy of the bias-correction method at both aggregated and finer temporal scales.

Table 3.1: Metric and index definitions

Metric/Index	Definition	Reference
MAE	Mean absolute error of quantiles of observed and raw model or bias-corrected model distributions. MAE is calculated using daily data (not spatially averaged) for the entire historical period using 10,000 estimated quantiles	[61]
MAE95	Mean absolute error of upper 5% of quantiles of observed and raw model or bias-corrected model distributions. MAE95 is calculated using daily data (not spatially averaged) for the entire historical period.	[65]
D90, D95, D99	Annual count of days for which mean daily precipitation exceeded the 90th 95th or 99th percentile	[173]
S90, S95, S99	Annual sum (mm) of mean daily precipitation on days in which mean daily precipitation exceeded the 90th , 95th, or 99th percentile.	[173]
TotalP	Annual sum (mm) of mean daily precipitation on days in which mean daily precipitation $\geq$ 0.1mm	[173]
WetDays	Annual count of days in which mean daily precipitation $\geq$ 0.1mm	[173]
SPI	Simple precipitation index (mm / day) calculated as TotalP / WetDays	[173]

3.4.1 Analyses

Bayesian one-way Analysis of Variance (ANOVA) models were used to determine if MAE and MAE95 differed significantly among raw model, EQM and EQM-LIN data. Separate ANOVA tests were conducted for MAE and MAE95. ANOVA tests were

conducted with data from all five cross-validated folds, as MAE and MAE95 values within folds can be considered subsamples. All analyses were conducted with the `RJags` package [174] in R. The response variables, MAE or MAE95 values, were log-transformed prior to analysis to ensure homogeneity of variances, an assumption of ANOVA models. The predictor variable for both ANOVA models was data type, a variable with three levels: raw model (Mod), EQM-LIN, and EQM. Credible intervals in the form of 95% highest posterior density (HPD) intervals were used to determine if the difference in posterior distributions was significantly different from 0. Credible intervals were constructed for all pairwise differences of posterior distributions of EQM-LIN, EQM, and raw model data. Credible intervals can be interpreted as follows: there is a 95% chance that the true pairwise difference in posterior distributions is contained within the interval, given the data. Therefore, if 0 is contained within the interval, the difference is not significant at the 95% confidence level. Full details on these analyses are provided in the Supplementary Materials.

Distributions of all nine ETCCDI indices calculated from EQM-, EQM-LIN-corrected, and raw model data were compared to those of observed data. Performance of bias-corrected and raw data relative to observed data was formally assessed using Kolmogorov-Smirnov (KS) tests [175]. The two-sample KS test is a non-parametric test that is used to assess the equality of two empirical distributions. It is sensitive to differences in both location and shape of the two ecdfs being compared and is often used in climatological studies [47], [176], [177]. Here, we applied the KS test three times for each ETCCDI index to determine the similarity of ecdfs between observed and EQM- and EQM-LIN-corrected data and between observed and raw model data. All tests were conducted with the two-sided null hypothesis that the samples being

compared belonged to a common distribution. The significance level, α, was set to 0.05; p-values below 0.05 indicate there is evidence that the two samples do not come from a common distribution. We acknowledge that the KS test has low power for small sample sizes (30 values or less) [178]. All KS tests in this study are performed on pairs of distributions composed of 30 annual values; thus, we use KS tests, along with visual inspection of boxplots, to guide our interpretation of results. In addition, to control for multiple comparisons, α was adjusted using the Holm-Bonferroni method [179] (details are shown in the Appendix).

3.5 Results

Overall, data bias-corrected with either EQM or EQM-LIN exhibited substantial improvements in both MAE and MAE95 compared to raw model data (Mod), but improvements were more pronounced for EQM-LIN. Both bias-correction methods generally improved ETCCDI indices compared to Mod, and EQM-LIN performed as well as or slightly better than EQM.

3.5.1 MAE and MAE95

MAEs of EQM- and EQM-LIN-corrected datasets were 0.704mm and 0.655mm, respectively, while MAE of Mod was 1.06mm (Figure 3.5 a). MAEs of both bias-corrected datasets were significantly lower than MAE of Mod. Monthly MAE values for EQM-LIN were overall slightly lower than those of EQM. The credible interval for the difference in MAE between EQM and EQM-LIN contained 0, indicating that although MAE of EQM-LIN was lower than that of EQM, the difference was not

significant at the 95% confidence level.

MAE95 values of EQM- and EQM-LIN-corrected model data and Mod were 3.45mm, 1.73mm, and 7.12mm, respectively. For EQM-LIN corrected data, MAE95 varied little among months; however, both raw model and EQM-corrected data exhibited substantial increases in MAE95 between months 8 and 11 (Figure 3.5 b). Similar to results for MAE, MAE95 values of both bias-corrected datasets were significantly lower than MAE95 of Mod. In contrast to results for MAE, 95% credible intervals for the difference in MAE95 of EQM and EQM-LIN indicated that MAE95 of EQM-LIN was significantly lower than MAE95 of EQM at the 95% confidence level (see Supplementary Materials for full details of ANOVA analysis).

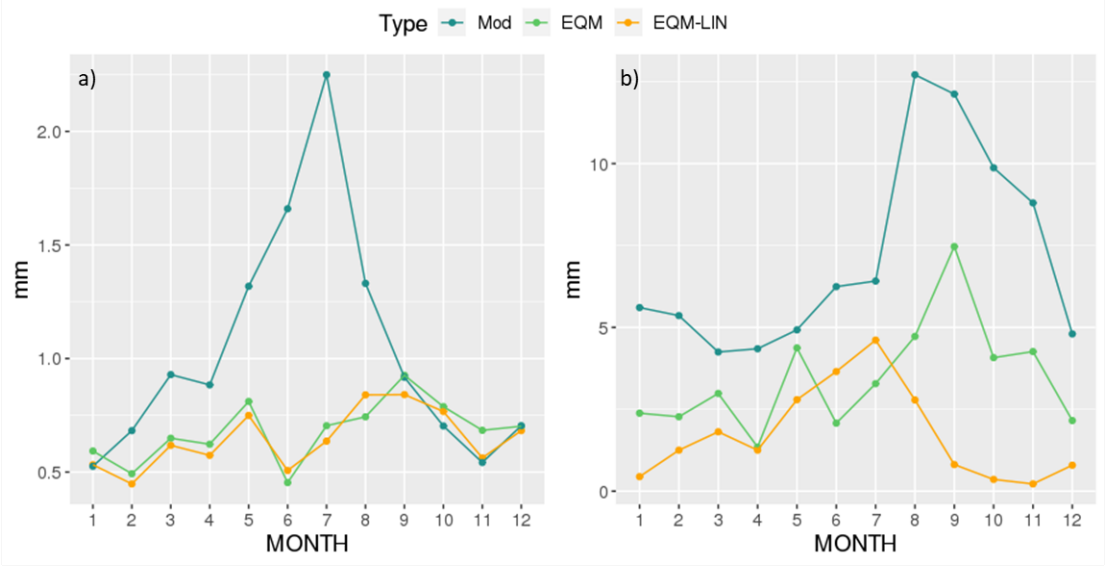

Figure 3.5: Monthly MAE (mm) (a) and MAE95 (mm) (b) for raw model (Mod), EQM- and EQM-LIN-corrected data. Please note the difference in y-axes limits for plots a and b.

3.5.2 ETCCDI indices

Distributions of ETCCDI indices for both bias-corrected datasets more closely resembled those of observed data compared to Mod, with EQM-LIN performing as good as or slightly better than EQM. Generally, mean and extreme total annual precipitation was overestimated in Mod, but Mod performed adequately in capturing extreme wet day frequency. While bias-correction resulted in the distributions of most ETCCDI indices becoming more similar to those of observed data, it also resulted in an underestimation of wet-day frequency. (See Appendix, Table 3.3 for selected summary statistics of ETTCDI index distributions for Mod, EQM, and EQM-LIN).

D and S indices Less extreme 'S' indices (S90 and S95) were substantially overestimated in Mod, and distributions of S90 and S95 calculated from Mod were significantly different from observed data (Figure 3.6 a; Table 3.2). The distribution of the more extreme S99 index was better represented in Mod and did not differ significantly from observed data. Both bias-correction methods provided minor improvements of the representation of S99 in Mod. Distributions of S90 and S95 were not significantly different for EQM-LIN as compared to observed data (Figure 3.6 a; Table 3.2). However, for EQM, the distribution of S95 differed significantly from that of observed data (Table 3.2). While both bias-correction methods were able to reduce the overestimation of total extreme annual rainfall exhibited in Mod, EQM-LIN slightly outperformed EQM.

Distributions of 'D' indices (D90, D95, and D99) were quite similar for Mod, bias-corrected, and observed data (Figure 3.6 b). P-values of KS tests for D90, D95, and D99 confirmed that distributions of Mod and bias-corrected data were not significantly

different from observed data (Table 3.2). These results show that the frequency of extreme precipitation days, D90, D95 and D99, are adequately represented in Mod and that bias-correction via either method does not adversely affect the representation of 'D' indices.

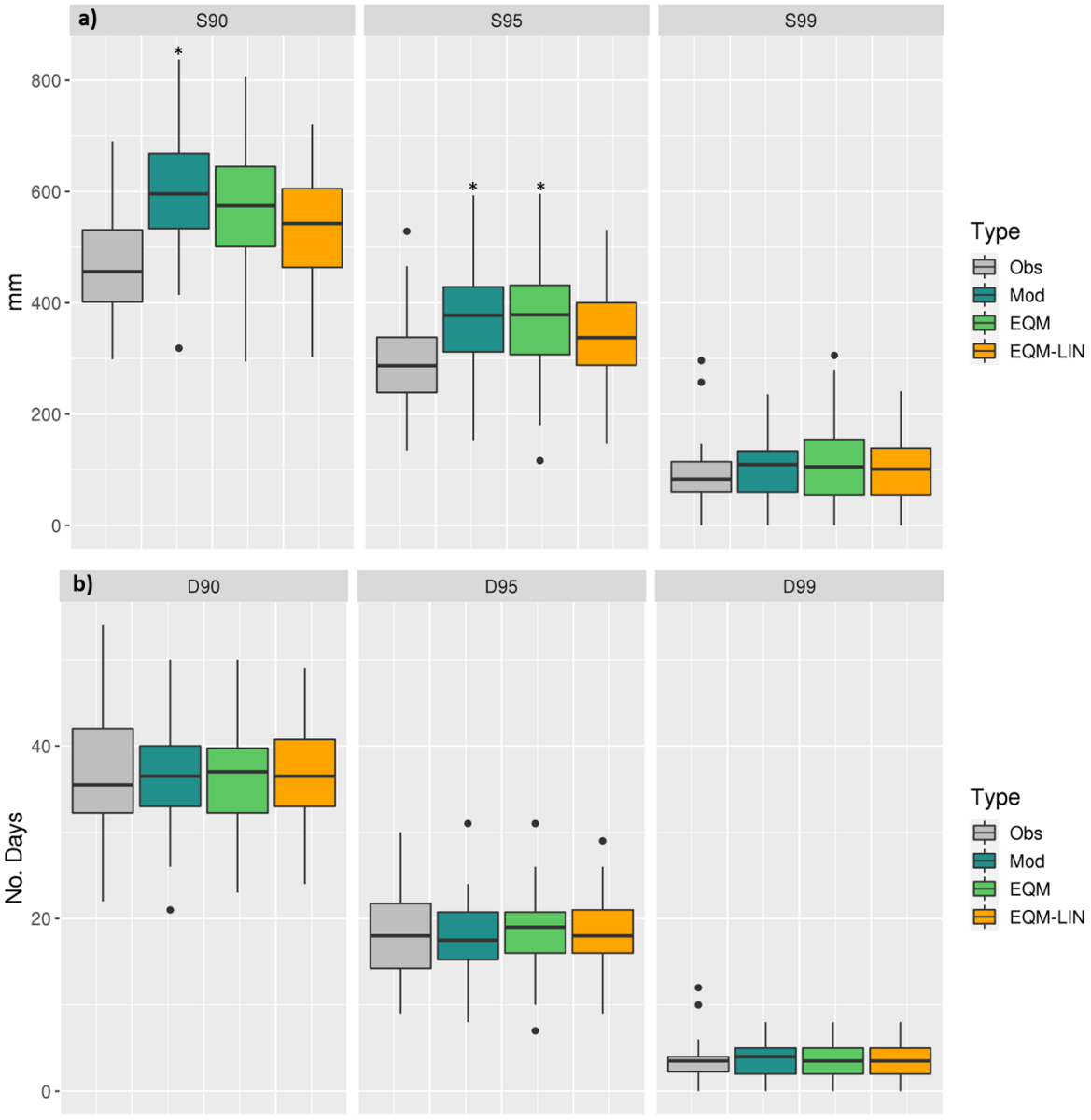

Figure 3.6: Boxplots of a) D90, D95, and D99 and b) S90, S95, and S99 for observed (Obs), model (Mod), EQM-, and EQM-LIN-corrected data. Each boxplot represents 30 values (ETCCDI indices are calculated annually). Significance of KS-tests of distributional similarity of Mod, EQM, or EQM-LIN compared to Obs at $\alpha = 0.05$, adjusted with the Holm-Bonferroni method, are indicated with (*).

TotalP, WetDays, and SPI TotalP was significantly overestimated in Mod ($p < 0.0001$), but distributions of TotalP calculated using either bias-corrected dataset were not significantly different from observed data ($p = 0.81$) (Figure 3.7; Table 3.2). Thus, both bias-correction methods were highly effective in correcting total annual precipitation.

The distribution of WetDays derived from Mod did not differ significantly from observed data ($p = 0.13$) (Table 3.2). However, WetDay distributions calculated from EQM- and EQM-LIN-corrected data were significantly underestimated relative to observed data ($p < 0.0001$) (Figure 3.7; Table 3.2). SPI was overestimated by Mod, due to the large overestimation of Total P; SPI was overestimated to a lesser degree, by EQM- and EQM-LIN-corrected data due to the underestimation of WetDays (Figure 3.7). Distributions of SPI calculated from EQM, EQM-LIN, and Mod all differed significantly from observed data (Table 3.2). Although bias-correction via either EQM-LIN or EQM results in underestimating WetDays, annual precipitation totals (TotalP) are effectively corrected. Moreover, while the distribution of WetDays is adequately represented in Mod, Mod contains an excessive number of low-precipitation occurrences relative to observed data (see Supplementary Materials, section 4). However, despite the underestimation of wet day frequency following bias-correction, precipitation intensity (SPI) is slightly improved compared to raw model data.

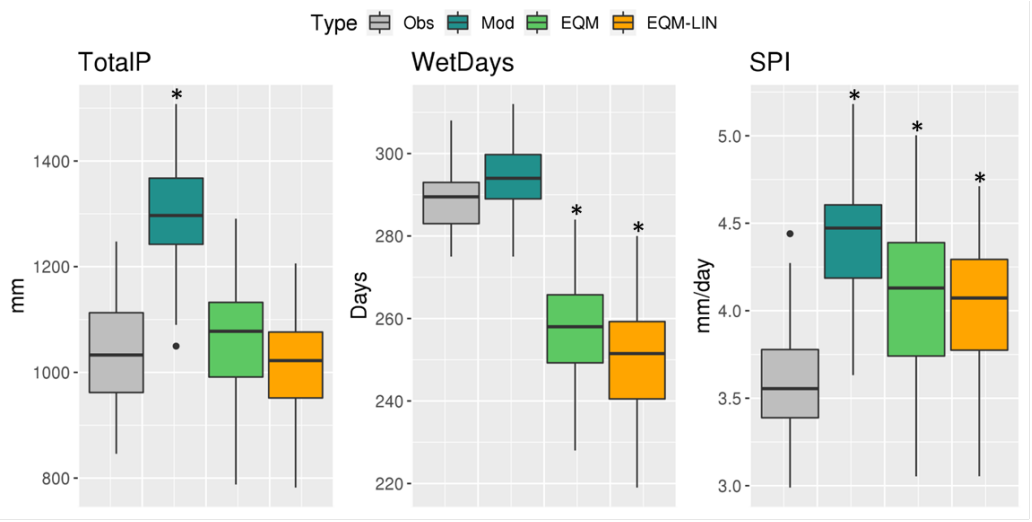

Figure 3.7: Boxplots of TotalP, WetDays, and SPI for observed (Obs), model (Mod), EQM-, and EQM-LIN-corrected data. Each bxplot represents 30 values (ETCCDI indices are calculated annually). Significance of KS-tests of distributional similarity of Mod, EQM, or EQM-LIN compared to Obs at $\alpha = 0.05$*, adjusted with the Holm-Bonferroni method, are indicated with (*).*

Table 3.2: Two-sample Kolmogorov-Smirnov (KS) test results for raw model (Mod), EQM-, and EQM-LIN-corrected distributions of ETCCDI indices compared to observed distributions of ETCCDI indices. D is the KS test statistic. P-values refer to a two-sided null hypothesis; p-values < 0.05 indicate that the distribution of a particular ETCCDI index for either Mod, EQM-LIN or EQM is significantly different from that of observed data at the 5% significance level. All ETCCDI index distributions consisted of 30 annual values. Significance of KS-tests at $\alpha = 0.05$, adjusted with the Holm-Bonferroni method, are indicated with (*).

S90	**D**	**p**
Mod	0.73	<0.0001*
EQM-LIN	0.33	0.07
EQM	0.40	0.02
S95	**D**	**p**
Mod	0.43	0.007*
EQM-LIN	0.33	0.007*
EQM	0.40	0.02
S99	**D**	**p**
Mod	0.30	.13
EQM-LIN	0.23	0.39
EQM	0.33	0.07
D90	**D**	**p**
Mod	0.17	0.80
EQM-LIN	0.13	0.95
EQM	0.17	0.80
D95	**D**	**p**
Mod	0.17	0.80
EQM-LIN	0.13	0.95
EQM	0.17	0.80
D99	**D**	**p**
Mod	0.10	1
EQM-LIN	0.10	1
EQM	0.10	1
TotalP	**D**	**p**
Mod	0.73	<0.0001*
EQM-LIN	0.17	0.808
EQM	0.17	0.808
WetDays	**D**	**p**
Mod	0.30	0.13
EQM-LIN	0.97	<0.0001*
EQM	0.90	<0.0001*
SPI	**D**	**p**
Mod	0.73	<0.0001*
EQM-LIN	0.53	0.0003*
EQM	0.60	<0.0001*

3.6 Discussion

Local-scale modeling efforts in hydrology, ecology, agriculture, and economics, as well as climate impact assessments, require high-resolution climate products. Since climate extremes exert a large influence on humans and the environment, it is crucial that extremes are accurately represented in climate products. An effective way to obtain high-resolution climate products is to statistically downscale and bias-correct dynamically downscaled output from an RCM. Bias-correction of precipitation extremes, in particular, is a difficult task. In this study, we developed a hybrid bias-correction method, EQM-LIN, that combines the efficacy of EQM for bias-correcting the bulk of raw model data, with a robust linear adjustment for correcting distributional tails. We found that EQM-LIN results in the accurate representation of mean and extreme precipitation. EQM-LIN outperformed EQM in terms of model skill (MAE and MAE95) and performed at least as well or better than EQM with respect to most ETCCDI climatological indices. Furthermore, our study indicates that a linear correction, as implemented in EQM-LIN, is resistant to overfitting and results in a more robust TF at higher quantiles, both of which can decrease uncertainty in bias-corrected data.

The substantial difference in performance between EQM-LIN and EQM with respect to model skill is due to the different ways in which TFs are constructed at extreme tails. In EQM, distributional tails are corrected with a flexible TF that closely interpolates the quantile-quantile map of raw and observed data. However, since data at extreme tails is, by definition, scarce and variable, the TF produced by EQM may be unstable and can result in a faulty correction on out-of-sample model

data [47], [144]. In our study, MAE95 values of EQM increased markedly between months 8 and 11, reaching a maximum in month 9, while those of EQM-LIN remained near 2.5mm (Figure 3.5 b). An inspection of training and testing datasets used during cross-validation reveals that often, the association between raw model and observed quantiles (the quantile-quantile map) was quite different between training and corresponding testing datasets. In such cases, EQM tended to overfit on training data, and consequently, the correction applied to testing data was unsuitable.

Figure 3.9 depicts such a scenario for month 9, when the difference in MAE95 between the two bias-correction methods was large. In Figure 3.9, the EQM TF constructed with training data (black dots) extends non-linearly above the one-to-one line and then increases sharply. The shape of the training TF indicates that, generally, raw model quantiles are too low relative to those of observed data. When the training TF is applied to test data, raw model values in the tails, especially, are increased. For instance, a raw model value of 58.6mm would be corrected to 81.8mm (Figure 3.9). However, the relationship between raw model and observed quantiles in the test data (blue dots), indicates that raw model quantiles are only slightly too high compared to observed quantiles (Figure 3.9). When raw model data in the test set are bias-corrected with the training TF, raw model values are increased too much relative to observed values (Figure 3.9). The quantile-quantile map of *corrected* model quantiles and observed quantiles (which should lie near or on the one-to-one line if the correction was satisfactory) is shifted far to the right of one-to-one line, indicating that corrected model values, especially in the tails, are too high. This example shows that the flexibility of EQM is also what makes it susceptible to overfitting on calibration data and supports other studies showing that EQM is sensitive to the choice of, and

can overfit on calibration data [8], [45], [65], [100], [144].

For the same scenario, EQM-LIN produces a linear TF at extreme tails with a slope of 1 and an intercept of δ (the constant correction factor) (Figure 3.10). Raw model values are adjusted by a constant, δ. Though this approach is less flexible than that of EQM, it produces more stable TFs and is less sensitive to training data. In Figure 3.10, the training TF for EQM-LIN (black dots) is linear and does not exhibit the fluctuations apparent in the training TF of EQM (Figure 3.9). The intercept (δ) of the TF in Figure 3.10 is slightly less than zero, which means that raw model values will be decreased by δ. The TF for EQM-LIN represents an appropriate correction, as model quantiles in the test dataset are, in fact, too high relative to observed quantiles (Figure 3.10, blue dots). For instance, the TF of EQM-LIN corrects a raw model value of 58.6mm to 58.1mm (Figure 3.10). Accordingly, the quantile-quantile map of corrected model quantiles and observed quantiles is close to the one-to-one line, indicating a satisfactory correction.

Figures 3.9 and 3.10 are representative of scenarios in which the relationship between raw model and observed quantiles differ between training and testing data and highlight the differences in bias-correction between EQM and EQM-LIN. In our study area, such scenarios are common in months when precipitation is variable and when extreme precipitation events are more likely (months 6-9). The difference in bias-correction between EQM-LIN and EQM can also be seen visually in downscaled, bias-corrected data over the study region. Figure 3.8 shows raw, downscaled, and corrected and downscaled precipitation data for one day in which mean precipitation exceeded the 95th percentile (September 12, 1986). Note that in Figure 3.8, EQM results in an increase of high precipitation values (bright pink regions), while EQM-LIN

results in a slight dampening of precipitation in the same regions. A model precipitation value of 52.14mm is transformed to 68.25mm using EQM and 51.28mm using EQM-LIN. The increase and dampening effects of EQM and EQM-LIN, respectively, in Figure 3.8 are a result of differences in EQM and EQM-LIN transfer functions.

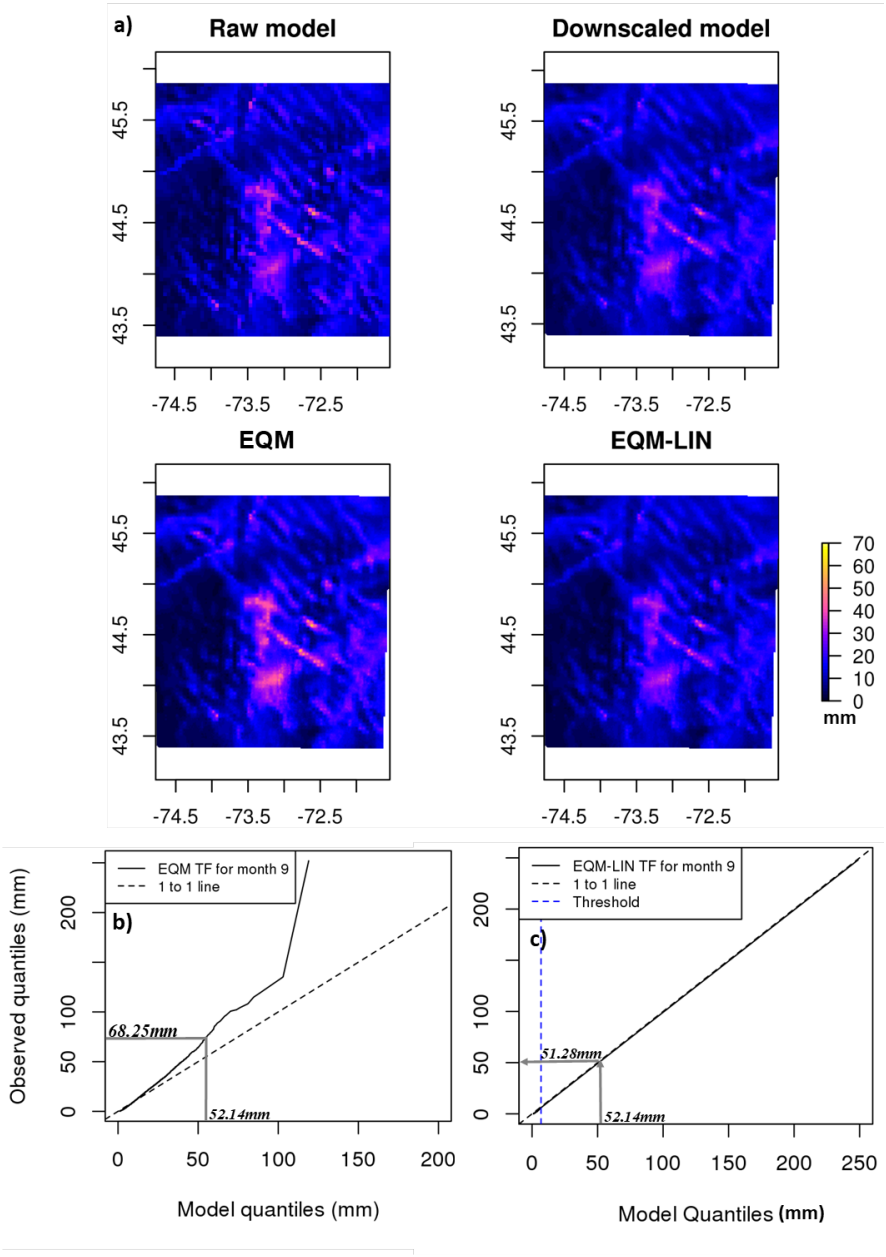

Figure 3.8: Raw model, downscaled raw model, and bias-corrected data for one day (September 12, 1986) (a) with corresponding TFs for EQM (b) and EQM-LIN (c). Plot (a) shows raw model (4km grid), downscaled raw model (1km grid), and downscaled and bias-corrected precipitation data (mm) for a day on which daily mean precipitation exceeded the 95th quantile (September 12, 1986). Plots (b) and (c) show the corresponding EQM and EQM-LIN TFs, respectively; in (b) and (c), gray lines indicate how EQM and EQM-LIN adjust the maximum model precipitation value for this day (52.14 mm) as an example. This figure is intended to provide an example of how the correction of raw model data differs between EQM and EQM-LIN.

Though EQM-LIN significantly outperformed EQM in terms of model skill (MAE and MAE95), results were not as dramatic for climatological (ETCCDI) indices. ETCCDI indices are calculated using spatially averaged, daily data, which reduces variation and may explain the similarity in performance of EQM and EQM-LIN for ETCCDI indices. Bias-correction via both EQM and EQM-LIN resulted in improvements over raw data for most indices. Though both bias-correction methods improved the overestimation of total annual mean precipitation (TotalP) as well as total extreme annual precipitation (Sum90) exhibited in raw model data, EQM-LIN performed slightly better than EQM for moderate extremes (Sum95). Raw model data adequately captured higher extremes (D99, S99); bias-correction provided slight improvement in the representation of S99.

Interestingly, the distribution of raw model wet day frequency (WetDays) was similar to that of observed data, while bias-correction via either method resulted in considerable underestimation of wet day frequency. The negative impact of bias-correction on wet day frequency is most likely due to the excessive number of low-precipitation occurrences ("drizzle effect") [44], [162] in raw model data. EQM, which is used to correct low-valued quantiles in both bias-correction methods, results in the majority of excessive low-precipitation days being set to zero. The underestimation of wet day frequency after bias-correction via EQM is not unusual; similar results were found by [98] and [180]. Moreover, although wet-day frequency appears to be adequately represented in raw model data, it comes at the expense of substantial overestimation of total annual precipitation (TotalP) and precipitation intensity (SPI). After bias-correction via either method, precipitation intensity is better represented, and the distribution of total annual precipitation is very close to that of observed

data. Thus, for most climatological indices, bias-correction via either method provides critical improvements to raw model data, especially with respect to extremes.

Conclusion

In this study, we show that a hybrid EQM approach for bias-correction (EQM-LIN), in which the majority of model data is corrected via EQM and extreme tails are corrected by a linear TF, resists overfitting on calibration data, increases overall and model skill, especially at extreme tails, and results in a better representation of climatological indices compared to conventional EQM. Our method is simple, intuitive, and easy to implement, making it a suitable alternative to EQM for bias-correcting historical and future climate simulations. Though we apply the method to precipitation data, we expect it could be applied to other climate variables as well. Future work might include adjusting the slope of the linear correction or using another function to construct the TF at extreme tails.

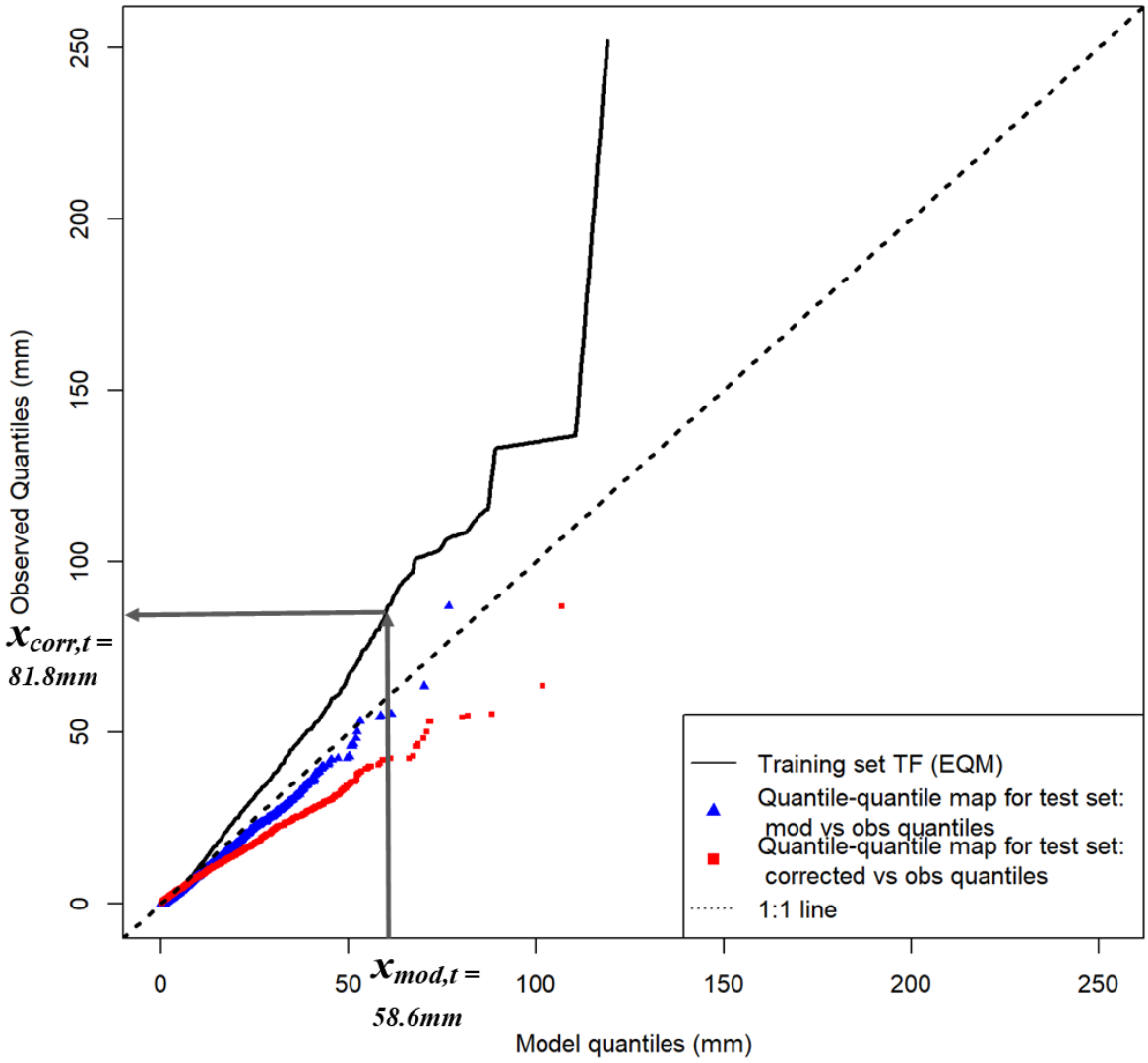

Figure 3.9: Construction of the EQM TF in a train-test scenario; data for this plot reflect one particular train-test fold used during cross-validation for month 9 (September). The TF obtained from training data is shown in black. The quantile-quantile map of model and observed data in the test set is shown in blue. The corrected quantile-quantile map (quantiles of corrected model data versus quantiles of observed data) in the test set are shown in red. $x_{mod,t}$ and $x_{corr,t}$ denote model and corrected model values, respectively, for day t. Gray arrows indicate how model data in the test set is corrected, based on the TF from training data.

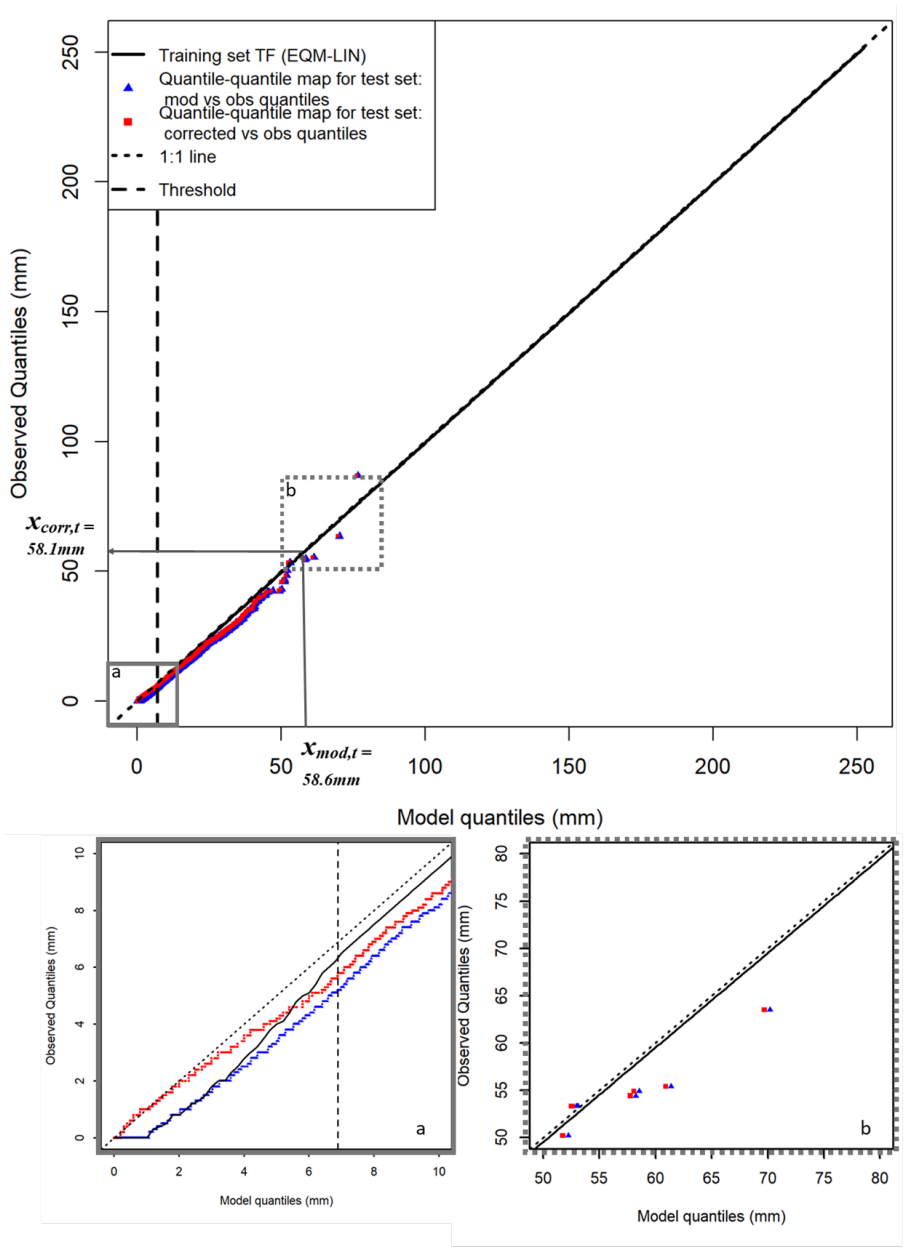

Figure 3.10: Construction of the EQM-LIN TF in a train-test scenario; data for this plot reflect one particular train-test fold used during cross-validation for month 9 (September). The TF obtained from training data is shown in black. The quantile-quantile map of model and observed data in the test set is shown in blue. The corrected quantile-quantile map (quantiles of corrected model data versus quantiles of observed data) in the test set are shown in red. $x_{mod,t}$ and $x_{corr,t}$ denote raw model and corrected model values, respectively, for day t. Gray arrows indicate how raw model data in the test set is corrected, based on the training-set TF. The threshold (dashed line), indicates the 79th quantile of model data (6.88mm). For ease of viewing, plot a) (gray box) shows the scenario at selected lower (0-10 mm) precipitation quantiles, and plot b) (gray dotted box) shows the scenario at selected extreme (50-80mm) precipitation quantiles.

3.7 Appendix

3.7.1 Estimating the threshold T and intercept δ

The first step for obtaining the threshold T is to estimate τ_{LIN} from the data. We chose τ_{LIN} to be 0.79, based on a grid search over a range of values in a five fold cross-validation approach. We chose the value of τ_{LIN} that resulted in the minimization of the mean absolute error of observed and model ecdfs above the 95th percentile (MAE95), [170] (section 3.4). It is crucial that τ_{LIN} be estimated using cross-validation; our result of $\tau_{LIN} = 0.79$ may not generalize to all data.

To obtain T, we must assume a fixed value of τ_{LIN}. The next steps involves the construction of ecdfs for observed and model data in the calibration period. Ecdfs are constructed using 10,000 quantiles evenly spaced between 0 and 1. Next, the threshold, T is computed as $ecdf_{mod}^{-1}(\tau_{LIN})$. Note that, T is the model precipitation value in mm corresponding to the quantile τ_{LIN} (whereas $0 \leq \tau_{LIN} \leq 1$).

To obtain δ, we assume that T has been calculated. Ecdfs of observed and model data are constructed using 10,000 quantiles evenly spaced between 0 and 1. Values in the ecdfs of model and observed data are sorted in increasing order. Note the rank of T within the sorted precipitation values of the model ecdf; the rank value will be denoted as R_T. For example, suppose $T = 12mm$ and the rank of T within the ecdf of model data is 5,000, then $R_T = 5000$.

Next, select the precipitation value from sorted, observed ecdf at rank R_T and denote this value as T_{obs}. The intercept of the linear TF, δ which represents the

constant correction, is calculated as the difference $T_{obs} - T$. Continuing with the example, suppose $T_{obs} = 9.1mm$; then $\delta = 9.1 - 12 = -2.9$. This means model extremes (all values $\geq T$) will be decreased by 2.9mm.

The constant correction at extremes, δ, is similar to the constant extrapolation correction used by [75]. However, here, the constant correction is the difference $T - T_{obs}$, whereas in [75], it is $ecdf_{obs}^{-1}(1.00) - ecdf_{mod}^{-1}(1.00)$.

3.7.2 KS Test

The KS test statistic, D is computed as

$$(D_n = \underset{x}{sup}|F_n(x) - G_n(x)|). \tag{3.3}$$

In (3.3), F_n and G_n are the two ecdfs being compared, n denotes the number of independent and identically distributed ordered values used to obtain F_n and G_n, and $\underset{x}{sup}$ is the supremum of the collection of n distances.

3.7.3 Holm-Bonferroni method for multiple comparisons

When multiple statistical comparisons are made, it is often necessary to adjust the Type I error rate (commonly referred to as the significance level or α). The Type I error rate is the probability of falsely rejecting the null hypotheses when it is, in fact, true (a false positive). In the context of multiple hypothesis testing, it is often desirable to adjust the family-wise error rate (FWER), the probability of rejecting

one null hypothesis in m hypothesis tests. The Holm-Bonferroni method is suitable when a less conservative adjustment of the FWER is preferred.

Suppose m hypothesis tests have been conducted, and m p-values have been calculated. The Holm-Bonferroni adjustment for the FWER involves two steps:

1. Order p-values from least to greatest and assign each p-value a rank from 1 to k, $k = 1 \dots m$

2. Find the smallest p-value such that $p_k < \frac{\alpha}{m+1-k}$.

If the condition in step 2 is true, the p-value is significant; if the condition in step 2 if false, the p-value is not significant.

3.7.4 Summary results for ETCCDI indices

Table 3.3 shows the 25th, 50th, and 75th quantiles for each data type (Mod, EQM, and EQM-LIN) and ETCCDI index.

Table 3.3: 25th (Q25), 50th (Q50), and 75th (Q75) quantiles of ETCCDI indices for observed data (Obs), raw model data (Mod), and EQM-, and EQM-LIN-corrected data during the calibration period (1976-2005). Each ETCCDI index was calculated using 30 annual values.

Data type	Q25	Q50	Q75
	Sum90		
Obs	401.63	456.20	531.18
Mod	533.62	595.78	668.31
EQM	501.05	574.38	645.14
EQM-LIN	463.71	542.41	605.23
	Sum95		
Obs	239.03	287.21	338.08
Mod	311.85	377.51	428.63
EQM	307.08	378.51	431.53
EQM-LIN	287.94	337.40	400.09
	Sum99		
Obs	60.29	83.17	114.33
Mod	311.85	377.51	428.63
EQM	55.09	105.16	154.46
EQM-LIN	55.10	101.02	138.54
	D90		
Obs	32.25	35.50	42.00
Mod	33.00	36.50	40.00
EQM	32.25	37.00	39.75
EQM-LIN	32.25	37.00	39.75
	D95		
Obs	14.25	18.00	21.75
Mod	15.25	17.50	20.75
EQM	16.00	19.00	20.75
EQM-LIN	16.00	18.00	21.00
	D99		
Obs	2.25	3.50	4.00
Mod	15.25	17.50	20.75
EQM	2.00	3.50	5.00
EQM-LIN	2.00	3.50	5.00
	TotalP		
Obs	961.92	1032.81	1112.80
Mod	1242.30	1296.84	1367.50
EQM	991.24	1077.74	1132.50
EQM-LIN	951.56	1022.26	1076.33
	WetDays		
Obs	283.00	289.50	293.00
Mod	289.00	294.00	299.75
EQM	249.25	258.00	265.75
EQM-LIN	240.50	251.50	259.25
	SPI		
Obs	3.39	3.55	3.78
Mod	4.19	4.47	4.61
EQM	3.74	4.13	4.34
EQM-LIN	3.77	4.07	4.29

References

[4] A. Zia, A. Bomblies, A. W. Schroth, *et al.*, "Coupled impacts of climate and land use change across a river–lake continuum: Insights from an integrated assessment model of lake champlain's missisquoi basin, 2000–2040," *Environmental Research Letters*, vol. 11, no. 11, p. 114 026, 2016.

[5] M. Ekström, M. R. Grose, and P. H. Whetton, "An appraisal of downscaling methods used in climate change research," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 6, no. 3, pp. 301–319, 2015.

[6] L. R. Leung, L. O. Mearns, F. Giorgi, and R. L. Wilby, "Regional climate research: Needs and opportunities," *Bulletin of the American Meteorological Society*, vol. 84, no. 1, pp. 89–95, 2003.

[7] D. Maraun, T. G. Shepherd, M. Widmann, *et al.*, "Towards process-informed bias correction of climate change simulations," *Nature Climate Change*, vol. 7, no. 11, p. 764, 2017.

[8] T. Lafon, S. Dadson, G. Buys, and C. Prudhomme, "Bias correction of daily precipitation simulated by a regional climate model: A comparison of methods," *International Journal of Climatology*, vol. 33, no. 6, pp. 1367–1381, 2013.

[9] A. J. Cannon, C. Piani, and S. Sippel, "Bias correction of climate model output for impact models," in *Climate Extremes and Their Implications for Impact and Risk Assessment*, Elsevier, 2020, pp. 77–104.

[10] I. Hanssen-Bauer, C. Achberger, R. Benestad, D. Chen, and E. Førland, "Statistical downscaling of climate scenarios over Scandinavia," *Climate Research*, vol. 29, no. 3, pp. 255–268, 2005.

[11] Z. A. Holden, J. T. Abatzoglou, C. H. Luce, and L. S. Baggett, "Empirical downscaling of daily minimum air temperature at very fine resolutions in complex terrain," *Agricultural and Forest Meteorology*, vol. 151, no. 8, pp. 1066–1073, 2011.

[13] P. Caldwell, H.-N. S. Chin, D. C. Bader, and G. Bala, "Evaluation of a WRF dynamical downscaling simulation over California," *Climatic change*, vol. 95, no. 3-4, pp. 499–521, 2009.

[15] F. Feser, B. Rockel, H. von Storch, J. Winterfeldt, and M. Zahn, "Regional climate models add value to global model data: A review and selected examples," *Bulletin of the American Meteorological Society*, vol. 92, no. 9, pp. 1181–1192, 2011.

[17] X. Gao, J. S. Pal, and F. Giorgi, "Projected changes in mean and extreme precipitation over the mediterranean region from a high resolution double nested rcm simulation," *Geophysical Research Letters*, vol. 33, no. 3, 2006.

[18] D. Maraun, "Bias correcting climate change simulations-a critical review," *Current Climate Change Reports*, vol. 2, no. 4, pp. 211–220, 2016.

[21] L. Mearns, F. Giorgi, P. Whetton, D. Pabon, M. Hulme, and M. Lal, "Guidelines for use of climate scenarios developed from regional climate model experiments," *Data Distribution Centre of the Intergovernmental Panel on Climate Change*, 2003.

[22] G. Fang, J. Yang, Y. Chen, and C. Zammit, "Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in china," *Hydrology and Earth System Sciences*, vol. 19, no. 6, pp. 2547–2559, 2015.

[31] J. M. Winter, B. Beckage, G. Bucini, R. M. Horton, and P. J. Clemins, "Development and evaluation of high-resolution climate simulations over the mountainous northeastern united states," *Journal of Hydrometeorology*, vol. 17, no. 3, pp. 881–896, 2016.

[44] R. Leander and T. A. Buishand, "Resampling of regional climate model output for the simulation of extreme river flows," *Journal of Hydrology*, vol. 332, no. 3-4, pp. 487–496, 2007.

[45] C. Piani, J. Haerter, and E. Coppola, "Statistical bias correction for daily precipitation in regional climate models over europe," *Theoretical and Applied Climatology*, vol. 99, no. 1-2, pp. 187–192, 2010.

[46] C. Teutschbein and J. Seibert, "Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods," *Journal of Hydrology*, vol. 456, pp. 12–29, 2012.

[47] A. J. Cannon, S. R. Sobie, and T. Q. Murdock, "Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?" *Journal of Climate*, vol. 28, no. 17, pp. 6938–6959, 2015.

[48] A. W. Wood, E. P. Maurer, A. Kumar, and D. P. Lettenmaier, "Long-range experimental hydrologic forecasting for the eastern united states," *Journal of Geophysical Research: Atmospheres*, vol. 107, no. D20, ACL–6, 2002.

[53] M. Jakob Themeßl, A. Gobiet, and A. Leuprecht, "Empirical-statistical downscaling and error correction of daily precipitation from regional climate models," *International Journal of Climatology*, vol. 31, no. 10, pp. 1530–1544, 2011.

[54] M. Shrestha, S. C. Acharya, and P. K. Shrestha, "Bias correction of climate models for hydrological modelling–are simple methods still useful?" *Meteorological Applications*, vol. 24, no. 3, pp. 531–539, 2017.

[55] D. R. Roberts, W. H. Wood, and S. J. Marshall, "Assessments of downscaled climate data with a high-resolution weather station network reveal consistent but predictable bias," *International Journal of Climatology*, vol. 39, no. 6, pp. 3091–3103, 2019.

[61] L. Gudmundsson, J. Bremnes, J. Haugen, and T. Engen-Skaugen, "Downscaling rcm precipitation to the station scale using statistical transformations–a comparison of methods," *Hydrology and Earth System Sciences*, vol. 16, no. 9, pp. 3383–3390, 2012.

[63] J.-H. Heo, H. Ahn, J.-Y. Shin, T. R. Kjeldsen, and C. Jeong, "Probability distributions for a quantile mapping technique for a bias correction of precipitation data: A case study to precipitation data under climate change," *Water*, vol. 11, no. 7, p. 1475, 2019.

[65] P. Reiter, O. Gutjahr, L. Schefczyk, G. Heinemann, and M. Casper, "Does applying quantile mapping to subsamples improve the bias correction of daily precipitation?" *International Journal of Climatology*, vol. 38, no. 4, pp. 1623–1633, 2018.

[66] O. Gutjahr and G. Heinemann, "Comparing precipitation bias correction methods for high-resolution regional climate simulations using cosmo-clm," *Theoretical and applied climatology*, vol. 114, no. 3, pp. 511–529, 2013.

[67] M. A. Ivanov and S. Kotlarski, "Assessing distribution-based climate model bias correction methods over an alpine domain: Added value and limitations," *International Journal of Climatology*, vol. 37, no. 5, pp. 2633–2653, 2017.

[69] J.-Y. Shin, T. Lee, T. Park, and S. Kim, "Bias correction of rcm outputs using mixture distributions under multiple extreme weather influences," *Theoretical and Applied Climatology*, vol. 137, no. 1, pp. 201–216, 2019.

[73] C. Miao, L. Su, Q. Sun, and Q. Duan, "A nonstationary bias-correction technique to remove bias in gcm simulations," *Journal of Geophysical Research: Atmospheres*, vol. 121, no. 10, pp. 5718–5735, 2016.

[74] N. Oceanic and A. Administration, *Climate data online search*, Accessed: 2017-09-30, 2018. [Online]. Available: https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND.

[75] J. Boé, L. Terray, F. Habets, and E. Martin, "Statistical and dynamical downscaling of the seine basin climate for hydro-meteorological studies," *International Journal of Climatology: A Journal of the Royal Meteorological Society*, vol. 27, no. 12, pp. 1643–1655, 2007.

[76] B. Livneh, T. J. Bohn, D. W. Pierce, *et al.*, "A spatially comprehensive, hydrometeorological data set for mexico, the us, and southern canada 1950–2013," *Scientific data*, vol. 2, no. 1, pp. 1–12, 2015.

[77] P. E. Thornton, M. M. Thornton, B. W. Mayer, *et al.*, "Daymet: Daily surface weather on a 1 km grid for north america, 1980-2008," *Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center for Biogeochemical Dynamics (DAAC)*, 2012.

[78] C. Daly, G. Taylor, W. Gibson, T. Parzybok, G. Johnson, and P. Pasteris, "High-quality spatial climate data sets for the united states and beyond," *Transactions of the ASAE*, vol. 43, no. 6, p. 1957, 2000.

[79] R. Behnke, S. Vavrus, A. Allstadt, T. Albright, W. E. Thogmartin, and V. C. Radeloff, "Evaluation of downscaled, gridded climate data for the conterminous united states," *Ecological applications*, vol. 26, no. 5, pp. 1338–1351, 2016.

[80] D. Walton and A. Hall, "An assessment of high-resolution gridded temperature datasets over california," *Journal of Climate*, vol. 31, no. 10, pp. 3789–3810, 2018.

[82] T. C. Peterson and R. S. Vose, "An overview of the global historical climatology network temperature database," *Bulletin of the American Meteorological Society*, vol. 78, no. 12, pp. 2837–2850, 1997.

[83] I. Durre, M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, "Comprehensive automated quality assurance of daily surface observations," *Journal of Applied Meteorology and Climatology*, vol. 49, no. 8, pp. 1615–1633, 2010.

[86] J. Rajczak, S. Kotlarski, N. Salzmann, and C. Schaer, "Robust climate scenarios for sites with sparse observations: A two-step bias correction approach," *International Journal of Climatology*, vol. 36, no. 3, pp. 1226–1243, 2016.

[98] H. Fowler, M. Ekström, S. Blenkinsop, and A. Smith, "Estimating change in extreme european precipitation using a multimodel ensemble," *Journal of Geophysical Research: Atmospheres*, vol. 112, no. D18, 2007.

[99] M. G. Grillakis, A. G. Koutroulis, and I. K. Tsanis, "Multisegment statistical bias correction of daily gcm precipitation output," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 8, pp. 3150–3162, 2013.

[100] M. F. Holthuijzen, B. Beckage, P. J. Clemins, D. Higdon, and J. M. Winter, "Constructing high-resolution, bias-corrected climate products: A comparison of methods," *Journal of Applied Meteorology and Climatology*, vol. 60, no. 4, pp. 455–475, 2021.

[101] A. Mamalakis, A. Langousis, R. Deidda, and M. Marrocu, "A parametric approach for simultaneous bias correction and high-resolution downscaling of climate model rainfall," *Water Resources Research*, vol. 53, no. 3, pp. 2149–2170, 2017.

[102] L. Gudmundsson, *Qmap: Statistical transformations for post-processing climate model output*, R package version 1.0-4, 2016.

[103] A. Gobiet, M. Suklitsch, and G. Heinrich, "The effect of empirical-statistical correction of intensity-dependent model errors on the temperature climate change signal," *Hydrology and Earth System Sciences*, vol. 19, no. 10, pp. 4055–4066, 2015.

[104] J. Hnilica, M. Hanel, and V. Puš, "Multisite bias correction of precipitation data from regional climate models," *International Journal of Climatology*, vol. 37, no. 6, pp. 2934–2946, 2017.

[105] M. G. Grillakis, A. G. Koutroulis, I. N. Daliakopoulos, and I. K. Tsanis, "A method to preserve trends in quantile mapping bias correction of climate modeled temperature," *Earth System Dynamics*, vol. 8, no. 3, p. 889, 2017.

[106] S. Tani and A. Gobiet, "Quantile mapping for improving precipitation extremes from regional climate models," *Journal of Agrometeorology*, vol. 21, no. 4, pp. 434–443, 2019.

[107] W. Yang, J. Andréasson, L. Phil Graham, J. Olsson, J. Rosberg, and F. Wetterhall, "Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies," *Hydrology Research*, vol. 41, no. 3-4, pp. 211–229, 2010.

[108] E. M. Laflamme, E. Linder, and Y. Pan, "Statistical downscaling of regional climate model output to achieve projections of precipitation extremes," *Weather and climate extremes*, vol. 12, pp. 15–23, 2016.

[109] D.-I. Kim, H.-H. Kwon, and D. Han, "Exploring the long-term reanalysis of precipitation and the contribution of bias correction to the reduction of uncertainty over south korea: A composite gamma-pareto distribution approach to the bias correction," *Hydrology and Earth System Sciences Discussions*, pp. 1–53, 2018.

[110] M. Luo, T. Liu, F. Meng, *et al.*, "Comparing bias correction methods used in downscaling precipitation and temperature from regional climate models: A case study from the kaidu river basin in western china," *Water*, vol. 10, no. 8, p. 1046, 2018.

[111] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels, *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.

[118] L. E. Flint and A. L. Flint, "Downscaling future climate scenarios to fine scales for hydrologic and ecological modeling and analysis," *Ecological Processes*, vol. 1, no. 1, p. 2, 2012.

[119] J. Franklin, F. W. Davis, M. Ikegami, *et al.*, "Modeling plant species distributions under future climates: How fine scale do climate projections need to be?" *Global change biology*, vol. 19, no. 2, pp. 473–483, 2013.

[122] D. Maraun, F. Wetterhall, A. Ireson, *et al.*, "Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user," *Reviews of Geophysics*, vol. 48, no. 3, 2010.

[128] W. C. Skamarock, J. B. Klemp, J. Dudhia, *et al.*, "A description of the advanced research wrf model," 2019, Accessed: 2019-03-04. [Online]. Available: https://opensky.ucar.edu/islandora/object/opensky:2898.

[130] H. Huang, J. M. Winter, E. C. Osterberg, *et al.*, "Simulating precipitation and temperature in the lake champlain basin using a regional climate model: Limitations and uncertainties," *Climate Dynamics*, vol. 54, no. 1-2, pp. 69–84, 2020.

[131] USGS, *The national map*, 2018. [Online]. Available: https://viewer.nationalmap.gov/basic/.

[132] T. Wang, A. Hamann, D. L. Spittlehouse, and T. Q. Murdock, "ClimateWNA—high-resolution spatial climate data for western North America," *Journal of Applied Meteorology and Climatology*, vol. 51, no. 1, pp. 16–29, 2012.

[139] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.R-project.org/.

[144] P. Berg, H. Feldmann, and H.-J. Panitz, "Bias correction of high resolution regional climate model data," *Journal of Hydrology*, vol. 448, pp. 80–92, 2012.

[145] S. Perkins, A. Pitman, N. Holbrook, and J. McAneney, "Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions," *Journal of climate*, vol. 20, no. 17, pp. 4356–4376, 2007.

[154] H. J. Fowler, S. Blenkinsop, and C. Tebaldi, "Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling," *International journal of climatology*, vol. 27, no. 12, pp. 1547–1578, 2007.

[156] C. B. Field, V. R. Barros, D. J. Dokken, *et al.*, "Contribution of working group ii to the fifth assessment report of the intergovernmental panel on climate change," *Climate change*, 2014.

[157] J. R. Lanzante, K. W. Dixon, D. Adams-Smith, M. J. Nath, and C. E. Whitlock, "Evaluation of some distributional downscaling methods as applied to daily precipitation with an eye towards extremes," *International Journal of Climatology*, vol. 41, no. 5, pp. 3186–3202, 2021.

[158] K. Hayhoe, C. P. Wake, T. G. Huntington, *et al.*, "Past and future changes in climate and hydrological indicators in the us northeast," *Climate Dynamics*, vol. 28, no. 4, pp. 381–407, 2007.

[159] H. Huang, J. M. Winter, E. C. Osterberg, R. M. Horton, and B. Beckage, "Total and extreme precipitation changes over the northeastern united states," *Journal of Hydrometeorology*, vol. 18, no. 6, pp. 1783–1798, 2017.

[160] D. W. Pierce, D. R. Cayan, E. P. Maurer, J. T. Abatzoglou, and K. C. Hegewisch, "Improved bias correction techniques for hydrological simulations of climate change," *Journal of Hydrometeorology*, vol. 16, no. 6, pp. 2421–2442, 2015.

[161] H. Hoffmann and T. Rath, "Meteorologically consistent bias correction of climate time series for agricultural models," *Theoretical and Applied Climatology*, vol. 110, no. 1, pp. 129–141, 2012.

[162] G. A. Baigorria, J. W. Jones, D.-W. Shin, A. Mishra, and J. J. O'Brien, "Assessing uncertainties in crop model simulations using daily bias-corrected regional circulation model outputs," *Climate Research*, vol. 34, no. 3, pp. 211–222, 2007.

[163] P. Friederichs and A. Hense, "Statistical downscaling of extreme precipitation events using censored quantile regression," *Monthly Weather Review*, vol. 135, no. 6, pp. 2365–2378, 2007.

[164] M.-J. Um, H. Kim, and J.-H. Heo, "Hybrid approach in statistical bias correction of projected precipitation for the frequency analysis of extreme events," *Advances in Water Resources*, vol. 94, pp. 278–290, 2016.

[165] M. Enayati, O. Bozorg-Haddad, J. Bazrafshan, S. Hejabi, and X. Chu, "Bias correction capabilities of quantile mapping methods for rainfall and temperature variables," *Journal of Water and Climate Change*, vol. 12, no. 2, pp. 401–419, 2021.

[166] C. Stager and M. Thill, "Climate change in the champlain basin: What natural resource managers can expect and do, the nature conservancy adirondack ch. and vt ch," *Rep., Keene Valley, NY*, 2010.

[167] A. Monaghan, D. Steinhoff, C. Bruyere, and D. Yates, "Ncar cesm global bias-corrected cmip5 output to support wrf/mpas research," *Research Data Archive National Center Atmospheric Research Computational Information Systems Laboratory, Boulder. DOI*, vol. 10, D6DJ5CN4, 2014.

[168] D. Bannister, A. Orr, S. K. Jain, *et al.*, "Bias correction of high-resolution regional climate model precipitation output gives the best estimates of precipitation in himalayan catchments," *Journal of Geophysical Research: Atmospheres*, vol. 124, no. 24, pp. 14 220–14 239, 2019.

[169] A. M. Wootten, K. W. Dixon, D. J. Adams-Smith, and R. A. McPherson, "Statistically downscaled precipitation sensitivity to gridded observation data and downscaling technique," *International Journal of Climatology*, vol. 41, no. 2, pp. 980–1001, 2021.

[170] P. Reiter, O. Gutjahr, L. Schefczyk, G. Heinemann, and M. Casper, "Bias correction of ensembles precipitation data with focus on the effect of the length of the calibration period," *Meteorologische Zeitschrift*, pp. 85–96, 2016.

[171] T. Peterson, "Climate change indices," *WMO bulletin*, vol. 54, no. 2, pp. 83–86, 2005.

[172] T. R. Karl, N. Nicholls, and A. Ghazi, "Clivar/gcos/wmo workshop on indices and indicators for climate extremes workshop summary," in *Weather and climate extremes*, Springer, 1999, pp. 3–7.

[173] L. Alexander, M. Donat, Y. Takayama, and H. Yang, "The climdex project: Creation of long-term global gridded products for the analysis of temperature and precipitation extremes," in *WCRP Open Science conference, Denver*, 2011.

[174] M. Plummer, A. Stukalov, M. Denwood, and M. M. Plummer, "Package 'rjags'," *Vienna, Austria*, 2016.

[175] N. Smirnov *et al.*, "Table for estimating the goodness of fit of empirical distributions," *Annals of mathematical statistics*, vol. 19, no. 2, pp. 279–281, 1948.

[176] E. A. Rosenberg, P. W. Keys, D. B. Booth, *et al.*, "Precipitation extremes and the impacts of climate change on stormwater infrastructure in washington state," *Climatic change*, vol. 102, no. 1, pp. 319–349, 2010.

[177] G. V. Tschöke, N. S. Kruk, P. I. B. de Queiroz, S. C. Chou, and W. C. de Sousa Junior, "Comparison of two bias correction methods for precipitation simulated with a regional climate model," *Theoretical and applied climatology*, vol. 127, no. 3, pp. 841–852, 2017.

[178] N. M. Razali, Y. B. Wah, *et al.*, "Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests," *Journal of statistical modeling and analytics*, vol. 2, no. 1, pp. 21–33, 2011.

[179] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian journal of statistics*, pp. 65–70, 1979.

[180] J. Martins, H. Fraga, A. Fonseca, and J. A. Santos, "Climate projections for precipitation and temperature indicators in the douro wine region: The importance of bias correction," *Agronomy*, vol. 11, no. 5, p. 990, 2021.

Chapter 4

Calibrating output from numerical climate models featuring a process convolution approach for correcting temporal dependence

4.1 Abstract

Output from numerical models at high spatial and temporal resolutions is critical for applications in a variety of disciplines, ranging from ecology to economics and agriculture. Output from climate models must be brought to a finer spatial resolution and calibrated with respect to observations prior to its use in modeling applications. The calibration of model output, referred to as bias-correction, is a challenging task due to relatively short calibration and long future time periods. Here, we propose a new bias-

correction method in which biases in temporal dependence are assessed and corrected using a multi-resolution process convolution approach. Biases in seasonal means and standard deviations are corrected via a series of linear transformations. The proposed method guarantees that there is no discontinuity in bias-corrected model output at the transition from the calibration to future time period and preserves the future climate change signal. We compare the proposed method to a widely used quantile-mapping bias-correction method, empirical quantile mapping (EQM). After adjustment via the process convolution approach, temporal dependence structures of model output resemble those of observed data. In addition, the proposed method better corrects biases of means and standard deviations compared to EQM. Finally, trends in seasonal means and standard deviations of future model data are preserved using the proposed method, while bias-correction via EQM causes statistical artifacts at extreme quantiles and results in a distortion of the seasonal standard deviation trend over time. While we apply the proposed method to calibrate daily maximum temperature simulations generated by a climate model, principles of this method, especially the use of process convolutions for adjusting temporal dependence, have a wide range of applications.

4.2 Introduction

Numerical models are widely used for simulating physical, atmospheric, and other Earth processes such as climate, ozone [181], CO_2 [182], $PM_{2.5}$ [183], and wind [184]. General circulation models (GCMs), generate simplified, numerical representations of the Earth's climate system [185], [186] and are crucial for understanding future

changes in climate processes. Output from from GCMs is routinely used for local environmental regulatory purposes and decision making [186] as well as climate-impact models [4]. However, GCMs can only resolve large-scale (100-200km) climate patterns and processes, so GCM simulations are of limited use for modeling applications at local or regional scales [5], [8]. For instance, ecological [40] and biological modeling such as forecasting changes in species distributions [187] and investigating ecosystem responses to a changing climate [188], require climate data products at fine spatial and temporal resolutions [11], [121] to capture the the relationships between living organisms and complex landscapes elements over time [187].

Before output from numerical climate models can be used in other modeling applications, it is typically calibrated with respect to observations and transformed to a finer spatial resolution in post-processing steps called bias-correction and downscaling, respectively. During downscaling, simulations from numerical models are transformed from a coarse to finer spatial resolution. Downscaling leverages fine-scale information, such as topography and elevation, to add value to the original outputs [5]. Statistical downscaling involves establishing statistical relationships between coarse-scale climate variables and local, observed data [122]. Statistical downscaling is computationally efficient, can be applied to different types of model output [21], and is especially useful in topographically complex terrain [10]. One drawback is that a substantial amount of observational data is necessary to derive statistical relationships, which may not always be available [14].

The bias-correction of numerical model output is a particularly challenging problem. Uncertainties in GCM output can arise from the simplification and limited understanding of climate processes and the temporal and spatial discretization of

continuous processes [185], [186], as well as future emissions scenarios [189]. Consequently, model output is characterized by some degree of systematic bias. The definition of bias may depend on the application but generally refers to discrepancies in properties (e.g. mean, variance, total rainfall, length of extreme heat days) between model and observed data [189]. During bias-correction, distributional properties (first, second, and potentially higher moments) of climate model output are corrected using observed data [8], [9]. The correction of future model output is informed by relationships between observed and model data over an often relatively short calibration period for which observations are available, whereas future climate projections are generated over periods of about 100 years. While it is generally agreed that model output in the future period should be bias-corrected to some extent [190], the reliability of future climate data, and the degree to which the future climate change signal (CCS) should be bias-corrected, has been debated [115], [189]. The CCS is defined as a measure of additive or multiplicative discrepancies between model statistics in future and calibration time periods [189]. The reliability of the CCS depends largely on the assumption of the stationarity of model biases over time [99], [105]. The so-called stationarity assumption implies that the correction applied during the calibration period will also be sensible in the future period. However, it has been shown that model biases may be non-stationary [7], [47], [105], [113], [114]. Despite uncertainty regarding stationarity of model biases and the degree to which future model simulations should be bias-corrected, there is a broad agreement that the applied correction should minimize distortions to the CCS [47], [113], [190].

Quantile-mapping (QM)-based methods are frequently used for bias-correction [22], [53], [73], [165]. In QM-based methods (also known as histogram mapping or

histogram equalization), a transfer function (TF) maps empirical quantiles of model data to quantiles of observed data, resulting in the correction of first, second, and potentially higher moments of model data. QM methods may be non-parametric (empirical quantile mapping) or parametric (distribution mapping) [46]. The flexibility and efficacy of empirical quantile mapping (EQM) make it a common choice for bias-correction [22], [53], [61], although its flexibility can cause overfitting on calibration data [8], [100]. Most bias-correction methods, including QM, operate under the stationarity assumption [113], [190]. Consequently, QM-based methods generally do not preserve the CCS [18] and may result in statistical artifacts or distortions to the CCS, especially at extreme quantiles [47]. Ultimately, these distortions can introduce greater uncertainty into bias-corrected data. In addtition, QM-based bias-correction methods cannot correct biases in temporal variability except for those at annual scales [191]. Various "trend-preserving" variants of EQM can account for non-stationarity of model errors and preserve the CCS [47], [73], [113], [116]. Generally, in these methods, bias-correction is still achieved by mapping model quantiles to observed quantiles, except that the difference (or ratio) of quantiles in calibration and future periods is incorporated into the process [47].

The correction of biases in temporal dependence has also received attention in the literature. Temporal dependence refers to the correlation of observations at previous and future times. In nested bias-correction (NBC) [91]–[93], temporal biases at multiple, pre-defined time scales are corrected. For instance, [91] used NBC to correct the mean, standard deviation, and lag-1 autocorrelation of daily GCM precipitation simulations at annual and monthly scales. In an extension to NBC, [94] performed the correction at daily, monthly, seasonal, annual, and tri-annual timescales. Though

the nested approaches in [91] and [94] resulted in better representation of temporal structures of GCM simulations compared to EQM, these approaches assume a temporal dependence structure (e.g. lag-1 linear autoregressive model) as well as stationarity of model errors and require the modeler to select the temporal scale(s) in which the correction should be reflected. Bias-correction of temporal dependence has also been performed by signal processing techniques. [95] used the Fast Fourier Transform to decompose observed and model time series; temporal dependence was then corrected via EQM in the frequency domain. This approach assumes stationarity of model errors, which may limit its use for bias-correcting future model simulations. In a trend-preserving approach, [96] decomposed observed and model time series using the discrete wavelet transform (DWT) and corrected biases in mean, standard deviation, and temporal variability in the frequency domain.

We propose a bias-correction approach that is similar to those proposed by [95] and [96] in that the method incorporates a signal processing technique (in this case, a process convolution approach) to correct biases in temporal dependence. Process convolution modeling has been applied to a variety of environmental problems, including the spatiotemporal modeling of ocean temperatures [192], dioxin concentrations [193], and fine air particulates [194]. To our knowledge, a process convolution approach has not been used for bias-correction. A process convolution approach is attractive for this purpose, because it allows for greater flexibility in the decomposition of component frequencies of a signal, and the correction is straightforward to apply. The proposed method corrects periodic and/or non-periodic temporal dependence of model output at various timescales via a process convolution approach; biases in seasonal means and standard deviations are corrected with linear transformations. The proposed

method ensures a smooth transition from the calibration to future time period in corrected model output and minimizes distortions to the CCS. As part of a high-resolution downscaling and bias-correction workflow, we apply the proposed method to historical (1976-2005) and future (2006-2099) daily maximum temperature simulations from a regional climate model. We compare the proposed method to the QM-based method, EQM, and evaluate the abilities of both methods in correcting distributional characteristics of model output in historical and future time periods. While we apply the proposed methods in the context of bias-correction, principles of the method, especially the process convolution approach for adjusting temporal dependence of signals, are generalizable to other variables or modeling applications.

This article is organized as follows: first, we introduce the study area and data. Next, we discuss the process convolution approach for adjusting biases in temporal dependence as well as the linear transformation approach to correct seasonal means and standard deviations of model data. Finally, we compare the performance of the proposed method to EQM and and discuss the implications.

4.3 Methods

4.3.1 Data

The study area, the Lake Champlain Basin, consists of parts of Vermont, New Hampshire, eastern New York and southern Quebec, Canada (Figure 4.1). Eleven watersheds drain into Lake Champlain, and the Green Mountains, Adirondack Mountains, and White Mountains span portions of the region [31].

Historical daily climate station data was obtained from the Global Historical Climate Network (GHCND) (https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND). GHCND data records are adjusted to account for changes in instrumentation and other anomalies [74], [82]. We retained only those stations with at least 70% complete records over the historical time period 1976-2005 (78 stations). We chose to use station data, rather than gridded data products (e.g. Livneh, [76]; Daymet, [77]; and PRISM, [78]), because interpolation algorithms used to create gridded climate products can introduce bias [79] and additional uncertainty when used for bias-correcting climate model output [80]. Gridded climate products can also misrepresent the variability of climate variables [195]. Station data represent direct climatological measurements and are available throughout the Northeast [82] [83]. We acknowledge that there is a spatial misalignment between gridded model data and point-based GHCND station data. In the study region, elevation has the most significant impact on temperature. The WRF model accounts for elevation at a 4km spatial resolution, which is adequate to capture the main effects of elevation within the study region. In addition, the effect of fine-scale (1km) elevation is incorporated via topographical downscaling ([31]), adding further value to model data. There are numerous studies in the bias-correction literature that treat point-based station and downscaled model data as equivalent (e.g [63], [66], [86]).

Daily historical (1976-2005) and future (2005-2099) simulations of two-meter maximum air temperature (TMAX) were generated by the Advanced Weather and Research Forecasting model (WRF) version 3.9.1, a regional climate model (RCM) [128]. WRF is a widely used numerical weather prediction system for both research and applied forecasting purposes [128]. Historical simulations (1976-2005)

were forced by bias-corrected Community Earth System Model 1 (CESM1) [167] under the RCP8.5 emissions scenario, which represents the "worst case" emissions scenario [196]. CESM1 historical simulations were dynamically downscaled with WRF to a 4km resolution using three one-way nests (36 km, 12 km, 4km). The 4km resolution WRF data were used for this study. A total of 4347 4km WRF grid cells cover the study area. WRF grid cell values represent averages over a 4km x 4km area. Additional WRF model details are included in the Supplementary Materials and a full description and evaluation of simulations can be found in [130]. Output from the WRF model will be referred to as "model data" or "model output".

WRF model data are characterized by a pronounced daily mean cold bias during colder periods of the year (approximately days 1-100 and 300-365) compared to observed data (Figure 4.2a). Over the same periods, the variance of WRF model data is larger compared to station observations; during warmer periods (days 120-250) the reverse is true (Figure 4.2b).

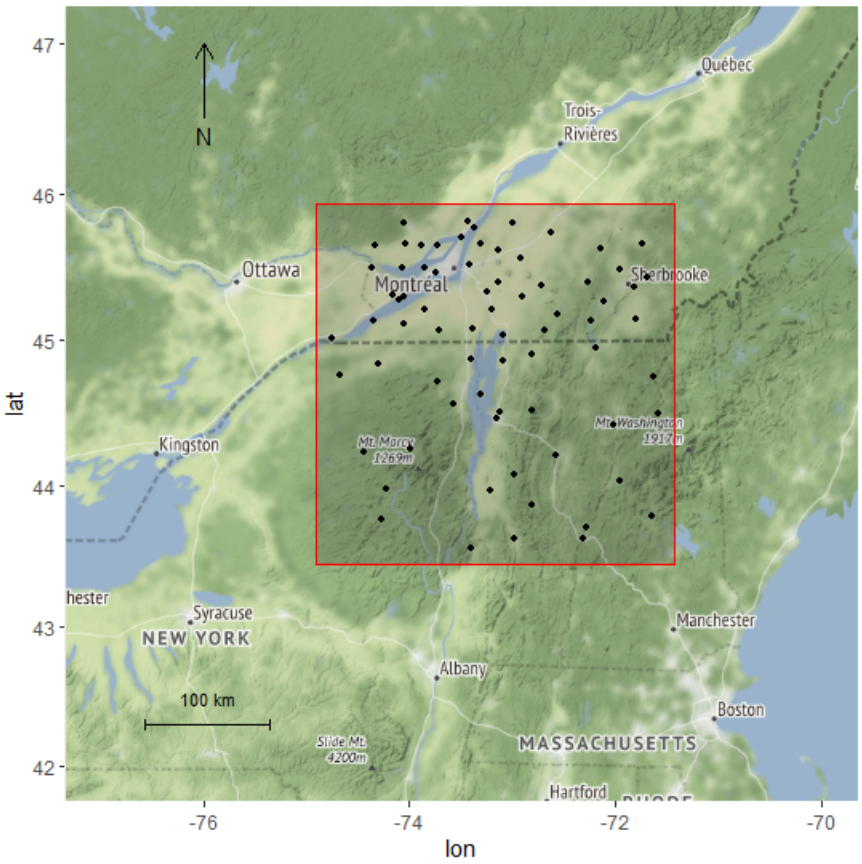

Figure 4.1: GHCND stations (black) within the study area (red). The size of the study area is approximately 13,251km^2.

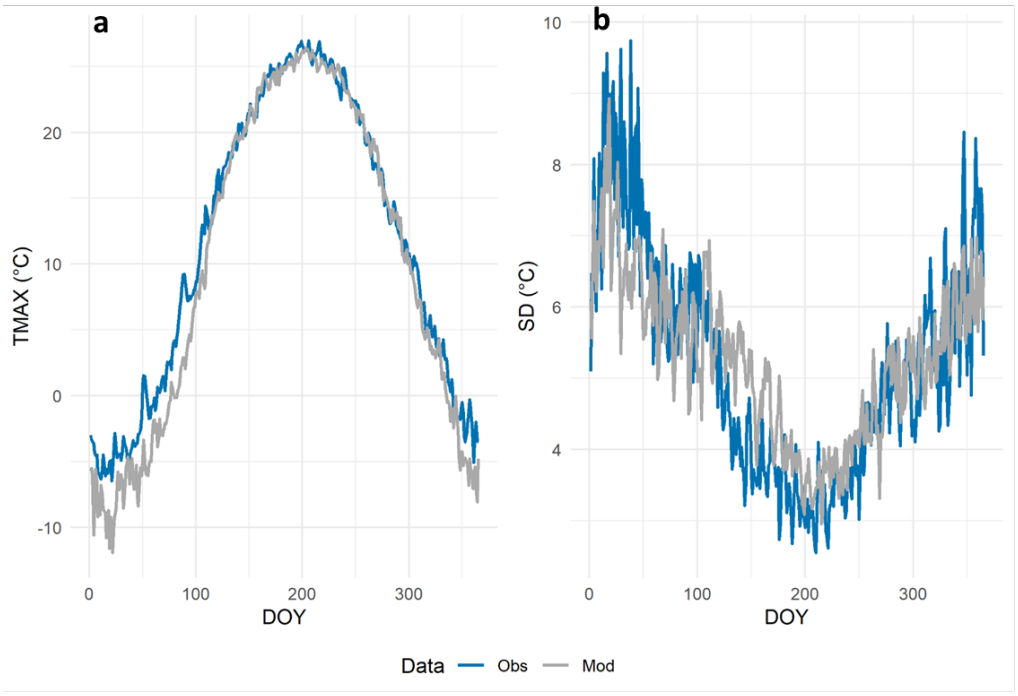

Figure 4.2: Daily means (°C) (a) and standard deviations (SD) (°C) (b) of WRF model (Mod) and observed data (Obs) over day of year (DOY) over 1976-2005. Means and SDs are calculated using WRF model data downscaled to the 78 GHCND station locations and GHCND station data.

4.3.2 Modeling

Historical and future model data are topographically downscaled to the 78 GHCND station locations for the purposes of constructing statistical models and calculating performance metrics. Topographic downscaling is a variation of inverse distance weighting (IDW) that can account for fine-scale topography via the incorporation of elevational lapse rates [31] (full details on topographic downscaling are provided in the Supplementary Material). To construct data products, model data were also downscaled to a 1km grid. Elevation estimates at each 1km grid cell were derived by interpolating elevation values from a 30m digital elevation model (DEM) [131]

via IDW, while elevations at GHCND station locations were derived from GHCND metadata [74]. The 1-km grid cell size was chosen based on resolution requirements for climate impact modeling efforts over the Lake Champlain Basin [31], [132]. The use of topographical downscaling is a simple, yet highly effective way to account for the effect of elevation on temperature variables and provides a spatial adjustment prior to bias-correction.

The proposed bias-correction method consists of two parts: a process convolution approach is used to correct the temporal dependence of model data and a series of linear transformations is employed to correct biases in seasonal means and standard deviations of model data. We refer to the proposed method as *Distribution Modification with Temporal Dependence Adjustment* (DMTA). However, the adjustment in temporal dependence is independent of the correction of distributional discrepancies in model data.

DMTA bias-correction consists of the following steps. It is important to note that the correction of temporal dependence and distributional bias is carried out on spatially-averaged data. The last step of the process involves adjusting spatially-explicit data to match the corrections made at the spatially-averaged scale.

1. Time series of daily, spatial averages for model and observed data over the 78 station locations are obtained for the historical and future time periods. The time series of model data will span historical and future time periods (1976-2099) (section 4.3.3).

2. Spatially-averaged time series from step 1 are processed such that both time series have seasonal means of 0 and standard deviations of 1. The removal of any seasonal or long term trends, as well as patterns in seasonal variance, is

necessary to obtain robust parameter estimates when fitting process convolution (PC) models (section 4.3.3).

3. PC models are fitted to processed, observed and model time series obtained in step 2. Using estimates from PC model fits, the model time series is adjusted such that the temporal dependence matches that of the observed time series (sections 4.3.4 and 4.3.5).

4. The results from step 3 are back-transformed by reversing the data processing steps in step 2. However, in the back-transformation, seasonal means and standard deviations (SDs) obtained from observed data are used, ensuring that back-transformed model data has seasonal means and SDs similar to those of observed data. The resulting back-transformed data represents spatially-averaged model data in which bias in temporal dependence and distribution (mean and SD) have been corrected (section 4.3.6).

5. To reflect the bias-correction of temporal dependence and distribution at spatially-explicit model values, a series of linear transformations is applied. After these linear transformations, daily means and SDs of spatially-explicit model data will reflect the bias-corrected model data obtained in step 4 (section 4.3.6).

The correction of temporal dependence is carried out separately from the correction of bias in means and SDs. This means that the process convolution approach for correcting temporal dependence could be applied within other bias correction methods (such as EQM) or in other applications entirely. The correction of bias in means and SDs may also be performed without the correction of bias in temporal dependence. Figure 4.3 shows the workflow of the DMTA bias-correction method as a schematic.

The following sections will describe in detail the processing steps for model and observed data (4.3.3), the process convolution approach for adjusting temporal dependence (4.3.4 and 4.3.5), and the final linear transformation necessary to adjust daily spatially-explicit model data (4.3.6).

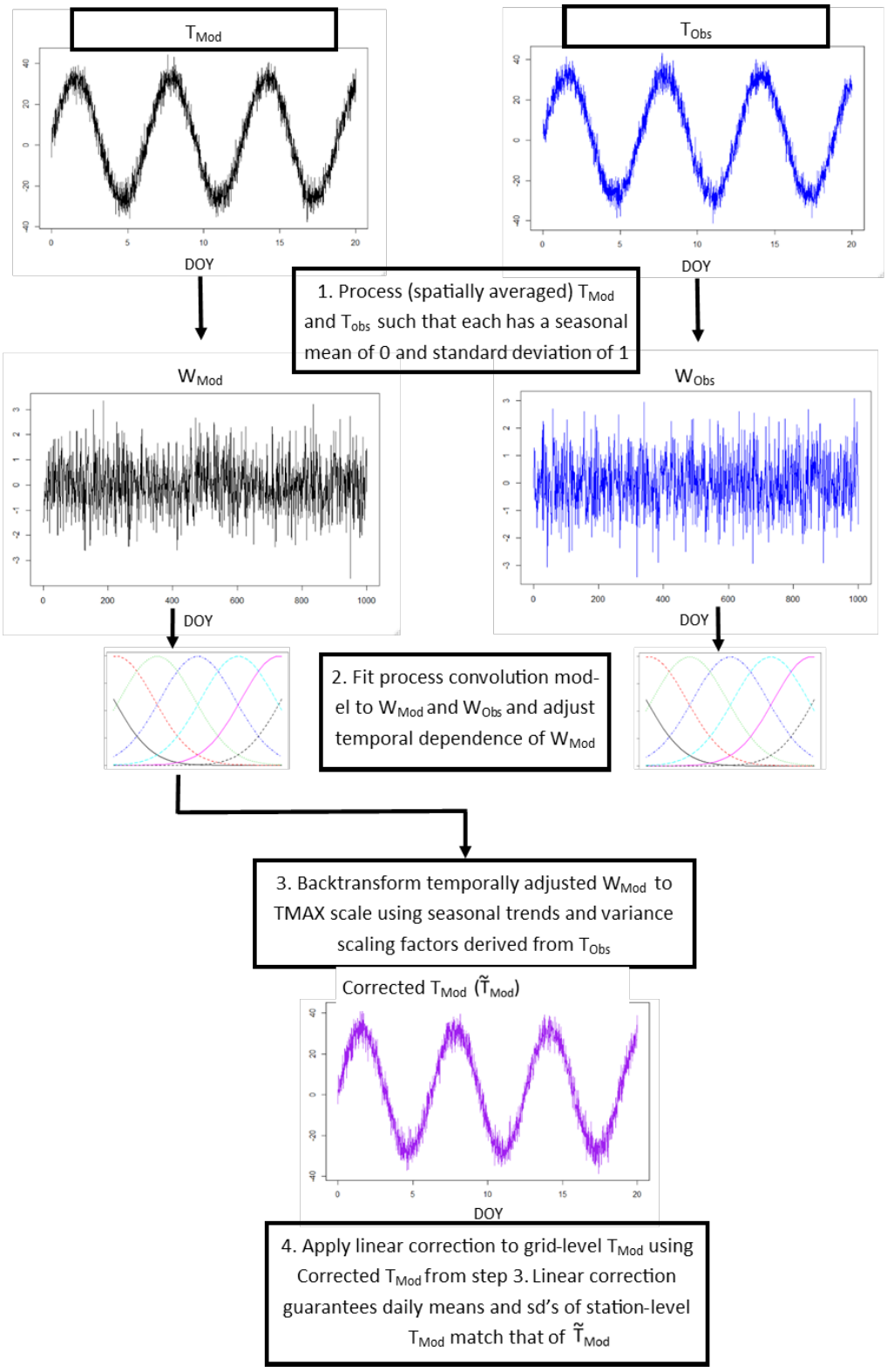

Figure 4.3: Workflow of the DMTA bias-correction method.

4.3.3 Data processing

In the following sections, subscripts Obs and Mod will refer to observed and model data, and $hist$ and fut subscripts will denote historical and future time periods. ϵ and μ are used to denote residual errors and means, respectively, and the subscript w will refer to residual variance. Subscripts will be utilized frequently to distinguish between parameter estimates related to observed or model data in future and historical time periods.

Let $\mathbf{T}_{Obs}$ and $\mathbf{T}_{Mod}$ denote spatially-averaged observed and model data, respectively. $\mathbf{T}_{Obs}$ represent the historical period (1976-2005), while $\mathbf{T}_{Mod}$ span years 1976-2099. $\mathbf{T}_{Obs}$ and $\mathbf{T}_{Mod}$ are processed according to (4.1) and (4.2), where $\mathbf{W}_{Mod}$ and $\mathbf{W}_{Obs}$ are the resulting time series with seasonal means of 0 and SDs of 1.

$$\mathbf{W}_{Mod} = \frac{(\mathbf{T}_{Mod} - \boldsymbol{\phi}_{Mod,hist} - \boldsymbol{\delta}_{Mod})}{\boldsymbol{\alpha}_{Mod}} \times \frac{1}{c_{Mod}} \tag{4.1}$$

$$\mathbf{W}_{Obs} = \frac{(\mathbf{T}_{Obs} - \boldsymbol{\phi}_{Mod} - \boldsymbol{\delta}_{Obs})}{\boldsymbol{\alpha}_{Obs}} \times \frac{1}{c_{Obs}}. \tag{4.2}$$

In (4.1) and (4.2), $\boldsymbol{\delta}_{Mod}$ and $\boldsymbol{\delta}_{Obs}$ are vectors of seasonal means and $\boldsymbol{\alpha}_{Mod}$ and $\boldsymbol{\alpha}_{Obs}$ are vectors of seasonal SDs, where elements of these vectors repeat every 365 values. The vectors $\boldsymbol{\delta}_{(\cdot)}$ and $\boldsymbol{\alpha}_{(\cdot)}$ contain estimated daily mean and standard deviation values for each day in a 365-day year (Figure 4.4) (the model for estimating $\boldsymbol{\delta}_{(\cdot)}$ and $\boldsymbol{\alpha}_{(\cdot)}$ is described below). c_{Obs} and c_{Mod} are constant scalars that scale the SD of the time series to be approximately 1. Scaling constants c_{Obs} and c_{Mod} were calculated after dividing by $\boldsymbol{\alpha}_{(\cdot)}$ and reflect the mean median absolute deviation (MAD) over a sliding window of length 100. $\boldsymbol{\phi}_{Mod}$ is the long-term estimated daily mean trend in $\mathbf{T}_{Mod}$.

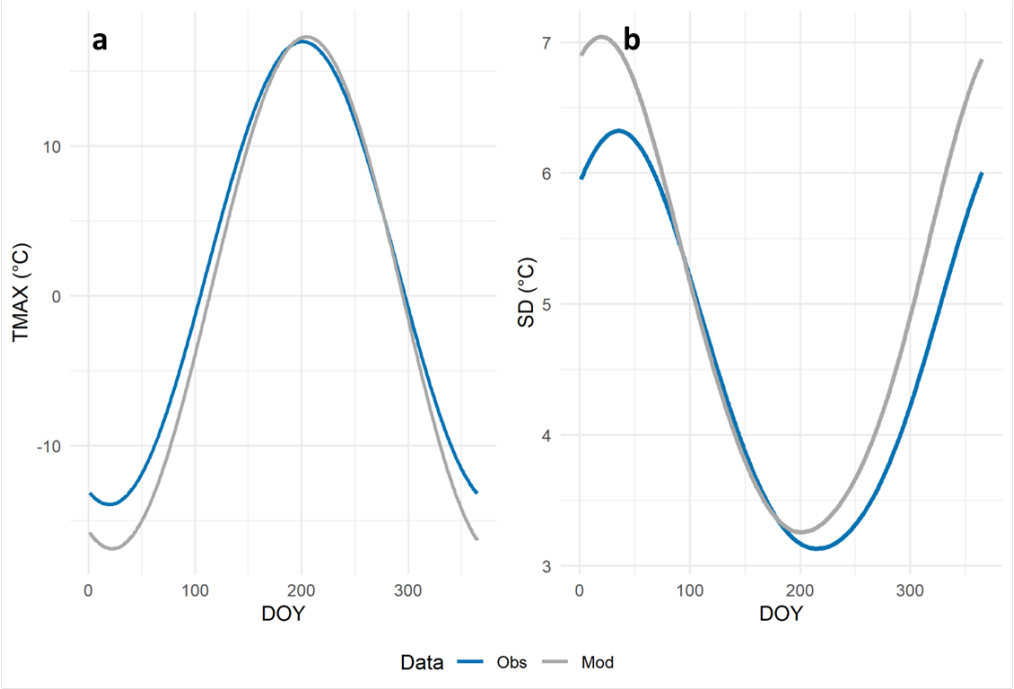

Figure 4.4: Fitted seasonal means ($\boldsymbol{\delta}_{(.)}$) (°C) (a) and SDs ($\boldsymbol{\alpha}_{(.)}$) (°C) (b) of WRF model (Mod) and observed data (Obs) via the HetGP model.

The long-term daily mean trend of $\mathbf{T}_{Mod}$ was estimated using a cubic regression spline with calendar day as the predictor variable. The number of knots (50) in the spline was chosen using the likelihood ratio test. Spline models were fitted using the `mgcv` package [197] in R. Estimation of $\boldsymbol{\phi}_{Mod}$ (4.5) over the entire time period (1976-2099) is important to ensure a smooth transition from the calibration to future time period.

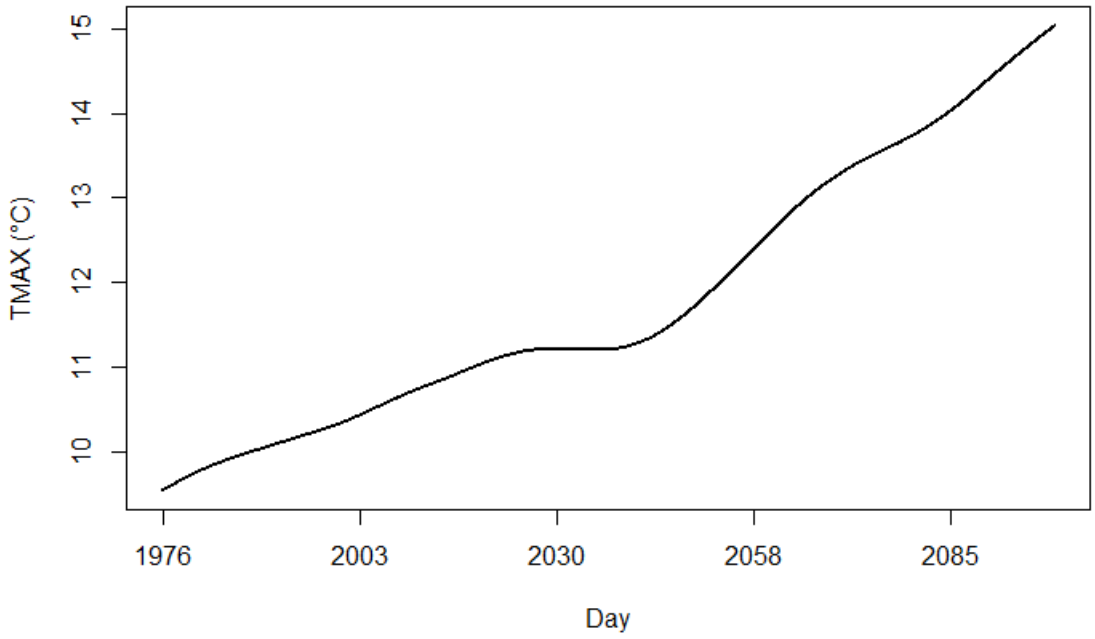

Figure 4.5: Estimated long-term daily mean trend (ϕ_{Mod}) of model data (1976-2099).

SDs of $\mathbf{T}_{Obs}$ and $\mathbf{T}_{Mod}$ exhibit considerable annual variation depending on day of year (Figure 4.2). To estimate seasonal means and SDs of $\mathbf{T}_{Obs}$ and $\mathbf{T}_{Mod}$, while accounting for the nonstationarity seasonal SDs, we used a heteroscedastic Gaussian process model (HetGP) [198]. In the implementation of the HetGP model by [199], replication is leveraged to model process noise, which also makes model fitting computationally efficient. As a result, the number of input locations is reduced from N to n, where $n << N$. The HetGP model is similar in idea to stochastic kriging [200] but is advantageous because estimates for variance and lengthscale parameters are obtained using maximum likelihood. Essentially, a HetGP is used for Gaussian process regression, except that process noise is allowed to vary over input space.

In a HetGP model, latent variance variables are used in lieu of a single nugget parameter. Let $\tau_1 ... \tau_n$ denote latent nuggets for n unique input locations. The τ_i are stored in a diagonal matrix $\boldsymbol{\Delta}_n$, and a GP prior is specified for their distribution [199]:

$$\boldsymbol{\Delta}_n \sim N(0, \nu(\boldsymbol{C}_n + g\boldsymbol{A}^{-1})). \tag{4.3}$$

In (4.3) ν is a scale parameter, $\mathbf{C}_n$ is a positive-definite correlation matrix, g is a nugget, and $\mathbf{A}$ is a diagonal matrix containing the number of replicates at each of the unique n locations. To obtain smoothed values of λ_i, $\boldsymbol{\Delta}_n$ is inserted into the GP prediction equation for the mean:

$$\boldsymbol{\Lambda}_n = \mathbf{C}_n(\mathbf{C}_n + g\mathbf{A}^{-1})^{-1}\boldsymbol{\Delta}_n.$$

Here, g is an overall nugget and controls the smoothness of the λ_i relative to the τ_i. We refer readers to [199] for a thorough description of HetGP modeling.

HetGP models were fitted to $\mathbf{T}_{Obs}$ and $\mathbf{T}_{Mod}$ (after subtraction of $\boldsymbol{\phi}_{Mod}$) using the `mleHetGP` function with the HetGP package [199] in R. Periodicity of $\boldsymbol{\delta}_{(\cdot)}$ and $\boldsymbol{\alpha}_{(\cdot)}$ were enforced during HetGP model fitting (see Supplementary Material for additional details). The resulting $\boldsymbol{\delta}_{(\cdot)}$ and $\boldsymbol{\alpha}_{(\cdot)}$ fitted by the HetGP model were adequately smooth and periodic, as required (Figure 4.4).

4.3.4 Process convolution modeling

Process convolution models represent a flexible and computationally efficient method for Gaussian process modeling [192]. In this study, we apply process convolution

modeling as an alternative to other signal decomposition procedures such as the discrete Fourier transform (DFT) or discrete wavelet transform (DWT) to gain insight into differences in temporal dependence between $\mathbf{T}_{Obs}$ and $\mathbf{T}_{Mod}$. In a PC approach, it is straightforward to determine if differences in temporal dependence exist between model and observed data. In addition, multivariate Gaussian process theory can be utilized to adjust model time series such that temporal dependence is similar to that of observed data.

Similar to the DFT or DWT, a PC can be used to decompose a time series into component frequencies. All three approaches are similar in that each decomposes a signal into a set of basis functions. One major disadvantage of the DFT is that information regarding time is lost, as the Fourier transform is an integral over time [201]. In addition, the Fourier transform is exclusively based on sine and cosine basis functions, which may be limiting for some types of signals. The DWT can decompose signals composed of periodic or non-periodic frequencies in which variation changes over time [201]. Though the DWT can represent a greater diversity of composite frequencies, the frequencies are typically represented by a dyadic scale. In a PC approach, a signal is decomposed via basis functions [192] that are specified by a defined kernel and its parameters. Though the Gaussian kernel is frequently used, others (e.g. spherical, exponential) also exist [192]. Parameters controlling the length scale are explicitly chosen to represent temporal frequencies present in a signal. The need to explicitly specify kernel parameters is not always an advantage, but it does allow for greater control over how signals are decomposed. Standard Gaussian process theory can then be used to adjust temporal dependence of composite frequencies.

The framework underlying PC modeling is the Gaussian process (GP), a proba-

bility distribution over possible functions. A flexible and efficient way to specify a GP $z(t)$ is by convolving a white noise process $x(t)$ with a smoothing kernel $k(t)$ [202]:

$$z(t) = \int_T k(u-t)x(u)d(u), \text{ for } t \in T, \tag{4.4}$$

for time t. The covariance function for $z(t)$ is only dependent on the distance vector $(d = t - t')$:

$$c(d) = \text{Cov}(z(t), z(t')) = \int_T k(u-t)k(u-t')du = \int_T k(u-d)k(u)du.$$

The process $z(t)$ is a 0-mean GP and is defined by specifications for the latent process $x(t)$ and the smoothing kernel $k(t)$. In practice, the theoretically continuous process $z(t)$ (4.4) is discretized, and the latent process $x(t)$ is specified to be non-zero at "knot" locations $\omega_1 \ldots \omega_m \in T$.

Let $x = (x_1 \ldots x_m)^T$ where $x_j = x(\omega_j), j = 1 \ldots m$ and let $x_j \sim N(0, \sigma^2_{x_j})$. The knots, ω_j, anchor the process $z(t)$ at specific locations. The process $z(t)$ can be expressed as

$$z(t) = \sum_{j=1}^{m} x_j k(t - \omega_j).$$

Thus, $z(t)$ is a 0-mean GP in which $Cov(z(t), z(t')) = \sigma_x \sum_{j=1}^{m} k(t-\omega_j)k(t'-\omega_j)$ and $\mathbf{K}\mathbf{K}^t = \mathbf{\Sigma}_z$, where $\mathbf{K}_{ij} = k_j(t_i)$. The discrete representation of $z(t)$ provides a good approximation to the continuous process, as long as the distance between knots is no larger than the standard deviation of the smoothing kernel $k(\cdot)$ [202], [203].

In this study, $k(\cdot, -\omega_j)$ is the Gaussian basis kernel centered at ω_j:

$$k_j(t) \propto 1/\sqrt{2\pi}\exp(-1/2(t-\omega_j)^2).$$

Let $\mathbf{w} = (w_1 \dots w_n)^T$ be observations recorded at times $t_1 \dots t_n \in T$. $\mathbf{w}$ can be modeled in a regression context as:

$$\mathbf{w} = \mathbf{K}\mathbf{x} + \boldsymbol{\epsilon}, \tag{4.5}$$

where $\mathbf{K}$ is a basis matrix. The elements in the basis matrix $\mathbf{K}$ are given by $\mathbf{K}_{i,j} = k_j(t_i - \omega_j x_j)$, $x \sim N(0, \sigma^2_{x_j})$, and $\boldsymbol{\epsilon} \sim N(0, \sigma^2_\epsilon I_n)$.

Model (4.5) can also be extended to incorporate multiple temporal resolutions [202]. In this approach, multiple resolutions can be captured by adjusting the length-scale parameter of $k(s)$. Since we use the Gaussian kernel in this application, the kernel standard deviation can be narrowed and widened to capture fine- and large-scale resolutions, respectively. $z(t)$ is expressed as the sum of p processes:

$$z(t) = \sum_{\ell=1}^{p} z_\ell(t).$$

Each of the p processes can be expressed as:

$$z_\ell(t) = \sum_{j=1}^{m_\ell} x_{\ell j} k(t - \omega_{\ell j}),$$

where m_ℓ is the number of knots in $\mathbf{K}^\ell$. Thus, each $z_\ell(t)$ captures additional detail. In the context of a multi-resolution PC, the regression model is:

$$\mathbf{w} = \sum_{\ell=1}^{p} \mathbf{K}^\ell \mathbf{x}_\ell + \boldsymbol{\epsilon}, \tag{4.6}$$

which reduces again to $\mathbf{w} = \mathbf{K}\mathbf{x} + \boldsymbol{\epsilon}$ (4.5), where

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_p \end{bmatrix}$$

and

$$\mathbf{K} = [\mathbf{K}^1 \dots \mathbf{K}^p]$$
$$\epsilon \sim N(0, \sigma_\epsilon^2 I_n)$$
$$x_\ell \sim N(0, \sigma_{x_\ell}^2 I_{m\ell}).$$

The elements of $\mathbf{K}^\ell$ are given by $K_{ij}^\ell = k(t_i - \omega_{\ell_j})$.

The multi-resolution PC is ideal for modeling processes composed of multiple sub-processes that may be periodic or non-periodic. To minimize edge effects, in fitting (4.5) it is important that basis matrices are constructed correctly. To minimize edge effects of non-periodic basis matrices, kernels were initiated at $2.25 \times \sigma_{x_\ell}^2$ and were extended to $n + 2.25 \times \sigma_{x_\ell}^2$, where n was the number of days used in model fitting, and $\sigma_{x_\ell}^2$ is the variance of x_ℓ. All basis matrices were standardized relative to a Normal(0, $\sigma_{x_\ell}^2$) distribution according to (4.7), such that the resulting covariance matrix ($K^\ell K^{\ell^T}$) had marginal variances of 1, which is assumed in model 4.5.

$$K_{i,j}^{\ell^*} = \frac{1}{\sigma_{x\ell}\sqrt{2\pi}e^{-\frac{x}{\sigma_{x\ell}}}}/\omega; \tag{4.7}$$

$$\omega = \sqrt{max(K^{\ell^*}(K^{\ell^*})^T)} \tag{4.8}$$

A Bayesian approach to fitting model (4.5) requires the specification of a likelihood function for $\mathbf{w}$ and priors for $\lambda_x = 1/\sigma^2_{x_\ell}$ and $\lambda_w = 1/\sigma^2_w$:

$$L(\mathbf{w}|\mathbf{x}, \lambda_w) \propto \lambda_w^{n/2} \exp(-\frac{1}{2}(\mathbf{w} - \mathbf{K}\mathbf{x})^T \lambda_w (\mathbf{w} - \mathbf{K}\mathbf{x})).$$

Priors, denoted by $\pi(\cdot)$, are designated as follows [192]:

$$\begin{aligned}
\pi(\mathbf{x}|\lambda_x) &\propto \lambda_x^{m/2} \exp(-\frac{1}{2}\lambda_x \Lambda_x x^T) \\
\pi(\lambda_x) &\sim \lambda_x^{a_x - 1} \exp(-0.0001\lambda_x) \\
\pi(\lambda_w) &\propto \lambda_w^{a_w - 1} \exp(-0.0001\lambda_w),
\end{aligned}$$

Here, λ_x is the precision (inverse variance) of $x_\ell, \ell = 1 \ldots p$. In addition, λ_ϵ is the precision of $\mathbf{w}$, Γ denotes the Gamma distribution, and

$$\mathbf{\Lambda_x} = \begin{bmatrix} \lambda_1 I_{m_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_p I_{m_p} \end{bmatrix}.$$

The posterior conditional distributions of $\mathbf{x}$, which is of most interest here, is given by:

$$\mathbf{x}|\lambda_x, \lambda_w, \mathbf{w} \sim N(\boldsymbol{\mu}, \mathbf{V}), \tag{4.9}$$

$$\boldsymbol{\mu} = (\lambda_w \mathbf{K}^T \mathbf{K} + \boldsymbol{\Lambda}_{\boldsymbol{x}})^{-1} \lambda_w \mathbf{K}^T \mathbf{w}, \tag{4.10}$$

$$\mathbf{V} = (\lambda_w \mathbf{K}^T \mathbf{K} + \boldsymbol{\Lambda}_{\boldsymbol{x}})^{-1}. \tag{4.11}$$

We fit the multi-resolution PC model (4.5) as a random effects model in which estimates of λ_x, and λ_w were obtained with maximum likelihood. Maximum likelihood was computationally more efficient than MCMC sampling, but given a large enough sample size, both approaches yield similar estimates [204]. PC models were fitted using the `lme` function from the `nlme` package in R [205].

Model (4.5) was fitted to $\mathbf{W}_{Mod}$ and $\mathbf{W}_{Obs}$; however, $\mathbf{W}_{Mod}$ was fitted in 30-year increments (1976-2005, 2006-2036, 2037-2067, and 2068-2099) due to computational limitations. Initially, PC models were fitted to $\mathbf{W}_{Mod}$ and $\mathbf{W}_{Obs}$ during the historical period, where $\mathbf{K}$ was composed of basis matrices with the following kernel standard deviations: 3.5, 7, 14, 21, 28, 35, 42, 48, 56, 73, 79, 100, 120, 140, 180, 210, and 1825 days. Based on exploratory analyses, we only included basis matrices representing non-periodic temporal trends at 180- and 3.5-day frequencies (e.g., Gaussian kernel standard deviations were 180 and 3.5 days). Gaussian kernel bases constructed with standard deviations of 180 and 3.5 days capture component processes relatively large (180-day) and fine (3.5-day) resolutions, respectively. In other words, Gaussian kernel bases constructed with standard deviation m can model component processes in which the covariance matrix is characterized by an $m-$day lengthscale parameter (the covariance matrix for any component process $z_\ell(t) = K^\ell K_\ell^T$). (We found that

estimates of $\sigma^2_{x_\ell}$ representative of all other temporal frequencies were nearly zero, and this was consistent for $\mathbf{W}_{Obs}$ and $\mathbf{W}_{Mod}$).

4.3.5 Process convolution approach for adjusting temporal dependence

Although PC model specification is facilitated by the use of precision rather than variance, we shall discuss inference using variance of the x_ℓ ($\sigma^2_{x_\ell}$).

It is clear from model (4.5) that elements of x_ℓ influence the relative importance of temporal trends represented by the $\mathbf{K}_\ell$. If a temporal trend, $z_\ell(t) = \mathbf{K}_\ell x_\ell$, is not well-represented, $z_\ell(t)$ is near 0, which means $\sigma^2_{x_\ell}$ is also near 0. Conversely, if some temporal trend $z_\ell(t)$ is well-represented in the signal, σ_{x_ℓ} will be nonzero. Thus, inference is focused on extracting information about the variance of the x_ℓ ($\sigma^2_{x_\ell}$) and ultimately the posterior conditional distribution of $\mathbf{x}$. The goal in this section is to reconstruct an adjusted version of $\mathbf{W}_{Mod}$ by scaling PC model fitted values and residual variance using estimates of $\sigma^2_{x_{\ell,Obs}}$ and $\sigma^2_{w,Obs}$, so that $\mathbf{W}_{Mod}$ exhibits temporal dependence characteristics of $\mathbf{W}_{Obs}$. After PC models were fitted to W_{Obs} and W_{Mod}, estimates of $\sigma^2_{x_{\ell,Obs}}$ and $\sigma^2_{x_{\ell,Mod}}$, as well as estimates of residual variances, ($\sigma^2_{w,Obs}$ and $\sigma^2_{w,Mod}$), were obtained. Posterior conditional means of x for PC models fitted to $\mathbf{W}_{Obs}$ and $\mathbf{W}_{Mod}$ were obtained using (4.10).

Next, the ratios between $\sigma^2_{x_{\ell,Obs}}$ and $\sigma^2_{x_{\ell,Mod}}$ is calculated as:

$$r_\ell = \frac{\sigma^2_{x_{\ell,Obs}}}{\sigma^2_{x_{\ell,Mod}}} \times c^*,$$

where $c^* = 0.96$ is a scaling constant necessary to obtain correct estimates of σ_{x_ℓ} when

$r_\ell > 1$. The scaling constant c^* is required, because the scaling of $\mu_{x,Mod}$ by the r_ℓ does not act linearly on variance. [206] showed that in random effects models (such as the PC implementation in this study), variances of random effects do not scale linearly. We found that adjusting r_ℓ by $c^* = 0.96$ resulted in a satisfactory adjustment of σ_{x_ℓ}. Since the technique presented here has not been used in the literature, we found that determining an appropriate value for $c^* = 0.96$ required testing values until we obtained the desired change in σ_{x_ℓ}. r_ℓ near 1 indicate that the strength of a component temporal process is present in similar magnitudes in observed and model data. Ratios greater than 1 indicate the magnitude of the component process is greater in $\mathbf{W}_{Obs}$ compared to $\mathbf{W}_{Mod}$, and the opposite is true for ratios less than 1. The temporal dependence of $\mathbf{W}_{Mod}$ is adjusted using results from model (4.5) fitted to $\mathbf{W}_{Obs}$ only during the historical period.

The ratio of residual variances $\sigma^2_{w,Obs}$ and $\sigma^2_{w,Mod}$ is calculated as:

$$r_w = \frac{\sigma^2_{w,Obs}}{\sigma^2_{w,Mod}}$$

The posterior mean of $\mathbf{x}_{Mod}$ ($\boldsymbol{\mu}_{x,Mod}$) is adjusted using the following expression:

$$\tilde{\boldsymbol{\mu}}_{x,Mod} = \boldsymbol{\mu}_{x,Mod} \times \begin{bmatrix} \mathbf{R}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{R}_p \end{bmatrix},$$

where $\mathbf{R}_1 \ldots \mathbf{R}_p$ are diagonal matrices with diagonal elements $r_1 \ldots r_p$.

Next, adjusted residuals of the PC model fitted to $\mathbf{W}_{Mod}$ are calculated as:

$$\tilde{\boldsymbol{\epsilon}}_{Mod} = (\mathbf{W}_{Mod} - \mathbf{K}\mu_{x,Mod}) \times r_w. \tag{4.12}$$

Now, the adjusted version of $\mathbf{W}_{Mod}$ ($\tilde{\mathbf{W}}_{Mod}$) is constructed by adding *adjusted* residuals (4.12) to *adjusted* fitted values (4.13):

$$\tilde{\mathbf{W}}_{Mod} = \mathbf{K}\tilde{\boldsymbol{\mu}}_{x,Mod} + \tilde{\boldsymbol{\epsilon}}_{Mod}. \tag{4.13}$$

The adjustment of temporal dependence may result in the increase or decrease of the overall variance of $\tilde{\mathbf{W}}_{Mod}$ with respect to $\mathbf{W}_{Mod}$, which may be undesirable. This can be resolved by performing the data processing steps again outlined in section 4.3.3 and obtaining new estimates of $\boldsymbol{\delta}_{Mod}$, $\boldsymbol{\alpha}_{Mod}$, $\boldsymbol{\phi}_{Mod}$, and c_{Mod}.

4.3.6 Correction of spatially-explicit model data

The previous section describes how a process convolution approach can be utilized to assess temporal dependence within $\mathbf{W}_{Mod}$ and $\mathbf{W}_{Obs}$ and how $\mathbf{W}_{Mod}$ can be adjusted such that it exhibits temporal dependence characteristics of $\mathbf{W}_{Obs}$. However, up to this point, only the temporal dependence of $\tilde{\mathbf{W}}_{Mod}$, which is at the spatially-averaged scale, has been corrected; distributional properties of $\tilde{\mathbf{W}}_{Mod}$ must also be corrected, and ultimately, those changes must be reflected in spatially-explicit model data. First, a back-transformation is applied to $\tilde{\mathbf{W}}_{Mod}$ so that the data is at the "TMAX" scale. This back-transformation ensures that the seasonal means and SDs of the resulting back-transformed time series are the same as those of observed data. The back-transformation is performed as follows:

1. $\tilde{\mathbf{W}}_{Mod}$ is multiplied by the constant scalar obtained from observed data during the calibration period, c_{Obs}.

2. The result from step 1 is multiplied by the seasonal SD trend of observed data,

$\boldsymbol{\alpha}_{Obs}$.

3. The long-term model trend $\boldsymbol{\phi}_{Mod}$ as well as the seasonal mean trend from observed data, $\boldsymbol{\delta}_{obs}$, are added to the result from step 2.

These steps can be expressed as:

$$\tilde{\mathbf{T}}_{Mod} = \tilde{\mathbf{W}}_{Mod} \times c_{Obs} \times (\boldsymbol{\alpha}_{Obs}) + \boldsymbol{\delta}_{Obs} + \boldsymbol{\phi}_{Mod}, \tag{4.14}$$

where $\tilde{\boldsymbol{T}}_{Mod}$ is bias-corrected, spatially averaged model data. The back-transformation in (4.14) ensures that seasonal means and standard deviations of $\tilde{\mathbf{T}}_{Mod}$ match those of $\mathbf{T}_{Obs}$, while the long-term term daily mean trend of $\mathbf{T}_{Mod}$ is preserved.

The final step in bias-correcting spatially-explicit model data involves another series of linear transformations applied to spatially-explicit model values. Linear transformations are applied to model data that have been topographically downscaled to the 78 GHCND station locations (or the fine-scale grid for the construction of high-resolution data products). The final linear transformation is applied to N locations (GHCND station locations or fine-scale grid cells) for each day $t, t = 1 \ldots T$ in historical and future time periods. Each uncorrected, downscaled model value at location $i, i = 1 \ldots N$ is corrected using the following linear transformation:

$$\begin{aligned} \mathbf{T}_{corr_{i,t}} &= b \times \mathbf{T}_{Mod_{i,t}} + a, \\ a &= \tilde{T}_{Mod,t} - T_{Mod,t} \times \frac{\alpha_{Obs,t}}{\alpha_{Mod,t}}, \\ b &= \frac{\alpha_{Obs,t}}{\alpha_{Mod,t}}, \end{aligned} \tag{4.15}$$

In (4.15) $T_{corr_{i,t}}$ is the corrected daily model value at grid cell i and day t. $T_{Mod_{i,t}}$ is the uncorrected model value at location i and day t; $\tilde{T}_{Mod,t}$ is the value of $\tilde{T}_{Mod}$ on day t, $T_{Mod,t}$ is the value of $\mathbf{T}_{Mod}$ on day t. The final linear transformation ensures that the daily spatially-averaged means and standard deviations of fine-scale grid-level model data matches those of corrected model data, $\tilde{\mathbf{T}}_{Mod}$, for all $t \in T$.

4.4 Validation and performance assessment

To quantify performance of the proposed method in correcting biases in the mean and SD, we compared it against EQM, a widely-used method for bias-correction. We implemented EQM and DMTA with and without the adjustment of temporal dependence.

4.4.1 Empirical quantile mapping

Bias-correction via empirical quantile mapping (EQM) is accomplished with monthly transfer functions (TFs), where each TF represents a mapping from empirical model quantiles to observed empirical quantiles for a particular month. The EQM TF is expressed by the empirical cumulative distribution function (ecdf) and its inverse (ecdf^{-1}).

$$T_{corr_{i,t}} = \text{ecdf}^{-1}_{Obs,m}(\text{ecdf}_{Mod,m}(T_{Mod_{i,t}})). \tag{4.16}$$

In (4.16), $T_{corr_{i,t}}$ is the corrected model TMAX value at location i on day t, $\text{ecdf}^{-1}_{obs,m}$ is the inverse ecdf of T_{Obs} for month m, and $\text{ecdf}_{Mod,m}$ is the ecdf T_{Mod} for month m, and $T_{Mod_{i,t}}$ is the (uncorrected) model value of TMAX at location i on day t. Monthly TFs were constructed using daily, spatially-explicit model and observed data from the calibration period for a given month. For each of the twelve TFs, 10,000 quantiles were estimated between 0 and 1; interpolation of the quantiles was accomplished with monotone Hermite splines using the `qmap` package [102] in R. For each month, the same TF was used to bias-correct all daily model data at the 78 station locations. Figure 4.9 shows an example of an EQM TF for month 12 and how it is used to transform model quantiles.

4.4.2 Evaluation of performance

The effectiveness of the process convolution approach for adjusting temporal dependence was assessed by comparing estimates of σ_{x_ℓ} and σ_w from PC models fitted to $\mathbf{W}_{Obs}$, $\mathbf{W}_{Mod}$, and $\tilde{\mathbf{W}}_{Mod}$. We report results on the SD, rather than variance scale for ease of interpretation. We also show results for PC models fitted to $\mathbf{W}_{Mod}$ and $\tilde{\mathbf{W}}_{Mod}$ over the four time periods (1976-2005, 2006-2036, 2037-2067, and 2068-2099) to assess the stationarity of temporal dependence over time.

4.4.3 Evaluating effectiveness of bias-correction

We implemented DMTA and EQM with and without correction of temporal dependence so as to achieve a fair comparison between the two methods. DMTA without temporal adjustment can be accomplished simply by following the data processing

steps (section 4.3.3) and applying the final linear transformation (4.15) to spatially-explicit model data (the process convolution approach is omitted). The implementation of EQM *with* temporal adjustment is discussed in Supplementary Material. EQM without temporal adjustment was done using daily, spatially-explicit model and observed data (4.16).

To assess how well both EQM and DMTA performed with respect to correcting model biases, we implemented two 10-fold cross validations (to account for cases in which temporal dependence was and was not corrected) during the calibration period (1976-2005). In the cross-validations, test datasets always consisted of consecutive years. A cross-validation allows for determining the suitability of a bias-correction method for future data, as out-of-sample data can be viewed as proxies for future data.

Cross-validation Two separate cross-validations were carried out for model data bias-corrected by EQM and DMTA, with and without temporal adjustment. For DMTA, model data during the historical calibration period (1976-2005), as well as all 10 training sets of observed data were processed according to steps in section 4.3.3 and fitted to model 4.5. During cross-validation, estimates of $\sigma^2_{x_{\ell,Obs}}$ and $\sigma^2_{w,Obs}$ corresponding to each observed training set were used to adjust temporal dependence of model data in corresponding test sets following steps in section 4.3.5. $\boldsymbol{\alpha}_{Obs}$ and $\boldsymbol{\delta}_{Obs}$ were estimated from observed training sets and were used in the back-transformation of corresponding model data in test sets (4.14). Finally, (4.15) was applied to (spatially-explicit) model data in test sets. Cross-validation for EQM with temporal adjustment is discussed in Supplementary Material.

When temporal dependence was not adjusted, cross-validation was carried out as

follows. First, model data over the calibration period were processed as described in section 4.3.3. For DMTA, in each fold, $\boldsymbol{\alpha}_{Obs}$ and $\boldsymbol{\delta}_{Obs}$ were estimated from observed training sets and used in the back-transformation of model data (4.14) in test sets. For EQM, spatially-explicit model and observed data were designated into training and testing sets for each fold. Model data in test sets were bias-corrected on a monthly basis using observed and model data in training sets.

Distributional similarity (model skill) of bias-corrected model and observed data was quantified with the mean absolute error (MAE) (4.17). MAE is often used in the climate literature to assess the performance of bias-correction methods [61].

$$MAE = \frac{1}{n}\sum_{i=n}^{n}|O_i - M_i| \tag{4.17}$$

In (4.17), O_i is i^{th} observed quantile, M_i is i^{th} model (or bias-corrected model) quantile, and n is the total number of quantiles. MAE values were calculated using 1,000 estimated quantiles on the interval $[0, 1]$. MAE values were calculated by day of year (1-365), and final MAE values are reported as the average value over all 10 cross-validation folds. MAE is bounded on $[0, \infty)$, where lower values of MAE indicate greater distributional similarity to observed data.

Performance was computed between raw model and observed data as well as between bias-corrected model and observed data in test sets for each fold. Numerical models are typically evaluated on their ability to simulate long-term climatology [189]. Climate model output over the historical period cannot be compared to observations on a daily basis, so metrics that quantify distributional similarity were chosen, rather than error metrics that quantify day-to-day errors.

4.5 Results

4.5.1 Correction of temporal dependence

Estimates of $\sigma_{x_{180}}$ were very similar for model, temporally-adjusted model, and observed data. For unadjusted model data, estimates of $\sigma_{x_{3.5}}$ were slightly lower ($\sigma_{3.5} =$ 0.81 to 0.87) than that of observed data ($\sigma_{3.5} = 0.98$) (Table 4.1). The consistently higher estimate of $\sigma_{x_{3.5}}$ for observed compared to model data indicates that temporal dependence of observed data at a fine temporal resolution (3.5 days) is slightly greater than that of model data. Estimates of σ_{x_w} of model data ($\sigma_{x_w} = 0.7072 - 0.7705$) were slightly greater than that of observed data ($\sigma_{x_w} = 0.7049$) (Table 4.1). Following temporal adjustment, all estimates of $\sigma_{x_{3.5}}$ and σ_{x_w} were similar to those obtained from models fitted to observed data (Table 4.1).

The correction of temporal dependence of model data is based on estimates of σ_{x_ℓ} from observed data during the calibration period, 1976-2005. Thus, this approach assumes that the σ_{x_ℓ} are stationary over time. We found that estimates of σ_{x_ℓ} obtained from PC models fitted to model data varied little over the four time periods during which PC models were fitted. This indicates that the temporal dependence of model data appear to be stationary over time. We also investigated if other kinds of temporal dependencies appeared in future data, but we found no notable changes.

Table 4.1: Estimates of $\sigma_{x_{3.5}}$, $\sigma_{x_{180}}$, and σ_w from PC models fitted to processed observed (Obs), unadjusted model (Mod), temporally-adjusted Mod (ModTA) for historical (1976-2005) and future (2006-2036, 2037-2067, and 2068-2099) time periods.

Data type	$\sigma_{x_{3.5}}$	$\sigma_{x_{180}}$	σ_w
Obs:	0.9817	0.0002	0.7049
Mod: 1976-2005	0.8771	0.0011	0.7705
Mod: 2006-2036	0.8751	0.0002	0.7529
Mod: 2037-2067	0.8189	0.0002	0.7503
Mod: 2068-2099	0.8524	0.0003	0.7072
ModTA: 1976-2005	0.9757	0.0005	0.6999
ModTA: 2006-2036	0.9791	0.0002	0.6995
ModTA: 2037-2067	0.9840	0.0002	0.6985
ModTA: 2068-2099	0.9670	0.0002	0.7013

4.5.2 Correction of seasonal means and standard deviations

MAE values resulting from model data with and without the correction of temporal dependence did not differ appreciably (see Figure S2 in Supplementary Material), indicating that correction of temporal dependence of model data does not adversely affect the correction of distributional characteristics. For uncorrected, EQM-, and DMTA-corrected model data, MAE values decreased between days 1 and 199 and increased between days 200 and 365 (Figure 4.6).

MAE values for uncorrected model data were noticeably greater than those for EQM- and DMTA-corrected model data during colder periods of the year (between days 1 and 100 and days 280-365), indicating that in colder periods, the distribution of model data is less similar to observed data. In the warmest periods of the year (days 150-230), MAE values of model data were slightly lower than or equal to those of

EQM- and DMTA-corrected model data (Figure 4.6). Bias-correction via DMTA was very effective in reducing MAE values during colder periods of the year (days 1-100 and days 280-365). However, for the same periods, bias-correction via EQM resulted in higher MAE values than those obtained from uncorrected model data (Figure 4.6). In warmer periods of the year (days 101-249), bias-correction via DMTA had little impact on MAE, and during days 101-160, DMTA resulted in larger MAE values than uncorrected model data. EQM resulted in larger values of MAE compared to uncorrected model data over the entire warm period (days 101-249).

Overall, DMTA resulted in more consistent reductions in MAE compared to EQM and DMTA provided improvements over uncorrected model data for most days of the year. While EQM improved distributional properties of model data for the coldest periods of the year, EQM resulted in greater MAE values than uncorrected model data during warmer periods of the year. EQM was more likely to overfit on training data during warmer periods of the year compared to DMTA.

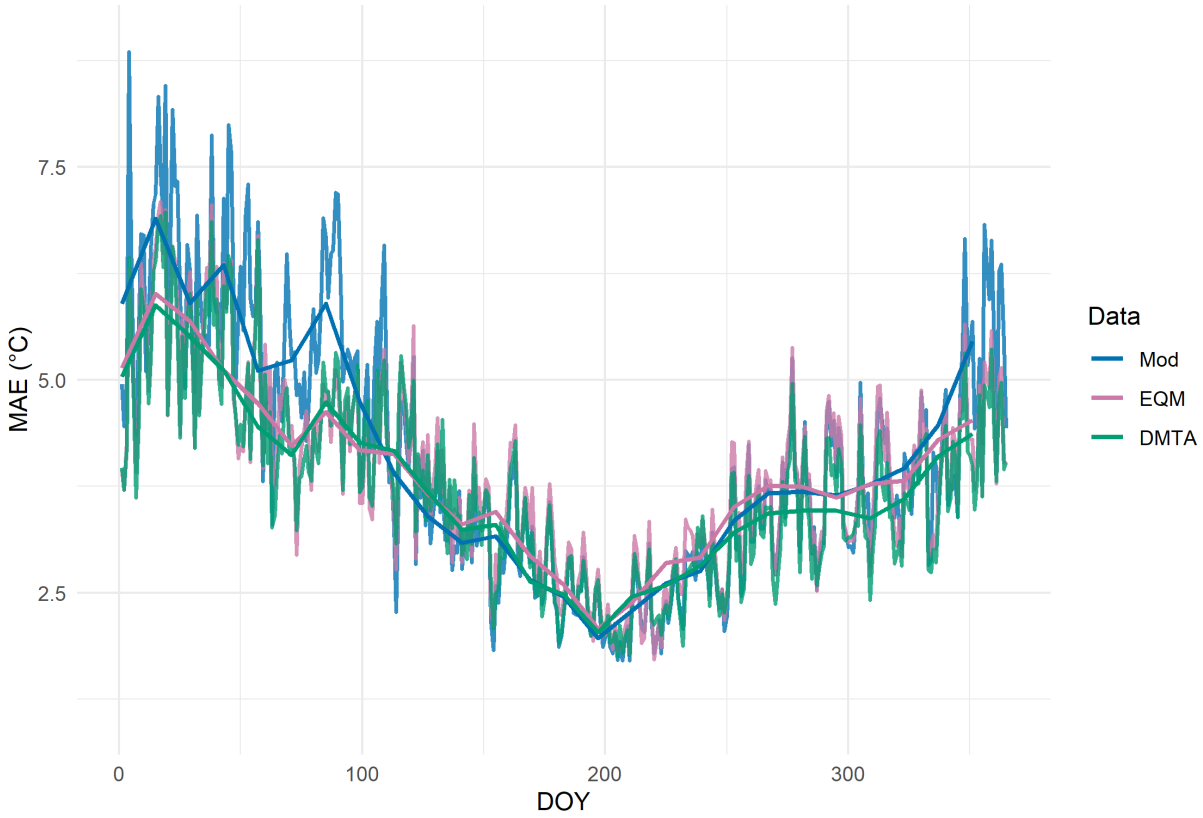

*Figure 4.6: Cross-validated results of MAE(°C) for uncorrected model (MOD), and EQM- and DMTA-corrected (*with *temporal adjustment) data relative to observed data during the historical period (1976-2005) over day of year (DOY). Averages over 14-day increments of DOY are overlaid to aid interpretation.*

Stationarity of seasonal means and standard deviations

While the stationarity of future model projections can generally not be assumed [47], [113], [116], many bias-correction methods, including EQM, work under this assumption [55]. Here we examine the stationarity of seasonal means and SDs in relation to bias-correction via DMTA and EQM. As shown in Figure 4.5, the daily mean of model simulations increases nonlinearly over the 124-year simulation period, Here, we examine the stationarity of seasonal means and SDs and how bias-correction

via DMTA and EQM affect the trend of seasonal means and SDs over time.

For uncorrected, EQM- and DMTA-corrected model data, the shapes of seasonal means do not change over all four time periods during which PC models were fitted, providing evidence that the seasonal mean of model data appears to be stationary over time (Figure 4.7 a, b, and c, respectively). Figure 4.7 also shows the seasonal mean for observed data for comparison (black dashed line). After bias-correction via DMTA and EQM, seasonal mean trends obtained over all four time periods very closely resemble the seasonal mean trend of observed data.

In contrast to seasonal means, the shape of seasonal SDs of model data changed considerably over time (Figure 4.8 a). In particular, the seasonal SD trend flattens slightly during colder periods of the year (approximately days 1-100 and days 300-365), indicating an overall decrease in the variability of daily TMAX (Figure 4.8 a). Though the "flattening" is apparent in the SD trends obtained from EQM-corrected model data (Figure 4.8 b), the overall shapes of the seasonal SDs over time are substantially different from those of observed data. In addition, the progressive decrease in the seasonal SD during colder periods over time is not preserved. In EQM-corrected data, the SD trend is "flattest" during the period 2036-2076 (Figure 4.8 b), while for uncorrected model data, the "flattest" SD trend occurs during the period 2067-2099 (Figure 4.8 a). After correction via DMTA the shapes of seasonal SDs over all four periods resemble those of observed data, and the same progressive "flattening" of the seasonal SD trend is preserved (Figure 4.8 c). Thus, not only does bias-correction via DMTA correct bias in the seasonal SD of model data, but it also preserves the "flattening" trend of the seasonal SD over time. Compared to EQM, DMTA better preserves the change in seasonal SD trends exhibited in uncorrected model data, while

also imparting characteristics of observed data.

These results imply that if the assumption of stationarity can be justified (as it can be for the seasonal mean), bias-correction methods such as EQM perform well. However, if the assumption of stationarity is violated (as it appears to be for seasonal SDs), DMTA can correct biases in means and SDs in addition to adequately preserving trends exhibited in uncorrected future model data.

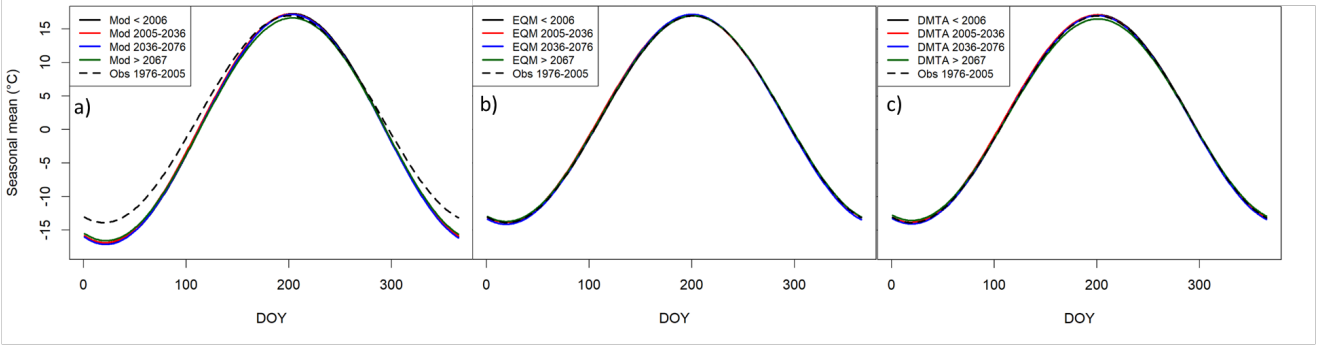

Figure 4.7: Seasonal means (°C) for uncorrected model data (a), EQM-corrected model data (b), and DMTA-corrected model data (c) over day of year (DOY) for time periods 1976-2005, 2006-2036, 2037-2067, and 2068-2099. Observed (Obs) seasonal means are denoted by the dashed black line.

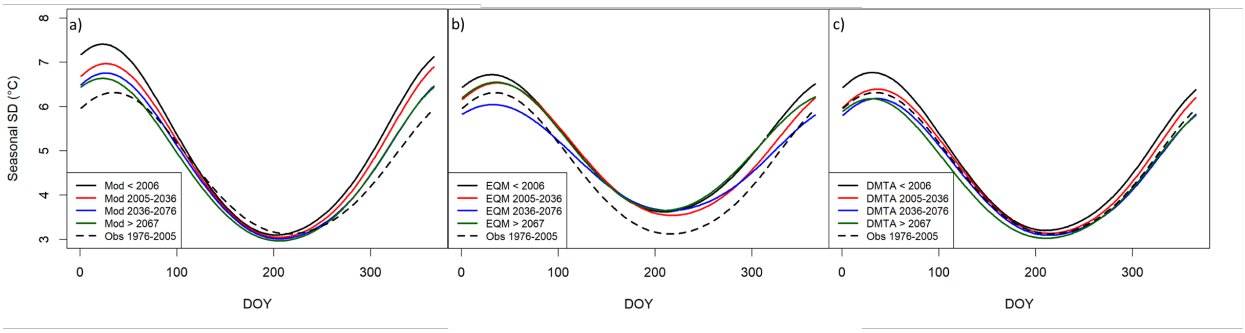

Figure 4.8: Seasonal standard deviations (SDs) (°C) for uncorrected model data (a), EQM-corrected model data (b), and DMTA-corrected model data (c) over day of year (DOY) for time periods 1976-2005, 2006-2036, 2037-2067, and 2068-2099. Observed (Obs) seasonal SDs are denoted by the dashed black line.

4.6 Discussion

In this study, we developed a novel method for bias-correcting simulations from a numerical climate model. The method features a process convolution approach for adjusting the temporal dependence of model data, and a series of linear transformations are used to correct the distribution of model data in both historical and future time periods. Additionally, the proposed method ensures that there is a smooth transition in correction from the calibration to future time period and important features of the future climate change signal (CCS), such as changes in the standard deviation, are preserved. Even without correcting a bias in temporal dependence, DMTA provides a more robust and realistic bias-correction of seasonal means and SDs compared to the widely-used quantile-mapping method EQM. DMTA is also less prone to overfitting on calibration data compared to EQM. While we apply the proposed method to daily maximum temperature simulations from a regional climate model, the process convolution approach for adjusting temporal dependence, together or separate from the distributional correction component, has a wide range of applications. There are several modeling aspects of the proposed bias-correction method that warrant further discussion. These aspects include computational resources and considerations for process convolution models, the selection of basis matrices in PC models and refining the temporal correction approach, and the need for an additional spatial correction.

The process convolution approach for adjusting temporal dependence of a signal is straightforward to implement. However, the approach has some limitations in terms of computational complexity. Process convolution models are computationally

expensive to fit, either as random effects models or in a Bayesian approach using a Gibbs sampler. Matrix inversion, necessary for both fitting methods, carries a computational cost of $O(n^3 \log_2 n)$. It is therefore not feasible to fit process convolution models with either large numbers ($>$ 30,000) of observations, or a large number and/or size of basis matrices. We found that fitting 30,000 observations with basis matrices representing short and long-term trends took approximately 8 hours on an NVIDIA DGX-1 system. Due to the high computational intensity of fitting PC models, it is important that only the most important component frequencies are represented in the basis matrices. We found choosing appropriate Gaussian kernel standard deviations for basis matrices to be a difficult task. A natural way to select appropriate kernel standard deviations would be a group shrinkage approach such as the group lasso or its Bayesian counterpart [207]. However, group lasso approaches were also computationally intensive and contradicted results from PC models fitted as random effects models and using a Gibbs sampler. We proceeded by fitting many PC models with various kernel standard deviations and retained those for which σ_x were nonzero. Further research is needed in the area of selecting kernel width parameters for PC models. Finally, because the temporal correction via a process convolution approach has never been done, there are some steps within the approach that should be refined. As described in section 4.3.5, the scaling of $\hat{\boldsymbol{\mu}}_{Mod}$ via ratios of the estimated σ_{x_ℓ} is not linear [206] and may require an additional scalar (c^*) to achieve the desired change in temporal dependence. Here, we tested several values of c^* before selecting the appropriate value. Further extensions to the approach could include a better way to select appropriate scaling constants during the temporal correction process.

DMTA provides the same correction, regardless of spatial location. Thus, when

spatial autocorelation is encountered the need for an additional spatial correction may arise. We considered the development of a secondary model to provide an additional spatial correction, but for our study area we did not find evidence of any spatial autocorrelation present in model data before or after bias-correction. The lack of residual spatial autocorrelation could be due to several factors. First, the WRF model accounts for topography and climate processes at the 4km scale and is therefore capable of capturing relatively fine-scale processes that could affect temperature (e.g. the effect of large bodies of water, elevation, topographical aspect, etc). Additionally, the use of topographical downscaling further adjusts WRF model data to represent the effects of fine-scale elevation. We surmise that for much larger study areas, an additional spatial model (such as that described in Appendix A4) may be necessary to correct biases that change over space in large-scale, complex landscapes.

While an in-depth investigation of the correction of extremes is outside the scope of this study, our results highlight some important issues related to the treatment of extremes during bias-correction. The proposed method relies on the climate model to accurately simulate extremes. While RCMs may be limited in their ability to capture extreme distributional tails of climate variables [44], [162], we did not incorporate an additional adjustment for extremes as we felt it would be problematic to adjust future extremes with the limited amount of data available for the calibration period. The treatment of extremes within DMTA warrants further study, and we are investigating this as part of future research.

The treatment of future extremes by EQM is also problematic. In EQM, the shape of the TF for extreme quantiles (> 95) can be very different from the shape of the TF at lower quantiles (e.g. Figure 4.9). Since observations representing extreme

tails are scarce and highly variable, the TF produced by EQM may be unstable at its tails. Consequently, the use of EQM for correcting future climate simulations can cause distortion of the CCS [47], [144] and statistical artifacts [208]. Likewise, bias-correction via EQM introduced inflated extremes in future data when the shape of the TF at upper extreme quantiles was very different from the majority of the TF (see Supplementary materials and Figures S2 and S3). While the effect of EQM tail TFs on extremes is not apparent in the correction of historical data, this may result in artifacts in future data that are a direct consequence of the shape of tail TFs, as well as the fact that there is a greater frequency of higher temperature values in the future.

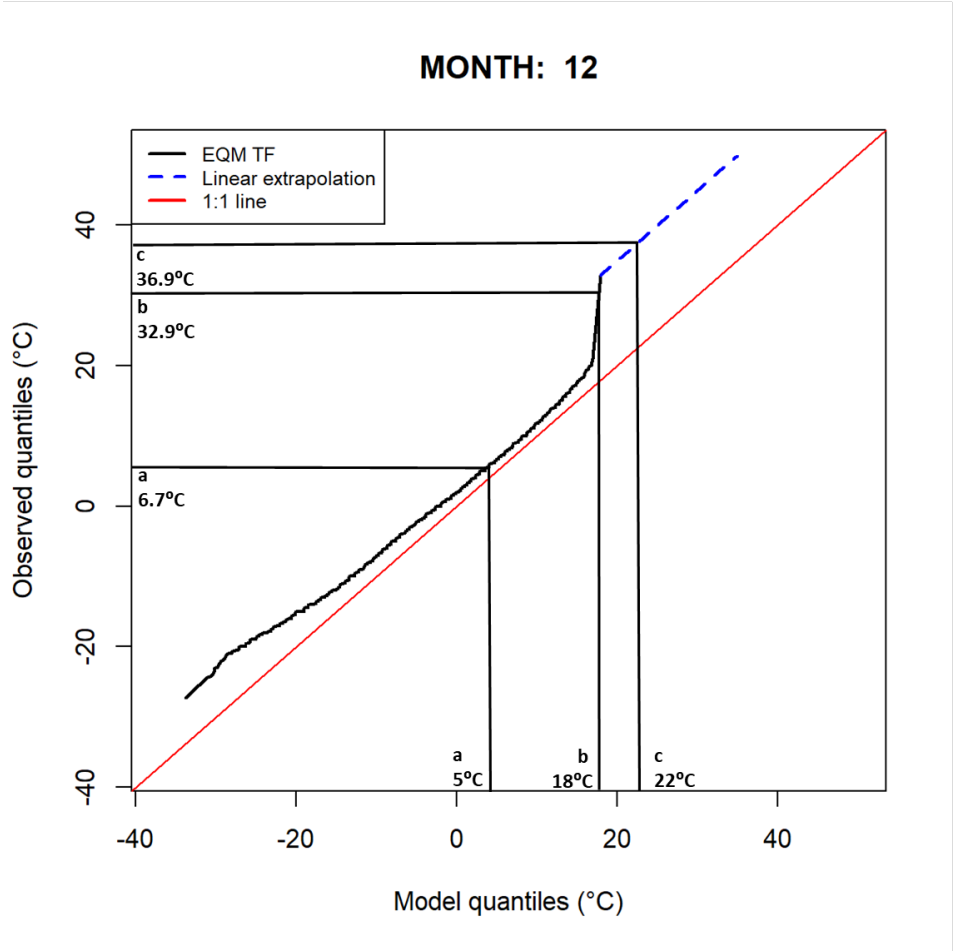

Figure 4.9: EQM transfer function (TF) (black line) for month 12 constructed using 1000 quantiles of observed and model data during the historical period (1976-2005). The shape of the TF beyond the range of data values is shown by the blue dashed line. The lines at a, b, and c show examples of how much model values are changed via the TF. At a), the TF increases a model value of 5° C to 6.7° C. At b) a model value increases from 18° C to 32.9° C, and at c) a model value of 22° C increases to 36.9° C. The correction at b) results in much larger increases in model values between 17.9 and 32.8° C, as the shape of the TF increases sharply. Furthermore, the correction at b) is unreliable, as that portion of the TF is based on interpolation of the maximum quantiles of model and observed data.

References

[4] A. Zia, A. Bomblies, A. W. Schroth, *et al.*, "Coupled impacts of climate and land use change across a river–lake continuum: Insights from an integrated assessment model of lake champlain's missisquoi basin, 2000–2040," *Environmental Research Letters*, vol. 11, no. 11, p. 114 026, 2016.

[5] M. Ekström, M. R. Grose, and P. H. Whetton, "An appraisal of downscaling methods used in climate change research," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 6, no. 3, pp. 301–319, 2015.

[7] D. Maraun, T. G. Shepherd, M. Widmann, *et al.*, "Towards process-informed bias correction of climate change simulations," *Nature Climate Change*, vol. 7, no. 11, p. 764, 2017.

[8] T. Lafon, S. Dadson, G. Buys, and C. Prudhomme, "Bias correction of daily precipitation simulated by a regional climate model: A comparison of methods," *International Journal of Climatology*, vol. 33, no. 6, pp. 1367–1381, 2013.

[9] A. J. Cannon, C. Piani, and S. Sippel, "Bias correction of climate model output for impact models," in *Climate Extremes and Their Implications for Impact and Risk Assessment*, Elsevier, 2020, pp. 77–104.

[10] I. Hanssen-Bauer, C. Achberger, R. Benestad, D. Chen, and E. Førland, "Statistical downscaling of climate scenarios over Scandinavia," *Climate Research*, vol. 29, no. 3, pp. 255–268, 2005.

[11] Z. A. Holden, J. T. Abatzoglou, C. H. Luce, and L. S. Baggett, "Empirical downscaling of daily minimum air temperature at very fine resolutions in complex terrain," *Agricultural and Forest Meteorology*, vol. 151, no. 8, pp. 1066–1073, 2011.

[14] R. L. Wilby, S. Charles, E. Zorita, B. Timbal, P. Whetton, and L. Mearns, "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 TGCIA*, vol. 27, 2004.

[18] D. Maraun, "Bias correcting climate change simulations-a critical review," *Current Climate Change Reports*, vol. 2, no. 4, pp. 211–220, 2016.

[21] L. Mearns, F. Giorgi, P. Whetton, D. Pabon, M. Hulme, and M. Lal, "Guidelines for use of climate scenarios developed from regional climate model experiments," *Data Distribution Centre of the Intergovernmental Panel on Climate Change*, 2003.

[22] G. Fang, J. Yang, Y. Chen, and C. Zammit, "Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in china," *Hydrology and Earth System Sciences*, vol. 19, no. 6, pp. 2547–2559, 2015.

[31] J. M. Winter, B. Beckage, G. Bucini, R. M. Horton, and P. J. Clemins, "Development and evaluation of high-resolution climate simulations over the mountainous northeastern united states," *Journal of Hydrometeorology*, vol. 17, no. 3, pp. 881–896, 2016.

[40] R. M. B. Harris, M. R. Grose, G. Lee, N. L. Bindoff, L. L. Porfirio, and P. Fox-Hughes, "Climate projections for ecologists," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 5, no. 5, pp. 621–637, 2014.

[44] R. Leander and T. A. Buishand, "Resampling of regional climate model output for the simulation of extreme river flows," *Journal of Hydrology*, vol. 332, no. 3-4, pp. 487–496, 2007.

[46] C. Teutschbein and J. Seibert, "Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods," *Journal of Hydrology*, vol. 456, pp. 12–29, 2012.

[47] A. J. Cannon, S. R. Sobie, and T. Q. Murdock, "Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?" *Journal of Climate*, vol. 28, no. 17, pp. 6938–6959, 2015.

[53] M. Jakob Themeßl, A. Gobiet, and A. Leuprecht, "Empirical-statistical downscaling and error correction of daily precipitation from regional climate models," *International Journal of Climatology*, vol. 31, no. 10, pp. 1530–1544, 2011.

[55] D. R. Roberts, W. H. Wood, and S. J. Marshall, "Assessments of downscaled climate data with a high-resolution weather station network reveal consistent but predictable bias," *International Journal of Climatology*, vol. 39, no. 6, pp. 3091–3103, 2019.

[61] L. Gudmundsson, J. Bremnes, J. Haugen, and T. Engen-Skaugen, "Downscaling rcm precipitation to the station scale using statistical transformations–a comparison of methods," *Hydrology and Earth System Sciences*, vol. 16, no. 9, pp. 3383–3390, 2012.

[63] J.-H. Heo, H. Ahn, J.-Y. Shin, T. R. Kjeldsen, and C. Jeong, "Probability distributions for a quantile mapping technique for a bias correction of precipitation data: A case study to precipitation data under climate change," *Water*, vol. 11, no. 7, p. 1475, 2019.

[66] O. Gutjahr and G. Heinemann, "Comparing precipitation bias correction methods for high-resolution regional climate simulations using cosmo-clm," *Theoretical and applied climatology*, vol. 114, no. 3, pp. 511–529, 2013.

[73] C. Miao, L. Su, Q. Sun, and Q. Duan, "A nonstationary bias-correction technique to remove bias in gcm simulations," *Journal of Geophysical Research: Atmospheres*, vol. 121, no. 10, pp. 5718–5735, 2016.

[74] N. Oceanic and A. Administration, *Climate data online search*, Accessed: 2017-09-30, 2018. [Online]. Available: https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND.

[76] B. Livneh, T. J. Bohn, D. W. Pierce, *et al.*, "A spatially comprehensive, hydrometeorological data set for mexico, the us, and southern canada 1950–2013," *Scientific data*, vol. 2, no. 1, pp. 1–12, 2015.

[77] P. E. Thornton, M. M. Thornton, B. W. Mayer, *et al.*, "Daymet: Daily surface weather on a 1 km grid for north america, 1980-2008," *Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center for Biogeochemical Dynamics (DAAC)*, 2012.

[78] C. Daly, G. Taylor, W. Gibson, T. Parzybok, G. Johnson, and P. Pasteris, "High-quality spatial climate data sets for the united states and beyond," *Transactions of the ASAE*, vol. 43, no. 6, p. 1957, 2000.

[79] R. Behnke, S. Vavrus, A. Allstadt, T. Albright, W. E. Thogmartin, and V. C. Radeloff, "Evaluation of downscaled, gridded climate data for the conterminous united states," *Ecological applications*, vol. 26, no. 5, pp. 1338–1351, 2016.

[80] D. Walton and A. Hall, "An assessment of high-resolution gridded temperature datasets over california," *Journal of Climate*, vol. 31, no. 10, pp. 3789–3810, 2018.

[82] T. C. Peterson and R. S. Vose, "An overview of the global historical climatology network temperature database," *Bulletin of the American Meteorological Society*, vol. 78, no. 12, pp. 2837–2850, 1997.

[83] I. Durre, M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, "Comprehensive automated quality assurance of daily surface observations," *Journal of Applied Meteorology and Climatology*, vol. 49, no. 8, pp. 1615–1633, 2010.

[86] J. Rajczak, S. Kotlarski, N. Salzmann, and C. Schaer, "Robust climate scenarios for sites with sparse observations: A two-step bias correction approach," *International Journal of Climatology*, vol. 36, no. 3, pp. 1226–1243, 2016.

[91] F. Johnson and A. Sharma, "A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations," *Water Resources Research*, vol. 48, no. 1, 2012.

[92] F. Johnson and A. Sharma, "Assessing future droughts in australia-a nesting model to correct for long-term persistence in general circulation model precipitation simulations," in *8th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation*, Modell., Simul. Soc. of Aust., and NZ, Cairns Queensland, Australia, 2009, pp. 3935–3941.

[93] F. Johnson, S. Westra, A. Sharma, and A. J. Pitman, "An assessment of gcm skill in simulating persistence across multiple time scales," *Journal of Climate*, vol. 24, no. 14, pp. 3609–3623, 2011.

[94] R. Mehrotra and A. Sharma, "An improved standardization procedure to remove systematic low frequency variability biases in gcm simulations," *Water Resources Research*, vol. 48, no. 12, 2012.

[95] H. Nguyen, R. Mehrotra, and A. Sharma, "Correcting for systematic biases in gcm simulations in the frequency domain," *Journal of Hydrology*, vol. 538, pp. 117–126, 2016.

[96] C. Kusumastuti, Z. Jiang, R. Mehrotra, and A. Sharma, "A signal processing approach to correct systematic bias in trend and variability in climate model simulations," *Geophysical Research Letters*, vol. 48, no. 13, e2021GL092953, 2021.

[99] M. G. Grillakis, A. G. Koutroulis, and I. K. Tsanis, "Multisegment statistical bias correction of daily gcm precipitation output," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 8, pp. 3150–3162, 2013.

[100] M. F. Holthuijzen, B. Beckage, P. J. Clemins, D. Higdon, and J. M. Winter, "Constructing high-resolution, bias-corrected climate products: A comparison of methods," *Journal of Applied Meteorology and Climatology*, vol. 60, no. 4, pp. 455–475, 2021.

[102] L. Gudmundsson, *Qmap: Statistical transformations for post-processing climate model output*, R package version 1.0-4, 2016.

[105] M. G. Grillakis, A. G. Koutroulis, I. N. Daliakopoulos, and I. K. Tsanis, "A method to preserve trends in quantile mapping bias correction of climate modeled temperature," *Earth System Dynamics*, vol. 8, no. 3, p. 889, 2017.

[113] M. B. Switanek, P. A. Troch, C. L. Castro, *et al.*, "Scaled distribution mapping: A bias correction method that preserves raw climate model projected changes," *Hydrology and Earth System Sciences*, vol. 21, no. 6, pp. 2649–2666, 2017.

[114] S. Hagemann, C. Chen, J. O. Haerter, J. Heinke, D. Gerten, and C. Piani, "Impact of a statistical bias correction on the projected hydrological changes obtained from three gcms and two hydrology models," *Journal of Hydrometeorology*, vol. 12, no. 4, pp. 556–578, 2011.

[115] M. A. Ivanov, J. Luterbacher, and S. Kotlarski, "Climate model biases and modification of the climate change signal by intensity-dependent bias correction," *Journal of Climate*, vol. 31, no. 16, pp. 6591–6610, 2018.

[116] H. Li, J. Sheffield, and E. F. Wood, "Bias correction of monthly precipitation and temperature fields from intergovernmental panel on climate change ar4 models using equidistant quantile matching," *Journal of Geophysical Research: Atmospheres*, vol. 115, no. D10, 2010.

[121] J. D. Fridley, "Downscaling climate over complex terrain: High finescale (less than 1000 m) spatial variation of near-ground temperatures in a montane forested landscape (great smoky mountains)," *Journal of Applied Meteorology and Climatology*, vol. 48, no. 5, pp. 1033–1049, 2009.

[122] D. Maraun, F. Wetterhall, A. Ireson, *et al.*, "Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user," *Reviews of Geophysics*, vol. 48, no. 3, 2010.

[128] W. C. Skamarock, J. B. Klemp, J. Dudhia, *et al.*, "A description of the advanced research wrf model," 2019, Accessed: 2019-03-04. [Online]. Available: https://opensky.ucar.edu/islandora/object/opensky:2898.

[130] H. Huang, J. M. Winter, E. C. Osterberg, *et al.*, "Simulating precipitation and temperature in the lake champlain basin using a regional climate model: Limitations and uncertainties," *Climate Dynamics*, vol. 54, no. 1-2, pp. 69–84, 2020.

[131] USGS, *The national map*, 2018. [Online]. Available: https://viewer.nationalmap.gov/basic/.

[132] T. Wang, A. Hamann, D. L. Spittlehouse, and T. Q. Murdock, "ClimateWNA—high-resolution spatial climate data for western North America," *Journal of Applied Meteorology and Climatology*, vol. 51, no. 1, pp. 16–29, 2012.

[144] P. Berg, H. Feldmann, and H.-J. Panitz, "Bias correction of high resolution regional climate model data," *Journal of Hydrology*, vol. 448, pp. 80–92, 2012.

[162] G. A. Baigorria, J. W. Jones, D.-W. Shin, A. Mishra, and J. J. O'Brien, "Assessing uncertainties in crop model simulations using daily bias-corrected regional circulation model outputs," *Climate Research*, vol. 34, no. 3, pp. 211–222, 2007.

[165] M. Enayati, O. Bozorg-Haddad, J. Bazrafshan, S. Hejabi, and X. Chu, "Bias correction capabilities of quantile mapping methods for rainfall and temperature variables," *Journal of Water and Climate Change*, vol. 12, no. 2, pp. 401–419, 2021.

[167] A. Monaghan, D. Steinhoff, C. Bruyere, and D. Yates, "Ncar cesm global bias-corrected cmip5 output to support wrf/mpas research," *Research Data Archive National Center Atmospheric Research Computational Information Systems Laboratory, Boulder. DOI*, vol. 10, D6DJ5CN4, 2014.

[181] V. J. Berrocal, A. E. Gelfand, and D. M. Holland, "A spatio-temporal downscaler for output from numerical models," *Journal of agricultural, biological, and environmental statistics*, vol. 15, no. 2, pp. 176–197, 2010.

[182] M. Kiel, C. W. O'Dell, B. Fisher, *et al.*, "How bias correction goes wrong: Measurement of x co 2 affected by erroneous surface pressure estimates," *Atmospheric Measurement Techniques*, vol. 12, no. 4, pp. 2241–2259, 2019.

[183] F.-Y. Cheng, C.-Y. Feng, Z.-M. Yang, *et al.*, "Evaluation of real-time pm2. 5 forecasts with the wrf-cmaq modeling system and weather-pattern-dependent bias-adjusted pm2. 5 forecasts in taiwan," *Atmospheric Environment*, vol. 244, p. 117 909, 2021.

[184] J. Zhang, P. Crippa, M. G. Genton, and S. Castruccio, "Assessing the reliability of wind power operations under a changing climate with a non-gaussian bias correction," *The Annals of Applied Statistics*, vol. 15, no. 4, pp. 1831–1849, 2021.

[185] K. Steele and C. Werndl, "Climate models, calibration, and confirmation," *The British Journal for the Philosophy of Science*, vol. 64, no. 3, pp. 609–635, 2013.

[186] D. Maraun and M. Widmann, *Statistical downscaling and bias correction for climate research*. Cambridge University Press, 2018.

[187] K. Tabor and J. W. Williams, "Globally downscaled climate projections for assessing the conservation impacts of climate change," *Ecological Applications*, vol. 20, no. 2, pp. 554–565, 2010.

[188] R. J. MacDonald, S. Boon, J. M. Byrne, M. D. Robinson, and J. B. Rasmussen, "Potential future climate effects on mountain hydrology, stream temperature, and native salmonid life history," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 71, no. 2, pp. 189–202, 2014.

[189] U. Ehret, E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert, "Hess opinions" should we apply bias correction to global and regional climate model data?"," *Hydrology and Earth System Sciences*, vol. 16, no. 9, pp. 3391–3404, 2012.

[190] D. Maraun, "Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue," *Journal of Climate*, vol. 26, no. 6, pp. 2137–2143, 2013.

[191] D. Maraun, R. Huth, J. M. Gutierrez, *et al.*, "The value perfect predictor experiment: Evaluation of temporal variability," *International Journal of Climatology*, vol. 39, no. 9, pp. 3786–3818, 2019.

[192] D. Higdon, "A process-convolution approach to modelling temperatures in the north atlantic ocean," *Environmental and Ecological Statistics*, vol. 5, no. 2, pp. 173–190, 1998.

[193] D. Higdon, J. Swall, and J. Kern, "Non-stationary spatial modeling," *Bayesian statistics*, vol. 6, no. 1, pp. 761–768, 1999.

[194] C. A. Calder, "Dynamic factor process convolution models for multivariate space–time data with application to air quality assessment," *Environmental and Ecological Statistics*, vol. 14, no. 3, pp. 229–247, 2007.

[195] S. Herrera, S. Kotlarski, P. M. Soares, *et al.*, "Uncertainty in gridded precipitation products: Influence of station density, interpolation method and grid resolution," *International Journal of Climatology*, vol. 39, no. 9, pp. 3717–3729, 2019.

[196] D. P. Van Vuuren, J. Edmonds, M. Kainuma, *et al.*, "The representative concentration pathways: An overview," *Climatic change*, vol. 109, no. 1, pp. 5–31, 2011.

[197] S. N. Wood, *Generalized additive models: an introduction with R*. CRC press, 2017.

[198] R. B. Gramacy, *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*. CRC Press, 2020.

[199] M. Binois and R. B. Gramacy, "Hetgp: Heteroskedastic gaussian process modeling and sequential design in r," *Journal of Statistical Software*, vol. 98, pp. 1–44, 2021.

[200] B. Ankenman, B. L. Nelson, and J. Staum, "Stochastic kriging for simulation metamodeling," in *2008 Winter Simulation Conference*, IEEE, 2008, pp. 362–370.

[201] A. Graps, "An introduction to wavelets," *IEEE computational science and engineering*, vol. 2, no. 2, pp. 50–61, 1995.

[202] D. Higdon, "Space and space-time modeling using process convolutions," in *Quantitative methods for current environmental issues*, Springer, 2002, pp. 37–56.

[203] C. A. Calder and N. Cressie, "Some topics in convolution-based spatial modeling," *Proceedings of the 56th Session of the International Statistics Institute*, pp. 22–29, 2007.

[204] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*. CRC press, 2013.

[205] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Core Team, *nlme: Linear and nonlinear mixed effects models*, R package version 3.1-144, 2020. [Online]. Available: https://CRAN.R-project.org/package=nlme.

[206] S. M. Butler and T. A. Louis, "Random effects models with non-parametric priors," *Statistics in medicine*, vol. 11, no. 14-15, pp. 1981–2000, 1992.

[207] X. Xu, M. Ghosh, *et al.*, "Bayesian variable selection and estimation for group lasso," *Bayesian Analysis*, vol. 10, no. 4, pp. 909–936, 2015.

[208] M. Holthuijzen, B. Beckage, D. Higdon, J. P. Clemins, and J. Winter, "A comparison of bias correction procedures for maximum daily temperature using sparsely distributed station data: Applications to high resolution downscaling," *Journal of Theoretical and Applied Climatology*, 2022.

Chapter 5

Conclusion

High-resolution, bias-corrected climate data products are used in a wide variety of disciplines, including agriculture, ecology, biology, and economics. General circulation models (GCMs) as well as regional climate models are very important for predicting past and future climate, but they cannot adequately capture local climate processes. Downscaling and bias-correction are typically applied to climate model output prior to its use in downstream modeling applications. Downscaling transforms data from a coarse spatial scale to a fine spatial scale, whereas bias-correction procedures adjust climate model output such that its statistical properties resemble those of observations. In this body of work, I developed computational workflows for constructing high-resolution, bias-corrected climate data products. Within these computational workflows, I addressed key issues related to bias-correction: spatial coherence, temporal coherence, and the handling of extremes.

In spatially coherent bias-correction, the correction is adjusted depending on the spatial location of model data. Spatially-coherent bias-correction can be very challenging if observed data is derived from climate stations, which are generally not

evenly distributed over spatial regions. I developed simple and easy-to-implement high-resolution computational workflows for downscaling and bias-correction in which observed data is derived from sparsely distributed climate stations (Chapter 2). In these workflows, bias-correction occurs either 1) at station locations, and resulting corrected model data is interpolated to a fine-scale grid, or 2) at the fine-scale grid using model and observed data that have been interpolated to the fine-scale grid. I also compared two bias-correction methods, empirical quantile mapping (EQM) and quantile mapping with linear transfer (LT) functions, in combination with two downscaling techniques, kriging and topographical downscaling (a variation of inverse distance weighting). Performance was most dependent on the efficacy of the bias-correction method and less dependent on the downscaling method. Downscaling using topographical downscaling was as accurate as the more sophisticated geostatistical method kriging.

The most appropriate bias-correction method for the construction of high-resolution temperature data products using station data depends on the intended use of the resulting data product. For historical products, improving day-to-day correspondence of simulated and observed data is likely to be of interest. This was achieved most effectively by bias-correction at the fine-scale grid using transfer functions obtained from temporally-ordered regressions generated at station locations. However, to achieve distributional similarity of simulated and observed data in historical data products, EQM performed best. For future data products, quantile mapping using simple LT functions constructed through rank-ordered regression provided more robust bias-correction of temperature variables than EQM, which is prone to overfitting on calibration data.

The application of spatially-coherent bias-correction is likely dependent on regional fine-scale topography, the size of the study area, the type of meteorological variable, the spatial resolution of model simulations, and unique, fine-scale landscape characteristics. The application of spatially-coherent bias-correction is relatively straightforward for well-behaved variables such as temperature. For censored and highly skewed variables such as precipitation, the process may become more complicated. For instance, interpolation methods must be modified to avoid negative values and to preserve the highly skewed nature of precipitation variables.

For larger study areas or those with complex regional topography, spatially-coherent bias correction would likely result in more accurate data products. Spatially coherent bias-correction would also improve the quality of data products when model simulations are only available at larger spatial scales ($>$ 4km). Spatially coherent bias-correction could help fill the gap between model output at coarser spatial resolutions and fine-scale topography and orography. Finally, more sophisticated spatial models, such as the Gaussian spatiotemporal model described in A4, may provide better uncertainty quantification regarding the bias-correction. However, some issues remain regarding the consistent estimation of spatial and temporal variance parameters and requires further work.

Temporal dependence (the correlation of past and future values) of climate impacts persistence attributes such as the length of hot and cold spells and dry and wet spells and also the frequency of extreme events. Temporal variability can also influence persistence attributes such as lag 1 and lag 2 autocorrelations and the distribution of $n-$day moving averages. These persistence attributes are extremely important for predicting the effects of climate change on agricultural production, farming, human

health, and energy. In chapter 4 I developed a method for temporally-coherent bias-correction that features a process convolution approach for correcting bias in temporal dependence, and a series of linear transformations are used to correct the distribution of model data in both historical and future time periods. Process convolution models can provide insight into the component frequencies of model and observed time series. In this approach, process convolution models are fitted to detrended, spatially averaged time series of model and observed data. The model time series is adjusted such that the temporal dependence matches that of the observed time series. Next, the adjusted model time series is back- transformed using seasonal means and standard deviations obtained from observed data. Finally, the spatially averaged, corrected model time series is used to correct spatially explicit model data. As part of a high-resolution downscaling and bias-correction workflow, I applied the proposed bias-correction method to historical (1976-2005) and future (2006-2099) maximum temperature simulations generated by an RCM. I compared the performances of the proposed method and EQM in correcting distributional discrepancies of model output using a cross-validation procedure. I found that after adjustment via the process convolution approach, temporal dependence structures of model output resembled those of observed data. The proposed method also corrected biases of means without the risk of overfitting. Moreover, trends in seasonal means and standard deviations of future model data were better preserved using the proposed method, while bias-correction via EQM caused statistical artifacts at extreme quantiles and resulted in a distortion of the seasonal standard deviation trend over time.

While the process of adjusting temporal dependence is straightforward, extensions to the process could increase its computational efficiency and ease of use. The pro-

posed bias-correction method was designed specifically for temperature variables. A process convolution approach may not be appropriate for highly skewed meteorological variables such as precipitation. Also, the type of temporal dependence (periodic or non-periodic) as well as length scale parameters of the kernel bases used in the process convolution model must be selected by the modeler. Suitable techniques for selecting appropriate kernel parameters are group shrinkage methods such as the group lasso or its Bayesian counterpart. However, I found that group lasso approaches were also computationally intensive and contradicted results from process convolution models fitted as random effects models. Thus, the fitting of process convolution models could be expedited by automating kernel parameter selection. It may also be possible to perform temporally coherent bias- correction using other signal processing techniques such as the discrete Fourier transform or discrete wavelet transform. These techniques are more efficient than process convolution modeling, but the adjustment of temporal dependence via modification of Fourier or wavelet coefficients is not straightforward. Further work would be necessary to incorporate these alternative techniques in a bias-correction method.

Since extremes have a disproportionately large effect on downstream modeling results, it is important that extremes are accurately represented in climate data products. EQM can generally correct distributional discrepancies between simulated climate variables and observed data. However, EQM may be sensitive to the choice of calibration period and is prone to overfitting, especially at extremes where data is scarce and highly variable. The handling of extremes is especially important for highly skewed variables such as precipitation. At distributional extremes, the transfer functions produced by EQM may be unstable, resulting in the inflation of deflation of

extremes and the introduction of additional uncertainty. To bias-correct precipitation extremes, hybrid quantile-mapping approaches have been proposed where extreme tails are modeled using heavy tailed distributions such as the generalized Pareto or the exponential. The drawback of such hybrid parametric methods is that the tails of model and observed data are assumed to follow a known distribution. If no known distribution can be adequately fitted to the extreme tails of model and observed data, bias-correction will be unreliable. I developed a different hybrid bias-correction method (EQM-LIN) that combines the efficacy of EQM for correcting lower quantiles with a robust linear correction for upper quantiles (Chapter 3). I applied both EQM and EQM-LIN to historical daily precipitation data simulated by an RCM over a region in the northeastern United States. I found that bias-correction via EQM-LIN resulted in an accurate representation of both mean and extreme precipitation. The linear correction proved resistant to overfitting on calibration data and provided a reliable correction at extreme tails. EQM-LIN outperformed EQM in terms of model skill and performed at least as well or better than EQM with respect to several climatological indices. In this research, EQM-LIN was applied to daily precipitation simulations, but we expect it could be applied to other meteorological variables without further modification. Further extensions to EQM-LIN could include an adjustment of the slope of the linear correction.

The construction of high-resolution, bias-corrected climate data products will become increasingly more important for modeling efforts in many different disciplines as climate change advances with more extreme weather conditions. In addition, such climate products will be indispensable for the development and implementation of strategies to mitigate and adapt to a changing climate. This research provides com-

putational workflows and methodologies that facilitate quality construction of reliable climate data products at fine spatial resolutions.

References

•

[1] L. C. B. Program, *Recreation resources*, Accessed: 2019-09-06, 2019. [Online]. Available: https://www.lcbp.org/culture-recreation/recreation-resources/.

[2] D. N. Karger, O. Conrad, J. Böhner, *et al.*, "Climatologies at high resolution for the earth's land surface areas," *Scientific data*, vol. 4, no. 1, pp. 1–20, 2017.

[3] A. Hamann and T. Wang, "Potential effects of climate change on ecosystem and tree species distribution in british columbia," *Ecology*, vol. 87, no. 11, pp. 2773–2786, 2006.

[4] A. Zia, A. Bomblies, A. W. Schroth, *et al.*, "Coupled impacts of climate and land use change across a river–lake continuum: Insights from an integrated assessment model of lake champlain's missisquoi basin, 2000–2040," *Environmental Research Letters*, vol. 11, no. 11, p. 114 026, 2016.

[5] M. Ekström, M. R. Grose, and P. H. Whetton, "An appraisal of downscaling methods used in climate change research," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 6, no. 3, pp. 301–319, 2015.

[6] L. R. Leung, L. O. Mearns, F. Giorgi, and R. L. Wilby, "Regional climate research: Needs and opportunities," *Bulletin of the American Meteorological Society*, vol. 84, no. 1, pp. 89–95, 2003.

[7] D. Maraun, T. G. Shepherd, M. Widmann, *et al.*, "Towards process-informed bias correction of climate change simulations," *Nature Climate Change*, vol. 7, no. 11, p. 764, 2017.

[8] T. Lafon, S. Dadson, G. Buys, and C. Prudhomme, "Bias correction of daily precipitation simulated by a regional climate model: A comparison of methods," *International Journal of Climatology*, vol. 33, no. 6, pp. 1367–1381, 2013.

[9] A. J. Cannon, C. Piani, and S. Sippel, "Bias correction of climate model output for impact models," in *Climate Extremes and Their Implications for Impact and Risk Assessment*, Elsevier, 2020, pp. 77–104.

[10] I. Hanssen-Bauer, C. Achberger, R. Benestad, D. Chen, and E. Førland, "Statistical downscaling of climate scenarios over Scandinavia," *Climate Research*, vol. 29, no. 3, pp. 255–268, 2005.

[11] Z. A. Holden, J. T. Abatzoglou, C. H. Luce, and L. S. Baggett, "Empirical downscaling of daily minimum air temperature at very fine resolutions in complex terrain," *Agricultural and Forest Meteorology*, vol. 151, no. 8, pp. 1066–1073, 2011.

[12] G. E. Liston and K. Elder, "A meteorological distribution system for high-resolution terrestrial modeling (micromet)," *Journal of Hydrometeorology*, vol. 7, no. 2, pp. 217–234, 2006.

[13] P. Caldwell, H.-N. S. Chin, D. C. Bader, and G. Bala, "Evaluation of a WRF dynamical downscaling simulation over California," *Climatic change*, vol. 95, no. 3-4, pp. 499–521, 2009.

[14] R. L. Wilby, S. Charles, E. Zorita, B. Timbal, P. Whetton, and L. Mearns, "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 TGCIA*, vol. 27, 2004.

[15] F. Feser, B. Rockel, H. von Storch, J. Winterfeldt, and M. Zahn, "Regional climate models add value to global model data: A review and selected examples," *Bulletin of the American Meteorological Society*, vol. 92, no. 9, pp. 1181–1192, 2011.

[16] F. Giorgi, C. Jones, G. R. Asrar, *et al.*, "Addressing climate information needs at the regional level: The cordex framework," *World Meteorological Organization Journal*, vol. 58, no. 3, p. 175, 2009.

[17] X. Gao, J. S. Pal, and F. Giorgi, "Projected changes in mean and extreme precipitation over the mediterranean region from a high resolution double nested rcm simulation," *Geophysical Research Letters*, vol. 33, no. 3, 2006.

[18] D. Maraun, "Bias correcting climate change simulations-a critical review," *Current Climate Change Reports*, vol. 2, no. 4, pp. 211–220, 2016.

[19] H. Fowler and M. Ekström, "Multi-model ensemble estimates of climate change impacts on uk seasonal precipitation extremes," *International Journal of Climatology: A Journal of the Royal Meteorological Society*, vol. 29, no. 3, pp. 385–416, 2009.

[20] R. Seager and G. A. Vecchi, "Greenhouse warming and the 21st century hydroclimate of southwestern North America," *Proceedings of the National Academy of Sciences*, vol. 107, no. 50, pp. 21 277–21 282, 2010.

[21] L. Mearns, F. Giorgi, P. Whetton, D. Pabon, M. Hulme, and M. Lal, "Guidelines for use of climate scenarios developed from regional climate model experiments," *Data Distribution Centre of the Intergovernmental Panel on Climate Change*, 2003.

[22] G. Fang, J. Yang, Y. Chen, and C. Zammit, "Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in china," *Hydrology and Earth System Sciences*, vol. 19, no. 6, pp. 2547–2559, 2015.

[23] R. Huth, "Statistical downscaling in central europe: Evaluation of methods and potential predictors," *Climate Res.*, vol. 13, no. 2, pp. 91–101, 1999.

[24] H. Kettle and R. Thompson, "Statistical downscaling in european mountains: Verification of reconstructed air temperature," *Climate Res.*, vol. 26, no. 2, pp. 97–112, 2004.

[25] J. T. Schoof and S. Pryor, "Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks," *International Journal of Climate*, vol. 21, no. 7, pp. 773–790, 2001.

[26] T. Vandal, E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, "Deepsd: Generating high resolution climate change projections through single image super-resolution," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2017, pp. 1663–1672.

[27] C. Hutengs and M. Vohland, "Downscaling land surface temperatures at regional scales with random forest regression," *Remote Sensing of Environment*, vol. 178, pp. 127–141, 2016.

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

[29] D. Jeong, A. St-Hilaire, T. Ouarda, and P. Gachon, "Comparison of transfer functions in statistical downscaling models for daily temperature and precipitation over canada," *Stochastic environmental research and risk assessment*, vol. 26, no. 5, pp. 633–653, 2012.

[30] R. L. Wilby and T. Wigley, "Downscaling general circulation model output: A review of methods and limitations," *Progress in physical geography*, vol. 21, no. 4, pp. 530–548, 1997.

[31] J. M. Winter, B. Beckage, G. Bucini, R. M. Horton, and P. J. Clemins, "Development and evaluation of high-resolution climate simulations over the mountainous northeastern united states," *Journal of Hydrometeorology*, vol. 17, no. 3, pp. 881–896, 2016.

[32] F. Wang, D. Tian, L. Lowe, L. Kalin, and J. Lehrter, "Deep learning for daily precipitation and temperature downscaling," *Water Resources Research*, vol. 57, no. 4, e2020WR029308, 2021.

[33] B. Pang, J. Yue, G. Zhao, and Z. Xu, "Statistical downscaling of temperature with the random forest model," *Advances in Meteorology*, vol. 2017, 2017.

[34] T. Vandal, E. Kodra, and A. R. Ganguly, "Intercomparison of machine learning methods for statistical downscaling: The case of daily and extreme precipitation," *Theoretical and Applied Climatology*, vol. 137, no. 1, pp. 557–570, 2019.

[35] D. Sachindra, K. Ahmed, M. M. Rashid, S. Shahid, and B. Perera, "Statistical downscaling of precipitation using machine learning techniques," *Atmospheric research*, vol. 212, pp. 240–258, 2018.

[36] X. Li, Z. Li, W. Huang, and P. Zhou, "Performance of statistical and machine learning ensembles for daily temperature downscaling," *Theoretical and Applied Climatology*, vol. 140, no. 1, pp. 571–588, 2020.

[37] J. Baño-Medina, R. Manzanas, and J. M. Gutiérrez, "Configuration and intercomparison of deep learning neural models for statistical downscaling," *Geoscientific Model Development*, vol. 13, no. 4, pp. 2109–2124, 2020.

[38] T. Bong, Y.-H. Son, S.-H. Yoo, and S.-W. Hwang, "Nonparametric quantile mapping using the response surface method–bias correction of daily precipitation," *Journal of Water and Climate Change*, vol. 9, no. 3, pp. 525–539, 2018.

[39] S. Pour, S. Harun, and S. Shahid, "Genetic programming for the downscaling of extreme rainfall events on the east coast of peninsular malaysia," *Atmosphere*, vol. 5, no. 4, pp. 914–936, 2014.

[40] R. M. B. Harris, M. R. Grose, G. Lee, N. L. Bindoff, L. L. Porfirio, and P. Fox-Hughes, "Climate projections for ecologists," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 5, no. 5, pp. 621–637, 2014.

[41] I. Takayabu, H. Kanamaru, K. Dairaku, R. Benestad, H. von Storch, and J. H. Christensen, "Reconsidering the quality and utility of downscaling," *Journal of the Meteorological Society of Japan. Ser. II*, vol. 94, pp. 31–45, 2016.

[42] G. Lenderink, A. Buishand, and W. v. Deursen, "Estimates of future discharges of the river rhine using two scenario methodologies: Direct versus delta approach," *Hydrology and Earth System Sciences*, vol. 11, no. 3, pp. 1145–1159, 2007.

[43] L. E. Hay, R. L. Wilby, and G. H. Leavesley, "A comparison of delta change and downscaled gcm scenarios for three mountainous basins in the united states 1," *JAWRA Journal of the American Water Resources Association*, vol. 36, no. 2, pp. 387–397, 2000.

[44] R. Leander and T. A. Buishand, "Resampling of regional climate model output for the simulation of extreme river flows," *Journal of Hydrology*, vol. 332, no. 3-4, pp. 487–496, 2007.

[45] C. Piani, J. Haerter, and E. Coppola, "Statistical bias correction for daily precipitation in regional climate models over europe," *Theoretical and Applied Climatology*, vol. 99, no. 1-2, pp. 187–192, 2010.

[46] C. Teutschbein and J. Seibert, "Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods," *Journal of Hydrology*, vol. 456, pp. 12–29, 2012.

[47] A. J. Cannon, S. R. Sobie, and T. Q. Murdock, "Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?" *Journal of Climate*, vol. 28, no. 17, pp. 6938–6959, 2015.

[48] A. W. Wood, E. P. Maurer, A. Kumar, and D. P. Lettenmaier, "Long-range experimental hydrologic forecasting for the eastern united states," *Journal of Geophysical Research: Atmospheres*, vol. 107, no. D20, ACL–6, 2002.

[49] E. P. Maurer and P. B. Duffy, "Uncertainty in projections of streamflow changes due to climate change in california," *Geophysical Research Letters*, vol. 32, no. 3, 2005.

[50] K. Hayhoe, C. Wake, B. Anderson, *et al.*, "Regional climate change projections for the northeast usa," *Mitigation and Adaptation Strategies for Global Change*, vol. 13, no. 5-6, pp. 425–436, 2008.

[51] A. W. Wood, L. R. Leung, V. Sridhar, and D. Lettenmaier, "Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs," *Climatic Change*, vol. 62, no. 1-3, pp. 189–216, 2004.

[52] J. C. Bennett, M. R. Grose, S. P. Corney, *et al.*, "Performance of an empirical bias-correction of a high-resolution climate dataset," *International Journal of Climatology*, vol. 34, no. 7, pp. 2189–2204, 2014.

[53] M. Jakob Themeßl, A. Gobiet, and A. Leuprecht, "Empirical-statistical downscaling and error correction of daily precipitation from regional climate models," *International Journal of Climatology*, vol. 31, no. 10, pp. 1530–1544, 2011.

[54] M. Shrestha, S. C. Acharya, and P. K. Shrestha, "Bias correction of climate models for hydrological modelling–are simple methods still useful?" *Meteorological Applications*, vol. 24, no. 3, pp. 531–539, 2017.

[55] D. R. Roberts, W. H. Wood, and S. J. Marshall, "Assessments of downscaled climate data with a high-resolution weather station network reveal consistent but predictable bias," *International Journal of Climatology*, vol. 39, no. 6, pp. 3091–3103, 2019.

[56] M. Shabalova, W. Van Deursen, and T. Buishand, "Assessing future discharge of the river rhine using regional climate model integrations and a hydrological model," *Climate research*, vol. 23, no. 3, pp. 233–246, 2003.

[57] R. W. Katz, "Extreme value theory for precipitation: Sensitivity analysis for climate change," *Advances in Water Resources*, vol. 23, no. 2, pp. 133–139, 1999.

[58] D. S. Wilks, *Statistical methods in the atmospheric sciences*. Academic press, 2011, vol. 100.

[59] P. Smitha, B. Narasimhan, K. Sudheer, and H. Annamalai, "An improved bias correction method of daily rainfall data using a sliding window technique for climate change impact assessment," *Journal of Hydrology*, vol. 556, pp. 100–118, 2018.

[60] O. Räty, J. Räisänen, and J. S. Ylhäisi, "Evaluation of delta change and bias correction methods for future daily precipitation: Intermodel cross-validation using ensembles simulations," *Climate dynamics*, vol. 42, no. 9-10, pp. 2287–2303, 2014.

[61] L. Gudmundsson, J. Bremnes, J. Haugen, and T. Engen-Skaugen, "Downscaling rcm precipitation to the station scale using statistical transformations–a comparison of methods," *Hydrology and Earth System Sciences*, vol. 16, no. 9, pp. 3383–3390, 2012.

[62] X. Qin and C. Dai, "Comparison of different quantile delta mapping schemes in frequency analysis of precipitation extremes over mainland southeast asia under climate change," *Journal of Hydrology*, p. 127 421, 2022.

[63] J.-H. Heo, H. Ahn, J.-Y. Shin, T. R. Kjeldsen, and C. Jeong, "Probability distributions for a quantile mapping technique for a bias correction of precipitation data: A case study to precipitation data under climate change," *Water*, vol. 11, no. 7, p. 1475, 2019.

[64] L. Ye, L. S. Hanson, P. Ding, D. Wang, and R. M. Vogel, "The probability distribution of daily precipitation at the point and catchment scales in the united states," *Hydrology and Earth System Sciences*, vol. 22, no. 12, pp. 6519–6531, 2018.

[65] P. Reiter, O. Gutjahr, L. Schefczyk, G. Heinemann, and M. Casper, "Does applying quantile mapping to subsamples improve the bias correction of daily precipitation?" *International Journal of Climatology*, vol. 38, no. 4, pp. 1623–1633, 2018.

[66] O. Gutjahr and G. Heinemann, "Comparing precipitation bias correction methods for high-resolution regional climate simulations using cosmo-clm," *Theoretical and applied climatology*, vol. 114, no. 3, pp. 511–529, 2013.

[67] M. A. Ivanov and S. Kotlarski, "Assessing distribution-based climate model bias correction methods over an alpine domain: Added value and limitations," *International Journal of Climatology*, vol. 37, no. 5, pp. 2633–2653, 2017.

[68] C. Volosciuk, D. Maraun, M. Vrac, and M. Widmann, "A combined statistical bias correction and stochastic downscaling method for precipitation," *Hydrology and Earth System Sciences*, vol. 21, no. 3, pp. 1693–1719, 2017.

[69] J.-Y. Shin, T. Lee, T. Park, and S. Kim, "Bias correction of rcm outputs using mixture distributions under multiple extreme weather influences," *Theoretical and Applied Climatology*, vol. 137, no. 1, pp. 201–216, 2019.

[70] L. E. Hay and M. Clark, "Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western united states," *Journal of Hydrology*, vol. 282, no. 1-4, pp. 56–75, 2003.

[71] S. McGinnis, D. Nychka, and L. O. Mearns, "A new distribution mapping technique for climate model bias correction," in *Machine learning and data mining approaches to climate science*, Springer, 2015, pp. 91–99.

[72] V. M. Garibay, M. W. Gitau, N. Kiggundu, D. Moriasi, and F. Mishili, "Evaluation of reanalysis precipitation data and potential bias correction methods for use in data-scarce areas," *Water Resources Management*, pp. 1–16, 2021.

[73] C. Miao, L. Su, Q. Sun, and Q. Duan, "A nonstationary bias-correction technique to remove bias in gcm simulations," *Journal of Geophysical Research: Atmospheres*, vol. 121, no. 10, pp. 5718–5735, 2016.

[74] N. Oceanic and A. Administration, *Climate data online search*, Accessed: 2017-09-30, 2018. [Online]. Available: https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND.

[75] J. Boé, L. Terray, F. Habets, and E. Martin, "Statistical and dynamical downscaling of the seine basin climate for hydro-meteorological studies," *International Journal of Climatology: A Journal of the Royal Meteorological Society*, vol. 27, no. 12, pp. 1643–1655, 2007.

[76] B. Livneh, T. J. Bohn, D. W. Pierce, *et al.*, "A spatially comprehensive, hydrometeorological data set for mexico, the us, and southern canada 1950–2013," *Scientific data*, vol. 2, no. 1, pp. 1–12, 2015.

[77] P. E. Thornton, M. M. Thornton, B. W. Mayer, *et al.*, "Daymet: Daily surface weather on a 1 km grid for north america, 1980-2008," *Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center for Biogeochemical Dynamics (DAAC)*, 2012.

[78] C. Daly, G. Taylor, W. Gibson, T. Parzybok, G. Johnson, and P. Pasteris, "High-quality spatial climate data sets for the united states and beyond," *Transactions of the ASAE*, vol. 43, no. 6, p. 1957, 2000.

[79] R. Behnke, S. Vavrus, A. Allstadt, T. Albright, W. E. Thogmartin, and V. C. Radeloff, "Evaluation of downscaled, gridded climate data for the conterminous united states," *Ecological applications*, vol. 26, no. 5, pp. 1338–1351, 2016.

[80] D. Walton and A. Hall, "An assessment of high-resolution gridded temperature datasets over california," *Journal of Climate*, vol. 31, no. 10, pp. 3789–3810, 2018.

[81] A. M. Wootten, K. W. Dixon, D. J. Adams-Smith, and R. A. McPherson, "Statistically downscaled precipitation sensitivity to gridded observation data and downscaling technique," *International Journal of Climatology*, 2020.

[82] T. C. Peterson and R. S. Vose, "An overview of the global historical climatology network temperature database," *Bulletin of the American Meteorological Society*, vol. 78, no. 12, pp. 2837–2850, 1997.

[83] I. Durre, M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, "Comprehensive automated quality assurance of daily surface observations," *Journal of Applied Meteorology and Climatology*, vol. 49, no. 8, pp. 1615–1633, 2010.

[84] J. F. Mejia, J. Huntington, B. Hatchett, D. Koracin, and R. G. Niswonger, "Linking global climate models to an integrated hydrologic model: Using an individual station downscaling approach," *J. Contemp. Water Res. Educ.*, vol. 147, no. 1, pp. 17–27, 2012.

[85] L. Poggio and A. Gimona, "Downscaling and correction of regional climate models outputs with a hybrid geostatistical approach," *Spat. Stat.*, vol. 14, pp. 4–21, 2015.

[86] J. Rajczak, S. Kotlarski, N. Salzmann, and C. Schaer, "Robust climate scenarios for sites with sparse observations: A two-step bias correction approach," *International Journal of Climatology*, vol. 36, no. 3, pp. 1226–1243, 2016.

[87] F. Lehner, I. Nadeem, and H. Formayer, "Evaluating quantile-based bias adjustment methods for climate change scenarios," *Hydrology and Earth System Sciences Discussions*, pp. 1–26, 2021.

[88] A. K. Mishra, V. P. Singh, and S. K. Jain, "Impact of global warming and climate change on social development," *Journal of Comparative Social Welfare*, vol. 26, no. 2-3, pp. 239–260, 2010.

[89] M. Clark, S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, "The schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields," *Journal of Hydrometeorology*, vol. 5, no. 1, pp. 243–262, 2004.

[90] M. Vrac and P. Friederichs, "Multivariate—intervariable, spatial, and temporal—bias correction," *Journal of Climate*, vol. 28, no. 1, pp. 218–237, 2015.

[91] F. Johnson and A. Sharma, "A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations," *Water Resources Research*, vol. 48, no. 1, 2012.

[92] F. Johnson and A. Sharma, "Assessing future droughts in australia-a nesting model to correct for long-term persistence in general circulation model precipitation simulations," in *8th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation*, Modell., Simul. Soc. of Aust., and NZ, Cairns Queensland, Australia, 2009, pp. 3935–3941.

[93] F. Johnson, S. Westra, A. Sharma, and A. J. Pitman, "An assessment of gcm skill in simulating persistence across multiple time scales," *Journal of Climate*, vol. 24, no. 14, pp. 3609–3623, 2011.

[94] R. Mehrotra and A. Sharma, "An improved standardization procedure to remove systematic low frequency variability biases in gcm simulations," *Water Resources Research*, vol. 48, no. 12, 2012.

[95] H. Nguyen, R. Mehrotra, and A. Sharma, "Correcting for systematic biases in gcm simulations in the frequency domain," *Journal of Hydrology*, vol. 538, pp. 117–126, 2016.

[96] C. Kusumastuti, Z. Jiang, R. Mehrotra, and A. Sharma, "A signal processing approach to correct systematic bias in trend and variability in climate model simulations," *Geophysical Research Letters*, vol. 48, no. 13, e2021GL092953, 2021.

[97] J. Haerter, S. Hagemann, C. Moseley, and C. Piani, "Climate model bias correction and the role of timescales," *Hydrology and Earth System Sciences*, vol. 15, no. 3, pp. 1065–1079, 2011.

[98] H. Fowler, M. Ekström, S. Blenkinsop, and A. Smith, "Estimating change in extreme european precipitation using a multimodel ensemble," *Journal of Geophysical Research: Atmospheres*, vol. 112, no. D18, 2007.

[99] M. G. Grillakis, A. G. Koutroulis, and I. K. Tsanis, "Multisegment statistical bias correction of daily gcm precipitation output," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 8, pp. 3150–3162, 2013.

[100] M. F. Holthuijzen, B. Beckage, P. J. Clemins, D. Higdon, and J. M. Winter, "Constructing high-resolution, bias-corrected climate products: A comparison of methods," *Journal of Applied Meteorology and Climatology*, vol. 60, no. 4, pp. 455–475, 2021.

[101] A. Mamalakis, A. Langousis, R. Deidda, and M. Marrocu, "A parametric approach for simultaneous bias correction and high-resolution downscaling of climate model rainfall," *Water Resources Research*, vol. 53, no. 3, pp. 2149–2170, 2017.

[102] L. Gudmundsson, *Qmap: Statistical transformations for post-processing climate model output*, R package version 1.0-4, 2016.

[103] A. Gobiet, M. Suklitsch, and G. Heinrich, "The effect of empirical-statistical correction of intensity-dependent model errors on the temperature climate change signal," *Hydrology and Earth System Sciences*, vol. 19, no. 10, pp. 4055–4066, 2015.

[104] J. Hnilica, M. Hanel, and V. Puš, "Multisite bias correction of precipitation data from regional climate models," *International Journal of Climatology*, vol. 37, no. 6, pp. 2934–2946, 2017.

[105] M. G. Grillakis, A. G. Koutroulis, I. N. Daliakopoulos, and I. K. Tsanis, "A method to preserve trends in quantile mapping bias correction of climate modeled temperature," *Earth System Dynamics*, vol. 8, no. 3, p. 889, 2017.

[106] S. Tani and A. Gobiet, "Quantile mapping for improving precipitation extremes from regional climate models," *Journal of Agrometeorology*, vol. 21, no. 4, pp. 434–443, 2019.

[107] W. Yang, J. Andréasson, L. Phil Graham, J. Olsson, J. Rosberg, and F. Wetterhall, "Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies," *Hydrology Research*, vol. 41, no. 3-4, pp. 211–229, 2010.

[108] E. M. Laflamme, E. Linder, and Y. Pan, "Statistical downscaling of regional climate model output to achieve projections of precipitation extremes," *Weather and climate extremes*, vol. 12, pp. 15–23, 2016.

[109] D.-I. Kim, H.-H. Kwon, and D. Han, "Exploring the long-term reanalysis of precipitation and the contribution of bias correction to the reduction of uncertainty over south korea: A composite gamma-pareto distribution approach to the bias correction," *Hydrology and Earth System Sciences Discussions*, pp. 1–53, 2018.

[110] M. Luo, T. Liu, F. Meng, *et al.*, "Comparing bias correction methods used in downscaling precipitation and temperature from regional climate models: A case study from the kaidu river basin in western china," *Water*, vol. 10, no. 8, p. 1046, 2018.

[111] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels, *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.

[112] O. Schabenberger and C. A. Gotway, *Statistical methods for spatial data analysis*. Chapman and Hall/CRC, 2017.

[113] M. B. Switanek, P. A. Troch, C. L. Castro, *et al.*, "Scaled distribution mapping: A bias correction method that preserves raw climate model projected changes," *Hydrology and Earth System Sciences*, vol. 21, no. 6, pp. 2649–2666, 2017.

[114] S. Hagemann, C. Chen, J. O. Haerter, J. Heinke, D. Gerten, and C. Piani, "Impact of a statistical bias correction on the projected hydrological changes obtained from three gcms and two hydrology models," *Journal of Hydrometeorology*, vol. 12, no. 4, pp. 556–578, 2011.

[115] M. A. Ivanov, J. Luterbacher, and S. Kotlarski, "Climate model biases and modification of the climate change signal by intensity-dependent bias correction," *Journal of Climate*, vol. 31, no. 16, pp. 6591–6610, 2018.

[116] H. Li, J. Sheffield, and E. F. Wood, "Bias correction of monthly precipitation and temperature fields from intergovernmental panel on climate change ar4 models using equidistant quantile matching," *Journal of Geophysical Research: Atmospheres*, vol. 115, no. D10, 2010.

[117] R. K. Srivastav, A. Schardong, and S. P. Simonovic, "Equidistance quantile matching method for updating idfcurves under climate change," *Water resources management*, vol. 28, no. 9, pp. 2539–2562, 2014.

[118] L. E. Flint and A. L. Flint, "Downscaling future climate scenarios to fine scales for hydrologic and ecological modeling and analysis," *Ecological Processes*, vol. 1, no. 1, p. 2, 2012.

[119] J. Franklin, F. W. Davis, M. Ikegami, *et al.*, "Modeling plant species distributions under future climates: How fine scale do climate projections need to be?" *Global change biology*, vol. 19, no. 2, pp. 473–483, 2013.

[120] J. W. Hansen, "Integrating seasonal climate prediction and agricultural models for insights into agricultural practice," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1463, pp. 2037–2047, 2005.

[121] J. D. Fridley, "Downscaling climate over complex terrain: High finescale (less than 1000 m) spatial variation of near-ground temperatures in a montane forested landscape (great smoky mountains)," *Journal of Applied Meteorology and Climatology*, vol. 48, no. 5, pp. 1033–1049, 2009.

[122] D. Maraun, F. Wetterhall, A. Ireson, *et al.*, "Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user," *Reviews of Geophysics*, vol. 48, no. 3, 2010.

[123] R. Haas and J. G. Pinto, "A combined statistical and dynamical approach for downscaling large-scale footprints of european windstorms," *Geophysical Research Letters*, vol. 39, no. 23, 2012.

[124] T. Engen-Skaugen, "Refinement of dynamically downscaled precipitation and temperature scenarios," *Climate Change*, vol. 84, no. 3-4, pp. 365–382, 2007.

[125] Z. Han, Y. Shi, J. Wu, Y. Xu, and B. Zhou, "Combined dynamical and statistical downscaling for high-resolution projections of multiple climate variables in the beijing–tianjin–hebei region of china," *Journal of Applied Meteorology and Climatology*, vol. 58, no. 11, pp. 2387–2403, 2019.

[126] D. A. Bishop and C. M. Beier, "Assessing uncertainty in high-resolution spatial climate data across the us northeast," *PloS one*, vol. 8, no. 8, e70260, 2013.

[127] C. Xu, J. Wang, and Q. Li, "A new method for temperature spatial interpolation based on sparse historical stations," *Journal of Climate*, vol. 31, no. 5, pp. 1757–1770, 2018.

[128] W. C. Skamarock, J. B. Klemp, J. Dudhia, *et al.*, "A description of the advanced research wrf model," 2019, Accessed: 2019-03-04. [Online]. Available: https://opensky.ucar.edu/islandora/object/opensky:2898.

[129] D. P. Dee, S. M. Uppala, A. Simmons, *et al.*, "The era-interim reanalysis: Configuration and performance of the data assimilation system," *Quarterly Journal of the royal meteorological society*, vol. 137, no. 656, pp. 553–597, 2011.

[130] H. Huang, J. M. Winter, E. C. Osterberg, *et al.*, "Simulating precipitation and temperature in the lake champlain basin using a regional climate model: Limitations and uncertainties," *Climate Dynamics*, vol. 54, no. 1-2, pp. 69–84, 2020.

[131] USGS, *The national map*, 2018. [Online]. Available: https://viewer.nationalmap.gov/basic/.

[132] T. Wang, A. Hamann, D. L. Spittlehouse, and T. Q. Murdock, "ClimateWNA—high-resolution spatial climate data for western North America," *Journal of Applied Meteorology and Climatology*, vol. 51, no. 1, pp. 16–29, 2012.

[133] R. G. Barry, *Mountain weather and climate*. Psychology Press, 1992.

[134] E. P. Boer, K. M. de Beurs, and A. D. Hartkamp, "Kriging and thin plate splines for mapping climate variables," *International Journal of Applied Earth Observation and Geoinformation*, vol. 3, no. 2, pp. 146–154, 2001.

[135] S. J. Jeffrey, J. O. Carter, K. B. Moodie, and A. R. Beswick, "Using spatial interpolation to construct a comprehensive archive of australian climate data," *Environmental Modeling and Software*, vol. 16, no. 4, pp. 309–330, 2001.

[136] C. K. Wikle, A. Zammit-Mangion, and N. Cressie, *Spatio-temporal Statistics with R*. CRC Press, 2019.

[137] C. Daly, "Guidelines for assessing the suitability of spatial climate data sets," *International Journal of Climatology*, vol. 26, no. 6, pp. 707–721, 2006.

[138] B. Gräler, E. Pebesma, and G. Heuvelink, "Spatio-temporal interpolation using gstat," *The R Journal*, vol. 8, pp. 204–218, 1 2016. [Online]. Available: https://journal.r-project.org/archive/2016/RJ-2016-014/index.html.

[139] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.R-project.org/.

[140] E. Varouchakis and D. Hristopulos, "Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins," *Environmental monitoring and assessment*, vol. 185, no. 1, pp. 1–19, 2013.

[141] N. Hofstra, M. Haylock, M. New, P. Jones, and C. Frei, "Comparison of six methods for the interpolation of daily, european climate data," *Journal of Geophysical Research: Atmospheres*, vol. 113, no. D21, 2008.

[142] G. A. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, 2012, vol. 329.

[143] J. Pilz and G. Spöck, "Why do we need and how should we implement bayesian kriging methods," *Stochastic Environmental Research and Risk Assessment*, vol. 22, no. 5, pp. 621–632, 2008.

[144] P. Berg, H. Feldmann, and H.-J. Panitz, "Bias correction of high resolution regional climate model data," *Journal of Hydrology*, vol. 448, pp. 80–92, 2012.

[145] S. Perkins, A. Pitman, N. Holbrook, and J. McAneney, "Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions," *Journal of climate*, vol. 20, no. 17, pp. 4356–4376, 2007.

[146] D. Keellings, "Evaluation of downscaled cmip5 model skill in simulating daily maximum temperature over the southeastern united states," *International Journal of Climatology*, vol. 36, no. 12, pp. 4172–4180, 2016.

[147] R. Bordoy and P. Burlando, "Bias correction of regional climate model simulations in a region of complex orography," *Journal of applied meteorology and climatology*, vol. 52, no. 1, pp. 82–101, 2013.

[148] R. Lenth, *Emmeans: Estimated marginal means, aka least-squares means*, R package version 1.4.4, 2020. [Online]. Available: https://CRAN.R-project.org/package=emmeans.

[149] T. R. Levine and C. R. Hullett, "Eta squared, partial eta squared, and misreporting of effect size in communication research," *Human Communication Research*, vol. 28, no. 4, pp. 612–625, 2002.

[150] K. E. Muller and B. L. Peterson, "Practical methods for computing power in testing the multivariate general linear hypothesis," *Computational Statistics and Data Analysis*, vol. 2, no. 2, pp. 143–158, 1984.

[151] X. Meng, S. Lyu, T. Zhang, *et al.*, "Simulated cold bias being improved by using modis time-varying albedo in the tibetan plateau in wrf model," *Environmental Research Letters*, vol. 13, no. 4, p. 044 028, 2018.

[152] C. Goodess, C. Anagnostopoulou, A. Bárdossy, *et al.*, "An intercomparison of statistical downscaling methods for europe and european regions–assessing their performance with respect to extreme temperature and precipitation events," *Climatic Change*, pp. 1–68, 2007.

[153] J. Friedman, T. Hastie, R. Tibshirani, *et al.*, *The elements of statistical learning*, 10. Springer series in statistics New York, 2001, vol. 1.

[154] H. J. Fowler, S. Blenkinsop, and C. Tebaldi, "Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling," *International journal of climatology*, vol. 27, no. 12, pp. 1547–1578, 2007.

[155] K. Stahl, R. Moore, J. Floyer, M. Asplin, and I. McKendry, "Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density," *Agricultural and Forest Meteorology*, vol. 139, no. 3-4, pp. 224–236, 2006.

[156] C. B. Field, V. R. Barros, D. J. Dokken, *et al.*, "Contribution of working group ii to the fifth assessment report of the intergovernmental panel on climate change," *Climate change*, 2014.

[157] J. R. Lanzante, K. W. Dixon, D. Adams-Smith, M. J. Nath, and C. E. Whitlock, "Evaluation of some distributional downscaling methods as applied to daily precipitation with an eye towards extremes," *International Journal of Climatology*, vol. 41, no. 5, pp. 3186–3202, 2021.

[158] K. Hayhoe, C. P. Wake, T. G. Huntington, *et al.*, "Past and future changes in climate and hydrological indicators in the us northeast," *Climate Dynamics*, vol. 28, no. 4, pp. 381–407, 2007.

[159] H. Huang, J. M. Winter, E. C. Osterberg, R. M. Horton, and B. Beckage, "Total and extreme precipitation changes over the northeastern united states," *Journal of Hydrometeorology*, vol. 18, no. 6, pp. 1783–1798, 2017.

[160] D. W. Pierce, D. R. Cayan, E. P. Maurer, J. T. Abatzoglou, and K. C. Hegewisch, "Improved bias correction techniques for hydrological simulations of climate change," *Journal of Hydrometeorology*, vol. 16, no. 6, pp. 2421–2442, 2015.

[161] H. Hoffmann and T. Rath, "Meteorologically consistent bias correction of climate time series for agricultural models," *Theoretical and Applied Climatology*, vol. 110, no. 1, pp. 129–141, 2012.

[162] G. A. Baigorria, J. W. Jones, D.-W. Shin, A. Mishra, and J. J. O'Brien, "Assessing uncertainties in crop model simulations using daily bias-corrected regional circulation model outputs," *Climate Research*, vol. 34, no. 3, pp. 211–222, 2007.

[163] P. Friederichs and A. Hense, "Statistical downscaling of extreme precipitation events using censored quantile regression," *Monthly Weather Review*, vol. 135, no. 6, pp. 2365–2378, 2007.

[164] M.-J. Um, H. Kim, and J.-H. Heo, "Hybrid approach in statistical bias correction of projected precipitation for the frequency analysis of extreme events," *Advances in Water Resources*, vol. 94, pp. 278–290, 2016.

[165] M. Enayati, O. Bozorg-Haddad, J. Bazrafshan, S. Hejabi, and X. Chu, "Bias correction capabilities of quantile mapping methods for rainfall and temperature variables," *Journal of Water and Climate Change*, vol. 12, no. 2, pp. 401–419, 2021.

[166] C. Stager and M. Thill, "Climate change in the champlain basin: What natural resource managers can expect and do, the nature conservancy adirondack ch. and vt ch," *Rep., Keene Valley, NY*, 2010.

[167] A. Monaghan, D. Steinhoff, C. Bruyere, and D. Yates, "Ncar cesm global bias-corrected cmip5 output to support wrf/mpas research," *Research Data Archive National Center Atmospheric Research Computational Information Systems Laboratory, Boulder. DOI*, vol. 10, D6DJ5CN4, 2014.

[168] D. Bannister, A. Orr, S. K. Jain, *et al.*, "Bias correction of high-resolution regional climate model precipitation output gives the best estimates of precipitation in himalayan catchments," *Journal of Geophysical Research: Atmospheres*, vol. 124, no. 24, pp. 14 220–14 239, 2019.

[169] A. M. Wootten, K. W. Dixon, D. J. Adams-Smith, and R. A. McPherson, "Statistically downscaled precipitation sensitivity to gridded observation data and downscaling technique," *International Journal of Climatology*, vol. 41, no. 2, pp. 980–1001, 2021.

[170] P. Reiter, O. Gutjahr, L. Schefczyk, G. Heinemann, and M. Casper, "Bias correction of ensembles precipitation data with focus on the effect of the length of the calibration period," *Meteorologische Zeitschrift*, pp. 85–96, 2016.

[171] T. Peterson, "Climate change indices," *WMO bulletin*, vol. 54, no. 2, pp. 83–86, 2005.

[172] T. R. Karl, N. Nicholls, and A. Ghazi, "Clivar/gcos/wmo workshop on indices and indicators for climate extremes workshop summary," in *Weather and climate extremes*, Springer, 1999, pp. 3–7.

[173] L. Alexander, M. Donat, Y. Takayama, and H. Yang, "The climdex project: Creation of long-term global gridded products for the analysis of temperature and precipitation extremes," in *WCRP Open Science conference, Denver*, 2011.

[174] M. Plummer, A. Stukalov, M. Denwood, and M. M. Plummer, "Package 'rjags'," *Vienna, Austria*, 2016.

[175] N. Smirnov *et al.*, "Table for estimating the goodness of fit of empirical distributions," *Annals of mathematical statistics*, vol. 19, no. 2, pp. 279–281, 1948.

[176] E. A. Rosenberg, P. W. Keys, D. B. Booth, *et al.*, "Precipitation extremes and the impacts of climate change on stormwater infrastructure in washington state," *Climatic change*, vol. 102, no. 1, pp. 319–349, 2010.

[177] G. V. Tschöke, N. S. Kruk, P. I. B. de Queiroz, S. C. Chou, and W. C. de Sousa Junior, "Comparison of two bias correction methods for precipitation simulated with a regional climate model," *Theoretical and applied climatology*, vol. 127, no. 3, pp. 841–852, 2017.

[178] N. M. Razali, Y. B. Wah, *et al.*, "Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests," *Journal of statistical modeling and analytics*, vol. 2, no. 1, pp. 21–33, 2011.

[179] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian journal of statistics*, pp. 65–70, 1979.

[180] J. Martins, H. Fraga, A. Fonseca, and J. A. Santos, "Climate projections for precipitation and temperature indicators in the douro wine region: The importance of bias correction," *Agronomy*, vol. 11, no. 5, p. 990, 2021.

[181] V. J. Berrocal, A. E. Gelfand, and D. M. Holland, "A spatio-temporal downscaler for output from numerical models," *Journal of agricultural, biological, and environmental statistics*, vol. 15, no. 2, pp. 176–197, 2010.

[182] M. Kiel, C. W. O'Dell, B. Fisher, *et al.*, "How bias correction goes wrong: Measurement of x co 2 affected by erroneous surface pressure estimates," *Atmospheric Measurement Techniques*, vol. 12, no. 4, pp. 2241–2259, 2019.

[183] F.-Y. Cheng, C.-Y. Feng, Z.-M. Yang, *et al.*, "Evaluation of real-time pm2. 5 forecasts with the wrf-cmaq modeling system and weather-pattern-dependent bias-adjusted pm2. 5 forecasts in taiwan," *Atmospheric Environment*, vol. 244, p. 117 909, 2021.

[184] J. Zhang, P. Crippa, M. G. Genton, and S. Castruccio, "Assessing the reliability of wind power operations under a changing climate with a non-gaussian bias correction," *The Annals of Applied Statistics*, vol. 15, no. 4, pp. 1831–1849, 2021.

[185] K. Steele and C. Werndl, "Climate models, calibration, and confirmation," *The British Journal for the Philosophy of Science*, vol. 64, no. 3, pp. 609–635, 2013.

[186] D. Maraun and M. Widmann, *Statistical downscaling and bias correction for climate research*. Cambridge University Press, 2018.

[187] K. Tabor and J. W. Williams, "Globally downscaled climate projections for assessing the conservation impacts of climate change," *Ecological Applications*, vol. 20, no. 2, pp. 554–565, 2010.

[188] R. J. MacDonald, S. Boon, J. M. Byrne, M. D. Robinson, and J. B. Rasmussen, "Potential future climate effects on mountain hydrology, stream temperature, and native salmonid life history," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 71, no. 2, pp. 189–202, 2014.

[189] U. Ehret, E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert, "Hess opinions" should we apply bias correction to global and regional climate model data?"," *Hydrology and Earth System Sciences*, vol. 16, no. 9, pp. 3391–3404, 2012.

[190] D. Maraun, "Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue," *Journal of Climate*, vol. 26, no. 6, pp. 2137–2143, 2013.

[191] D. Maraun, R. Huth, J. M. Gutierrez, *et al.*, "The value perfect predictor experiment: Evaluation of temporal variability," *International Journal of Climatology*, vol. 39, no. 9, pp. 3786–3818, 2019.

[192] D. Higdon, "A process-convolution approach to modelling temperatures in the north atlantic ocean," *Environmental and Ecological Statistics*, vol. 5, no. 2, pp. 173–190, 1998.

[193] D. Higdon, J. Swall, and J. Kern, "Non-stationary spatial modeling," *Bayesian statistics*, vol. 6, no. 1, pp. 761–768, 1999.

[194] C. A. Calder, "Dynamic factor process convolution models for multivariate space–time data with application to air quality assessment," *Environmental and Ecological Statistics*, vol. 14, no. 3, pp. 229–247, 2007.

[195] S. Herrera, S. Kotlarski, P. M. Soares, *et al.*, "Uncertainty in gridded precipitation products: Influence of station density, interpolation method and grid resolution," *International Journal of Climatology*, vol. 39, no. 9, pp. 3717–3729, 2019.

[196] D. P. Van Vuuren, J. Edmonds, M. Kainuma, *et al.*, "The representative concentration pathways: An overview," *Climatic change*, vol. 109, no. 1, pp. 5–31, 2011.

[197] S. N. Wood, *Generalized additive models: an introduction with R*. CRC press, 2017.

[198] R. B. Gramacy, *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*. CRC Press, 2020.

[199] M. Binois and R. B. Gramacy, "Hetgp: Heteroskedastic gaussian process modeling and sequential design in r," *Journal of Statistical Software*, vol. 98, pp. 1–44, 2021.

[200] B. Ankenman, B. L. Nelson, and J. Staum, "Stochastic kriging for simulation metamodeling," in *2008 Winter Simulation Conference*, IEEE, 2008, pp. 362–370.

[201] A. Graps, "An introduction to wavelets," *IEEE computational science and engineering*, vol. 2, no. 2, pp. 50–61, 1995.

[202] D. Higdon, "Space and space-time modeling using process convolutions," in *Quantitative methods for current environmental issues*, Springer, 2002, pp. 37–56.

[203] C. A. Calder and N. Cressie, "Some topics in convolution-based spatial modeling," *Proceedings of the 56th Session of the International Statistics Institute*, pp. 22–29, 2007.

[204] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*. CRC press, 2013.

[205] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Core Team, *nlme: Linear and nonlinear mixed effects models*, R package version 3.1-144, 2020. [Online]. Available: https://CRAN.R-project.org/package=nlme.

[206] S. M. Butler and T. A. Louis, "Random effects models with non-parametric priors," *Statistics in medicine*, vol. 11, no. 14-15, pp. 1981–2000, 1992.

[207] X. Xu, M. Ghosh, *et al.*, "Bayesian variable selection and estimation for group lasso," *Bayesian Analysis*, vol. 10, no. 4, pp. 909–936, 2015.

[208] M. Holthuijzen, B. Beckage, D. Higdon, J. P. Clemins, and J. Winter, "A comparison of bias correction procedures for maximum daily temperature using sparsely distributed station data: Applications to high resolution downscaling," *Journal of Theoretical and Applied Climatology*, 2022.

[209] S. Banerjee, B. P. Carlin, and A. E. Gelfand, *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, 2004.

[210] A. Finley, *Spnngp*, Online, Jul. 2017. [Online]. Available: https://cran.r-project.org/web/packages/spNNGP/spNNGP.pdf.

[211] J. W. Tukey, "Comparing individual means in the analysis of variance," *Biometrics*, pp. 99–114, 1949.

[212] D.-I. Kim, H.-H. Kwon, and D. Han, "Bias correction of daily precipitation over south korea from the long-term reanalysis using a composite gamma-pareto distribution approach," *Hydrology Research*, vol. 50, no. 4, pp. 1138–1161, 2019.

[213] Y. Li, W. Cai, and E. Campbell, "Statistical modeling of extreme rainfall in southwest western australia," *Journal of climate*, vol. 18, no. 6, pp. 852–863, 2005.

[214] E. Gilleland and R. W. Katz, "Extremes 2.0: An extreme value analysis package in r," *Journal of Statistical Software*, vol. 72, no. 1, pp. 1–39, 2016.

[215] D. Nychka, C. Wikle, and J. A. Royle, "Multiresolution models for nonstationary spatial covariance functions," *Statistical Modelling*, vol. 2, no. 4, pp. 315–331, 2002.

[216] R. Varadhan, H. W. Borchers, and M. R. Varadhan, *Package 'dfoptim'*, 2016.

[217] O. Stegle, C. Lippert, J. M. Mooij, N. Larence, and K. Borgwardt, "Efficient inference in matrix-variate gaussian models with iid observation noise," 2011.

[218] H. Zhang, "Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 250–261, 2004.

[219] N. J. Higham, *Analysis of the Cholesky decomposition of a semi-definite matrix*. Oxford University Press, 1990.

A1. Chapter 2 supplementary material

Supplementary tables

1. Table S1. Pairwise comparisons for RMSE ANOVA model: *Month* × *Method*
2. Table S2. Pairwise comparisons for PSS ANOVAmodel: *Month* × *Method*
3. Table S3. Pairwise comparisons for RMSE ANOVA model: *Bias_correction_-years* × *Method*
4. Table S4. Pairwise comparisons for PSS ANOVA model: *Bias_correction_years* × *Method*
5. Table S5. Pairwise comparisons for RMSE ANOVA model: *Elevation* × *Method*
6. Table S6. Physics settings and details for the WRF model

Table 5.1: Table S1. Pairwise comparisons for the interaction Month × Method *in the RMSE ANOVA model*

contrast	estimate	SE	df	t.ratio	p.value
Month = 1					
EQM_krig - EQM_IDW	-0.0358	0.0508	204	-0.705	1.0000
EQM_grid - EQM_IDW	-0.0386	0.0508	204	-0.760	1.0000
EQM_grid - EQM_krig	-0.0028	0.0508	204	-0.055	1.0000
LT_grid - EQM_IDW	-0.1421	0.0508	204	-2.795	0.0854
LT_grid - EQM_krig	-0.1062	0.0508	204	-2.090	0.5682
LT_grid - EQM_grid	-0.1034	0.0508	204	-2.035	0.6472
LTQM_grid_C - EQM_IDW	-0.0414	0.0508	204	-0.815	1.0000
LTQM_grid_C - EQM_krig	-0.0056	0.0508	204	-0.110	1.0000
LTQM_grid_C - EQM_grid	-0.0028	0.0508	204	-0.055	1.0000
LTQM_grid_C - LT_grid	0.1006	0.0508	204	1.980	0.7361
LTQM_grid_V - EQM_IDW	-0.0307	0.0508	204	-0.604	1.0000
LTQM_grid_V - EQM_krig	0.0051	0.0508	204	0.100	1.0000
LTQM_grid_V - EQM_grid	0.0079	0.0508	204	0.155	1.0000
LTQM_grid_V - LT_grid	0.1113	0.0508	204	2.190	0.4447
LTQM_grid_V - LTQM_grid_C	0.0107	0.0508	204	0.210	1.0000
Month = 2					
EQM_krig - EQM_IDW	-0.0007	0.0508	204	-0.014	1.0000
EQM_grid - EQM_IDW	-0.0231	0.0508	204	-0.454	1.0000
EQM_grid - EQM_krig	-0.0223	0.0508	204	-0.440	1.0000
LT_grid - EQM_IDW	-0.1898	0.0508	204	-3.733	0.0037
LT_grid - EQM_krig	-0.1891	0.0508	204	-3.719	0.0039
LT_grid - EQM_grid	-0.1667	0.0508	204	-3.279	0.0184
LTQM_grid_C - EQM_IDW	-0.0468	0.0508	204	-0.920	1.0000
LTQM_grid_C - EQM_krig	-0.0460	0.0508	204	-0.906	1.0000
LTQM_grid_C - EQM_grid	-0.0237	0.0508	204	-0.466	1.0000
LTQM_grid_C - LT_grid	0.1430	0.0508	204	2.813	0.0808
LTQM_grid_V - EQM_IDW	-0.0487	0.0508	204	-0.957	1.0000
LTQM_grid_V - EQM_krig	-0.0479	0.0508	204	-0.943	1.0000
LTQM_grid_V - EQM_grid	-0.0256	0.0508	204	-0.503	1.0000
LTQM_grid_V - LT_grid	0.1411	0.0508	204	2.776	0.0903
LTQM_grid_V - LTQM_grid_C	-0.0019	0.0508	204	-0.037	1.0000

Table S1 cont.

Month = 3					
EQM_krig - EQM_IDW	0.0087	0.0508	204	0.171	1.0000
EQM_grid - EQM_IDW	-0.0360	0.0508	204	-0.707	1.0000
EQM_grid - EQM_krig	-0.0446	0.0508	204	-0.878	1.0000
LT_grid - EQM_IDW	-0.2269	0.0508	204	-4.463	0.0002
LT_grid - EQM_krig	-0.2356	0.0508	204	-4.634	0.0001
LT_grid - EQM_grid	-0.1909	0.0508	204	-3.756	0.0034
LTQM_grid_C - EQM_IDW	-0.0521	0.0508	204	-1.025	1.0000
LTQM_grid_C - EQM_krig	-0.0608	0.0508	204	-1.196	1.0000
LTQM_grid_C - EQM_grid	-0.0162	0.0508	204	-0.318	1.0000
LTQM_grid_C - LT_grid	0.1748	0.0508	204	3.438	0.0107
LTQM_grid_V - EQM_IDW	-0.0580	0.0508	204	-1.141	1.0000
LTQM_grid_V - EQM_krig	-0.0667	0.0508	204	-1.312	1.0000
LTQM_grid_V - EQM_grid	-0.0221	0.0508	204	-0.434	1.0000
LTQM_grid_V - LT_grid	0.1689	0.0508	204	3.322	0.0159
LTQM_grid_V - LTQM_grid_C	-0.0059	0.0508	204	-0.116	1.0000
Month = 4					
EQM_krig - EQM_IDW	0.0126	0.0508	204	0.248	1.0000
EQM_grid - EQM_IDW	-0.0205	0.0508	204	-0.404	1.0000
EQM_grid - EQM_krig	-0.0331	0.0508	204	-0.652	1.0000
LT_grid - EQM_IDW	-0.1101	0.0508	204	-2.165	0.4728
LT_grid - EQM_krig	-0.1227	0.0508	204	-2.413	0.2505
LT_grid - EQM_grid	-0.0895	0.0508	204	-1.762	1.0000
LTQM_grid_C - EQM_IDW	0.1601	0.0508	204	3.150	0.0282
LTQM_grid_C - EQM_krig	0.1475	0.0508	204	2.902	0.0618
LTQM_grid_C - EQM_grid	0.1806	0.0508	204	3.553	0.0071
LTQM_grid_C - LT_grid	0.2702	0.0508	204	5.315	<.0001
LTQM_grid_V - EQM_IDW	0.1469	0.0508	204	2.890	0.0640
LTQM_grid_V - EQM_krig	0.1343	0.0508	204	2.642	0.1331
LTQM_grid_V - EQM_grid	0.1674	0.0508	204	3.294	0.0175
LTQM_grid_V - LT_grid	0.2570	0.0508	204	5.055	<.0001
LTQM_grid_V - LTQM_grid_C	-0.0132	0.0508	204	-0.260	1.0000

Table S1 cont.

Month = 5					
EQM_krig - EQM_IDW	0.0234	0.0508	204	0.460	1.0000
EQM_grid - EQM_IDW	-0.0093	0.0508	204	-0.183	1.0000
EQM_grid - EQM_krig	-0.0327	0.0508	204	-0.644	1.0000
LT_grid - EQM_IDW	-0.0178	0.0508	204	-0.351	1.0000
LT_grid - EQM_krig	-0.0412	0.0508	204	-0.811	1.0000
LT_grid - EQM_grid	-0.0085	0.0508	204	-0.168	1.0000
LTQM_grid_C - EQM_IDW	0.1026	0.0508	204	2.018	0.6740
LTQM_grid_C - EQM_krig	0.0792	0.0508	204	1.557	1.0000
LTQM_grid_C - EQM_grid	0.1119	0.0508	204	2.201	0.4330
LTQM_grid_C - LT_grid	0.1204	0.0508	204	2.369	0.2818
LTQM_grid_V - EQM_IDW	0.2000	0.0508	204	3.934	0.0017
LTQM_grid_V - EQM_krig	0.1766	0.0508	204	3.474	0.0094
LTQM_grid_V - EQM_grid	0.2093	0.0508	204	4.118	0.0008
LTQM_grid_V - LT_grid	0.2178	0.0508	204	4.285	0.0004
LTQM_grid_V - LTQM_grid_C	0.0974	0.0508	204	1.917	0.8499

Table S1 cont.

Month = 6					
EQM_krig - EQM_IDW	-0.0152	0.0508	204	-0.300	1.0000
EQM_grid - EQM_IDW	-0.0260	0.0508	204	-0.512	1.0000
EQM_grid - EQM_krig	-0.0108	0.0508	204	-0.212	1.0000
LT_grid - EQM_IDW	-0.0445	0.0508	204	-0.875	1.0000
LT_grid - EQM_krig	-0.0293	0.0508	204	-0.576	1.0000
LT_grid - EQM_grid	-0.0185	0.0508	204	-0.364	1.0000
LTQM_grid_C - EQM_IDW	0.0233	0.0508	204	0.458	1.0000
LTQM_grid_C - EQM_krig	0.0385	0.0508	204	0.758	1.0000
LTQM_grid_C - EQM_grid	0.0493	0.0508	204	0.970	1.0000
LTQM_grid_C - LT_grid	0.0678	0.0508	204	1.334	1.0000
LTQM_grid_V - EQM_IDW	0.1182	0.0508	204	2.325	0.3160
LTQM_grid_V - EQM_krig	0.1334	0.0508	204	2.624	0.1401
LTQM_grid_V - EQM_grid	0.1442	0.0508	204	2.837	0.0753
LTQM_grid_V - LT_grid	0.1627	0.0508	204	3.200	0.0239
LTQM_grid_V - LTQM_grid_C	0.0949	0.0508	204	1.866	0.9512
Month = 7					
EQM_krig - EQM_IDW	0.0057	0.0508	204	0.113	1.0000
EQM_grid - EQM_IDW	-0.0355	0.0508	204	-0.698	1.0000
EQM_grid - EQM_krig	-0.0412	0.0508	204	-0.810	1.0000
LT_grid - EQM_IDW	-0.0054	0.0508	204	-0.106	1.0000
LT_grid - EQM_krig	-0.0111	0.0508	204	-0.218	1.0000
LT_grid - EQM_grid	0.0301	0.0508	204	0.592	1.0000
LTQM_grid_C - EQM_IDW	-0.0442	0.0508	204	-0.869	1.0000
LTQM_grid_C - EQM_krig	-0.0499	0.0508	204	-0.982	1.0000
LTQM_grid_C - EQM_grid	-0.0087	0.0508	204	-0.171	1.0000
LTQM_grid_C - LT_grid	-0.0388	0.0508	204	-0.763	1.0000
LTQM_grid_V - EQM_IDW	0.1112	0.0508	204	2.188	0.4468
LTQM_grid_V - EQM_krig	0.1055	0.0508	204	2.076	0.5877
LTQM_grid_V - EQM_grid	0.1467	0.0508	204	2.886	0.0648
LTQM_grid_V - LT_grid	0.1166	0.0508	204	2.294	0.3423
LTQM_grid_V - LTQM_grid_C	0.1554	0.0508	204	3.057	0.0380

Table S1 cont.

Month = 8					
EQM_krig - EQM_IDW	-0.0058	0.0508	204	-0.113	1.0000
EQM_grid - EQM_IDW	-0.0466	0.0508	204	-0.917	1.0000
EQM_grid - EQM_krig	-0.0409	0.0508	204	-0.804	1.0000
LT_grid - EQM_IDW	-0.0105	0.0508	204	-0.207	1.0000
LT_grid - EQM_krig	-0.0047	0.0508	204	-0.093	1.0000
LT_grid - EQM_grid	0.0361	0.0508	204	0.710	1.0000
LTQM_grid_C - EQM_IDW	0.0321	0.0508	204	0.631	1.0000
LTQM_grid_C - EQM_krig	0.0378	0.0508	204	0.744	1.0000
LTQM_grid_C - EQM_grid	0.0787	0.0508	204	1.547	1.0000
LTQM_grid_C - LT_grid	0.0426	0.0508	204	0.837	1.0000
LTQM_grid_V - EQM_IDW	0.3961	0.0508	204	7.792	<.0001
LTQM_grid_V - EQM_krig	0.4019	0.0508	204	7.905	<.0001
LTQM_grid_V - EQM_grid	0.4427	0.0508	204	8.709	<.0001
LTQM_grid_V - LT_grid	0.4066	0.0508	204	7.998	<.0001
LTQM_grid_V - LTQM_grid_C	0.3640	0.0508	204	7.161	<.0001
Month = 9					
EQM_krig - EQM_IDW	0.0063	0.0508	204	0.124	1.0000
EQM_grid - EQM_IDW	-0.0206	0.0508	204	-0.405	1.0000
EQM_grid - EQM_krig	-0.0269	0.0508	204	-0.529	1.0000
LT_grid - EQM_IDW	0.0030	0.0508	204	0.060	1.0000
LT_grid - EQM_krig	-0.0032	0.0508	204	-0.064	1.0000
LT_grid - EQM_grid	0.0236	0.0508	204	0.465	1.0000
LTQM_grid_C - EQM_IDW	-0.0615	0.0508	204	-1.209	1.0000
LTQM_grid_C - EQM_krig	-0.0677	0.0508	204	-1.332	1.0000
LTQM_grid_C - EQM_grid	-0.0409	0.0508	204	-0.804	1.0000
LTQM_grid_C - LT_grid	-0.0645	0.0508	204	-1.269	1.0000
LTQM_grid_V - EQM_IDW	0.1249	0.0508	204	2.458	0.2223
LTQM_grid_V - EQM_krig	0.1187	0.0508	204	2.334	0.3084
LTQM_grid_V - EQM_grid	0.1455	0.0508	204	2.863	0.0696
LTQM_grid_V - LT_grid	0.1219	0.0508	204	2.398	0.2608
LTQM_grid_V - LTQM_grid_C	0.1864	0.0508	204	3.667	0.0047

Table S1 cont.

Month = 10					
EQM_krig - EQM_IDW	0.0061	0.0508	204	0.120	1.0000
EQM_grid - EQM_IDW	-0.0041	0.0508	204	-0.081	1.0000
EQM_grid - EQM_krig	-0.0102	0.0508	204	-0.201	1.0000
LT_grid - EQM_IDW	-0.0792	0.0508	204	-1.558	1.0000
LT_grid - EQM_krig	-0.0853	0.0508	204	-1.679	1.0000
LT_grid - EQM_grid	-0.0751	0.0508	204	-1.477	1.0000
LTQM_grid_C - EQM_IDW	0.0097	0.0508	204	0.192	1.0000
LTQM_grid_C - EQM_krig	0.0036	0.0508	204	0.071	1.0000
LTQM_grid_C - EQM_grid	0.0139	0.0508	204	0.273	1.0000
LTQM_grid_C - LT_grid	0.0889	0.0508	204	1.750	1.0000
LTQM_grid_V - EQM_IDW	0.0578	0.0508	204	1.137	1.0000
LTQM_grid_V - EQM_krig	0.0517	0.0508	204	1.017	1.0000
LTQM_grid_V - EQM_grid	0.0619	0.0508	204	1.218	1.0000
LTQM_grid_V - LT_grid	0.1370	0.0508	204	2.695	0.1143
LTQM_grid_V - LTQM_grid_C	0.0481	0.0508	204	0.945	1.0000

Table S1 cont.

Month = 11					
EQM_krig - EQM_IDW	0.0261	0.0508	204	0.513	1.0000
EQM_grid - EQM_IDW	-0.0258	0.0508	204	-0.508	1.0000
EQM_grid - EQM_krig	-0.0519	0.0508	204	-1.021	1.0000
LT_grid - EQM_IDW	-0.1856	0.0508	204	-3.651	0.0050
LT_grid - EQM_krig	-0.2117	0.0508	204	-4.164	0.0007
LT_grid - EQM_grid	-0.1598	0.0508	204	-3.143	0.0288
LTQM_grid_C - EQM_IDW	-0.0608	0.0508	204	-1.195	1.0000
LTQM_grid_C - EQM_krig	-0.0868	0.0508	204	-1.708	1.0000
LTQM_grid_C - EQM_grid	-0.0349	0.0508	204	-0.687	1.0000
LTQM_grid_C - LT_grid	0.1249	0.0508	204	2.456	0.2232
LTQM_grid_V - EQM_IDW	-0.0257	0.0508	204	-0.506	1.0000
LTQM_grid_V - EQM_krig	-0.0518	0.0508	204	-1.019	1.0000
LTQM_grid_V - EQM_grid	0.0001	0.0508	204	0.002	1.0000
LTQM_grid_V - LT_grid	0.1599	0.0508	204	3.145	0.0286
LTQM_grid_V - LTQM_grid_C	0.0350	0.0508	204	0.689	1.0000
Month = 12					
EQM_krig - EQM_IDW	-0.0073	0.0508	204	-0.144	1.0000
EQM_grid - EQM_IDW	-0.0614	0.0508	204	-1.208	1.0000
EQM_grid - EQM_krig	-0.0541	0.0508	204	-1.064	1.0000
LT_grid - EQM_IDW	-0.3990	0.0508	204	-7.849	<.0001
LT_grid - EQM_krig	-0.3916	0.0508	204	-7.704	<.0001
LT_grid - EQM_grid	-0.3376	0.0508	204	-6.641	<.0001
LTQM_grid_C - EQM_IDW	-0.2600	0.0508	204	-5.114	<.0001
LTQM_grid_C - EQM_krig	-0.2526	0.0508	204	-4.969	<.0001
LTQM_grid_C - EQM_grid	-0.1985	0.0508	204	-3.906	0.0019
LTQM_grid_C - LT_grid	0.1390	0.0508	204	2.735	0.1018
LTQM_grid_V - EQM_IDW	-0.2493	0.0508	204	-4.904	<.0001
LTQM_grid_V - EQM_krig	-0.2420	0.0508	204	-4.760	0.0001
LTQM_grid_V - EQM_grid	-0.1879	0.0508	204	-3.696	0.0042
LTQM_grid_V - LT_grid	0.1497	0.0508	204	2.944	0.0542
LTQM_grid_V - LTQM_grid_C	0.0107	0.0508	204	0.210	1.0000

Results are averaged over the levels of: *Bias_correction_years*, *Elevation*
P value adjustment: Bonferroni method for 15 tests

Table 5.2: Table S2. Pairwise comparisons for the interaction Month × Method *in the PSS ANOVA model*

contrast	estimate	SE	df	t.ratio	p.value
Month = 1					
EQM_IDW - EQM_grid	-0.0026	0.0074	210	-0.357	1.0000
EQM_krig - EQM_grid	-0.0004	0.0074	210	-0.058	1.0000
EQM_krig - EQM_IDW	0.0022	0.0074	210	0.299	1.0000
LT_grid - EQM_grid	-0.0525	0.0074	210	-7.139	<.0001
LT_grid - EQM_IDW	-0.0499	0.0074	210	-6.782	<.0001
LT_grid - EQM_krig	-0.0521	0.0074	210	-7.081	<.0001
LTQM_grid_C - EQM_grid	-0.0173	0.0074	210	-2.355	0.2918
LTQM_grid_C - EQM_IDW	-0.0147	0.0074	210	-1.997	0.7061
LTQM_grid_C - EQM_krig	-0.0169	0.0074	210	-2.296	0.3396
LTQM_grid_C - LT_grid	0.0352	0.0074	210	4.784	<.0001
LTQM_grid_V - EQM_grid	-0.0152	0.0074	210	-2.070	0.5952
LTQM_grid_V - EQM_IDW	-0.0126	0.0074	210	-1.713	1.0000
LTQM_grid_V - EQM_krig	-0.0148	0.0074	210	-2.011	0.6833
LTQM_grid_V - LT_grid	0.0373	0.0074	210	5.069	<.0001
LTQM_grid_V - LTQM_grid_C	0.0021	0.0074	210	0.285	1.0000
Month = 2					
EQM_IDW - EQM_grid	0.0046	0.0074	210	0.629	1.0000
EQM_krig - EQM_grid	0.0068	0.0074	210	0.927	1.0000
EQM_krig - EQM_IDW	0.0022	0.0074	210	0.299	1.0000
LT_grid - EQM_grid	-0.0516	0.0074	210	-7.015	<.0001
LT_grid - EQM_IDW	-0.0562	0.0074	210	-7.644	<.0001
LT_grid - EQM_krig	-0.0584	0.0074	210	-7.942	<.0001
LTQM_grid_C - EQM_grid	-0.0469	0.0074	210	-6.371	<.0001
LTQM_grid_C - EQM_IDW	-0.0515	0.0074	210	-7.000	<.0001
LTQM_grid_C - EQM_krig	-0.0537	0.0074	210	-7.298	<.0001
LTQM_grid_C - LT_grid	0.0047	0.0074	210	0.644	1.0000
LTQM_grid_V - EQM_grid	-0.0467	0.0074	210	-6.344	<.0001
LTQM_grid_V - EQM_IDW	-0.0513	0.0074	210	-6.972	<.0001
LTQM_grid_V - EQM_krig	-0.0535	0.0074	210	-7.271	<.0001
LTQM_grid_V - LT_grid	0.0049	0.0074	210	0.671	1.0000
LTQM_grid_V - LTQM_grid_C	0.0002	0.0074	210	0.027	1.0000

Table S2 cont.

Month = 3					
EQM_IDW - EQM_grid	-0.0029	0.0074	210	-0.400	1.0000
EQM_krig - EQM_grid	-0.0008	0.0074	210	-0.112	1.0000
EQM_krig - EQM_IDW	0.0021	0.0074	210	0.289	1.0000
LT_grid - EQM_grid	-0.0925	0.0074	210	-12.577	<.0001
LT_grid - EQM_IDW	-0.0896	0.0074	210	-12.177	<.0001
LT_grid - EQM_krig	-0.0917	0.0074	210	-12.466	<.0001
LTQM_grid_C - EQM_grid	-0.0700	0.0074	210	-9.509	<.0001
LTQM_grid_C - EQM_IDW	-0.0670	0.0074	210	-9.108	<.0001
LTQM_grid_C - EQM_krig	-0.0691	0.0074	210	-9.397	<.0001
LTQM_grid_C - LT_grid	0.0226	0.0074	210	3.068	0.0365
LTQM_grid_V - EQM_grid	-0.0730	0.0074	210	-9.921	<.0001
LTQM_grid_V - EQM_IDW	-0.0701	0.0074	210	-9.520	<.0001
LTQM_grid_V - EQM_krig	-0.0722	0.0074	210	-9.809	<.0001
LTQM_grid_V - LT_grid	0.0195	0.0074	210	2.656	0.1275
LTQM_grid_V - LTQM_grid_C	-0.0030	0.0074	210	-0.412	1.0000
Month = 4					
EQM_IDW - EQM_grid	0.0010	0.0074	210	0.131	1.0000
EQM_krig - EQM_grid	-0.0043	0.0074	210	-0.584	1.0000
EQM_krig - EQM_IDW	-0.0053	0.0074	210	-0.715	1.0000
LT_grid - EQM_grid	-0.0407	0.0074	210	-5.530	<.0001
LT_grid - EQM_IDW	-0.0417	0.0074	210	-5.661	<.0001
LT_grid - EQM_krig	-0.0364	0.0074	210	-4.946	<.0001
LTQM_grid_C - EQM_grid	-0.0186	0.0074	210	-2.523	0.1857
LTQM_grid_C - EQM_IDW	-0.0195	0.0074	210	-2.654	0.1285
LTQM_grid_C - EQM_krig	-0.0143	0.0074	210	-1.939	0.8078
LTQM_grid_C - LT_grid	0.0221	0.0074	210	3.007	0.0444
LTQM_grid_V - EQM_grid	-0.0114	0.0074	210	-1.553	1.0000
LTQM_grid_V - EQM_IDW	-0.0124	0.0074	210	-1.684	1.0000
LTQM_grid_V - EQM_krig	-0.0071	0.0074	210	-0.969	1.0000
LTQM_grid_V - LT_grid	0.0293	0.0074	210	3.977	0.0014
LTQM_grid_V - LTQM_grid_C	0.0071	0.0074	210	0.970	1.0000

Table S2 cont.

Month = 5					
EQM_IDW - EQM_grid	0.0048	0.0074	210	0.648	1.0000
EQM_krig - EQM_grid	0.0016	0.0074	210	0.224	1.0000
EQM_krig - EQM_IDW	-0.0031	0.0074	210	-0.425	1.0000
LT_grid - EQM_grid	-0.0476	0.0074	210	-6.464	<.0001
LT_grid - EQM_IDW	-0.0523	0.0074	210	-7.113	<.0001
LT_grid - EQM_krig	-0.0492	0.0074	210	-6.688	<.0001
LTQM_grid_C - EQM_grid	0.0077	0.0074	210	1.041	1.0000
LTQM_grid_C - EQM_IDW	0.0029	0.0074	210	0.392	1.0000
LTQM_grid_C - EQM_krig	0.0060	0.0074	210	0.817	1.0000
LTQM_grid_C - LT_grid	0.0552	0.0074	210	7.505	<.0001
LTQM_grid_V - EQM_grid	0.0093	0.0074	210	1.270	1.0000
LTQM_grid_V - EQM_IDW	0.0046	0.0074	210	0.621	1.0000
LTQM_grid_V - EQM_krig	0.0077	0.0074	210	1.046	1.0000
LTQM_grid_V - LT_grid	0.0569	0.0074	210	7.734	<.0001
LTQM_grid_V - LTQM_grid_C	0.0017	0.0074	210	0.229	1.0000

Table S2 cont.

Month = 6					
EQM_IDW - EQM_grid	0.0026	0.0074	210	0.356	1.0000
EQM_krig - EQM_grid	0.0097	0.0074	210	1.315	1.0000
EQM_krig - EQM_IDW	0.0071	0.0074	210	0.960	1.0000
LT_grid - EQM_grid	-0.0175	0.0074	210	-2.375	0.2770
LT_grid - EQM_IDW	-0.0201	0.0074	210	-2.730	0.1030
LT_grid - EQM_krig	-0.0272	0.0074	210	-3.690	0.0043
LTQM_grid_C - EQM_grid	0.0310	0.0074	210	4.219	0.0005
LTQM_grid_C - EQM_IDW	0.0284	0.0074	210	3.863	0.0022
LTQM_grid_C - EQM_krig	0.0214	0.0074	210	2.904	0.0613
LTQM_grid_C - LT_grid	0.0485	0.0074	210	6.593	<.0001
LTQM_grid_V - EQM_grid	0.0247	0.0074	210	3.355	0.0141
LTQM_grid_V - EQM_IDW	0.0221	0.0074	210	3.000	0.0454
LTQM_grid_V - EQM_krig	0.0150	0.0074	210	2.040	0.6389
LTQM_grid_V - LT_grid	0.0422	0.0074	210	5.730	<.0001
LTQM_grid_V - LTQM_grid_C	-0.0064	0.0074	210	-0.863	1.0000
Month = 7					
EQM_IDW - EQM_grid	0.0051	0.0074	210	0.698	1.0000
EQM_krig - EQM_grid	-0.0073	0.0074	210	-0.998	1.0000
EQM_krig - EQM_IDW	-0.0125	0.0074	210	-1.696	1.0000
LT_grid - EQM_grid	-0.0231	0.0074	210	-3.137	0.0293
LT_grid - EQM_IDW	-0.0282	0.0074	210	-3.835	0.0025
LT_grid - EQM_krig	-0.0157	0.0074	210	-2.139	0.5041
LTQM_grid_C - EQM_grid	0.0217	0.0074	210	2.955	0.0522
LTQM_grid_C - EQM_IDW	0.0166	0.0074	210	2.258	0.3751
LTQM_grid_C - EQM_krig	0.0291	0.0074	210	3.954	0.0016
LTQM_grid_C - LT_grid	0.0448	0.0074	210	6.092	<.0001
LTQM_grid_V - EQM_grid	0.0140	0.0074	210	1.899	0.8844
LTQM_grid_V - EQM_IDW	0.0088	0.0074	210	1.201	1.0000
LTQM_grid_V - EQM_krig	0.0213	0.0074	210	2.897	0.0625
LTQM_grid_V - LT_grid	0.0371	0.0074	210	5.036	<.0001
LTQM_grid_V - LTQM_grid_C	-0.0078	0.0074	210	-1.056	1.0000

Table S2 cont.

Month = 8					
EQM_IDW - EQM_grid	-0.0001	0.0074	210	-0.017	1.0000
EQM_krig - EQM_grid	0.0039	0.0074	210	0.529	1.0000
EQM_krig - EQM_IDW	0.0040	0.0074	210	0.545	1.0000
LT_grid - EQM_grid	-0.0396	0.0074	210	-5.375	<.0001
LT_grid - EQM_IDW	-0.0394	0.0074	210	-5.358	<.0001
LT_grid - EQM_krig	-0.0434	0.0074	210	-5.904	<.0001
LTQM_grid_C - EQM_grid	0.0013	0.0074	210	0.174	1.0000
LTQM_grid_C - EQM_IDW	0.0014	0.0074	210	0.191	1.0000
LTQM_grid_C - EQM_krig	-0.0026	0.0074	210	-0.355	1.0000
LTQM_grid_C - LT_grid	0.0408	0.0074	210	5.549	<.0001
LTQM_grid_V - EQM_grid	-0.0068	0.0074	210	-0.924	1.0000
LTQM_grid_V - EQM_IDW	-0.0067	0.0074	210	-0.907	1.0000
LTQM_grid_V - EQM_krig	-0.0107	0.0074	210	-1.453	1.0000
LTQM_grid_V - LT_grid	0.0328	0.0074	210	4.451	0.0002
LTQM_grid_V - LTQM_grid_C	-0.0081	0.0074	210	-1.098	1.0000
Month = 9					
EQM_IDW - EQM_grid	0.0039	0.0074	210	0.528	1.0000
EQM_krig - EQM_grid	0.0048	0.0074	210	0.647	1.0000
EQM_krig - EQM_IDW	0.0009	0.0074	210	0.119	1.0000
LT_grid - EQM_grid	-0.0090	0.0074	210	-1.225	1.0000
LT_grid - EQM_IDW	-0.0129	0.0074	210	-1.753	1.0000
LT_grid - EQM_krig	-0.0138	0.0074	210	-1.872	0.9384
LTQM_grid_C - EQM_grid	0.0271	0.0074	210	3.679	0.0045
LTQM_grid_C - EQM_IDW	0.0232	0.0074	210	3.151	0.0279
LTQM_grid_C - EQM_krig	0.0223	0.0074	210	3.032	0.0410
LTQM_grid_C - LT_grid	0.0361	0.0074	210	4.904	<.0001
LTQM_grid_V - EQM_grid	0.0308	0.0074	210	4.192	0.0006
LTQM_grid_V - EQM_IDW	0.0270	0.0074	210	3.663	0.0047
LTQM_grid_V - EQM_krig	0.0261	0.0074	210	3.544	0.0073
LTQM_grid_V - LT_grid	0.0399	0.0074	210	5.417	<.0001
LTQM_grid_V - LTQM_grid_C	0.0038	0.0074	210	0.512	1.0000

Table S2 cont.

Month = 10					
EQM_IDW - EQM_grid	0.0057	0.0074	210	0.781	1.0000
EQM_krig - EQM_grid	0.0102	0.0074	210	1.379	1.0000
EQM_krig - EQM_IDW	0.0044	0.0074	210	0.599	1.0000
LT_grid - EQM_grid	-0.0211	0.0074	210	-2.869	0.0681
LT_grid - EQM_IDW	-0.0269	0.0074	210	-3.649	0.0050
LT_grid - EQM_krig	-0.0313	0.0074	210	-4.248	0.0005
LTQM_grid_C - EQM_grid	0.0260	0.0074	210	3.536	0.0075
LTQM_grid_C - EQM_IDW	0.0203	0.0074	210	2.755	0.0957
LTQM_grid_C - EQM_krig	0.0159	0.0074	210	2.157	0.4826
LTQM_grid_C - LT_grid	0.0471	0.0074	210	6.405	<.0001
LTQM_grid_V - EQM_grid	0.0254	0.0074	210	3.457	0.0099
LTQM_grid_V - EQM_IDW	0.0197	0.0074	210	2.676	0.1206
LTQM_grid_V - EQM_krig	0.0153	0.0074	210	2.077	0.5851
LTQM_grid_V - LT_grid	0.0465	0.0074	210	6.325	<.0001
LTQM_grid_V - LTQM_grid_C	-0.0006	0.0074	210	-0.080	1.0000

Month = 11					
EQM_IDW - EQM_grid	0.0013	0.0074	210	0.181	1.0000
EQM_krig - EQM_grid	0.0066	0.0074	210	0.903	1.0000
EQM_krig - EQM_IDW	0.0053	0.0074	210	0.722	1.0000
LT_grid - EQM_grid	-0.0177	0.0074	210	-2.410	0.2519
LT_grid - EQM_IDW	-0.0191	0.0074	210	-2.591	0.1534
LT_grid - EQM_krig	-0.0244	0.0074	210	-3.314	0.0163
LTQM_grid_C - EQM_grid	0.0159	0.0074	210	2.162	0.4763
LTQM_grid_C - EQM_IDW	0.0146	0.0074	210	1.981	0.7336
LTQM_grid_C - EQM_krig	0.0093	0.0074	210	1.259	1.0000
LTQM_grid_C - LT_grid	0.0336	0.0074	210	4.572	0.0001
LTQM_grid_V - EQM_grid	0.0167	0.0074	210	2.271	0.3620
LTQM_grid_V - EQM_IDW	0.0154	0.0074	210	2.090	0.5667
LTQM_grid_V - EQM_krig	0.0101	0.0074	210	1.368	1.0000
LTQM_grid_V - LT_grid	0.0345	0.0074	210	4.682	0.0001
LTQM_grid_V - LTQM_grid_C	0.0008	0.0074	210	0.110	1.0000
Month = 12					
EQM_IDW - EQM_grid	-0.0024	0.0074	210	-0.332	1.0000
EQM_krig - EQM_grid	-0.0036	0.0074	210	-0.488	1.0000
EQM_krig - EQM_IDW	-0.0012	0.0074	210	-0.157	1.0000
LT_grid - EQM_grid	-0.0136	0.0074	210	-1.844	0.9984
LT_grid - EQM_IDW	-0.0111	0.0074	210	-1.513	1.0000
LT_grid - EQM_krig	-0.0100	0.0074	210	-1.356	1.0000
LTQM_grid_C - EQM_grid	0.0049	0.0074	210	0.662	1.0000
LTQM_grid_C - EQM_IDW	0.0073	0.0074	210	0.993	1.0000
LTQM_grid_C - EQM_krig	0.0085	0.0074	210	1.150	1.0000
LTQM_grid_C - LT_grid	0.0184	0.0074	210	2.506	0.1947
LTQM_grid_V - EQM_grid	0.0073	0.0074	210	0.989	1.0000
LTQM_grid_V - EQM_IDW	0.0097	0.0074	210	1.320	1.0000
LTQM_grid_V - EQM_krig	0.0109	0.0074	210	1.477	1.0000
LTQM_grid_V - LT_grid	0.0208	0.0074	210	2.833	0.0759
LTQM_grid_V - LTQM_grid_C	0.0024	0.0074	210	0.327	1.0000

Results are averaged over the levels of: *Bias_correction_years*
P value adjustment: bonferroni method for 15 tests

Table 5.3: Pairwise comparisons for the interaction Bias_correction_years × Method *in the RMSE ANOVA model*

contrast	estimate	SE	df	t.ratio	p.value
Bias_correction_years = 1980-1989					
EQM_krig - EQM_IDW	-0.0062	0.0208	204	-0.300	1.0000
EQM_grid - EQM_IDW	-0.0219	0.0208	204	-1.053	1.0000
EQM_grid - EQM_krig	-0.0156	0.0208	204	-0.753	1.0000
LT_grid - EQM_IDW	-0.1697	0.0208	204	-8.179	<.0001
LT_grid - EQM_krig	-0.1635	0.0208	204	-7.879	<.0001
LT_grid - EQM_grid	-0.1479	0.0208	204	-7.126	<.0001
LTQM_grid_C - EQM_IDW	-0.0557	0.0208	204	-2.685	0.1177
LTQM_grid_C - EQM_krig	-0.0495	0.0208	204	-2.385	0.2696
LTQM_grid_C - EQM_grid	-0.0339	0.0208	204	-1.632	1.0000
LTQM_grid_C - LT_grid	0.1140	0.0208	204	5.494	<.0001
LTQM_grid_V - EQM_IDW	0.0135	0.0208	204	0.652	1.0000
LTQM_grid_V - EQM_krig	0.0197	0.0208	204	0.951	1.0000
LTQM_grid_V - EQM_grid	0.0354	0.0208	204	1.705	1.0000
LTQM_grid_V - LT_grid	0.1833	0.0208	204	8.830	<.0001
LTQM_grid_V - LTQM_grid_C	0.0692	0.0208	204	3.337	0.0151
Bias_correction_years = 1980-2014					
EQM_krig - EQM_IDW	0.0102	0.0208	204	0.492	1.0000
EQM_grid - EQM_IDW	-0.0361	0.0208	204	-1.738	1.0000
EQM_grid - EQM_krig	-0.0463	0.0208	204	-2.230	0.4023
LT_grid - EQM_IDW	-0.0649	0.0208	204	-3.127	0.0304
LT_grid - EQM_krig	-0.0751	0.0208	204	-3.619	0.0056
LT_grid - EQM_grid	-0.0288	0.0208	204	-1.389	1.0000
LTQM_grid_C - EQM_IDW	0.0159	0.0208	204	0.767	1.0000
LTQM_grid_C - EQM_krig	0.0057	0.0208	204	0.274	1.0000
LTQM_grid_C - EQM_grid	0.0520	0.0208	204	2.505	0.1957
LTQM_grid_C - LT_grid	0.0808	0.0208	204	3.894	0.0020
LTQM_grid_V - EQM_IDW	0.1103	0.0208	204	5.313	<.0001
LTQM_grid_V - EQM_krig	0.1000	0.0208	204	4.821	<.0001
LTQM_grid_V - EQM_grid	0.1463	0.0208	204	7.051	<.0001
LTQM_grid_V - LT_grid	0.1752	0.0208	204	8.440	<.0001
LTQM_grid_V - LTQM_grid_C	0.0944	0.0208	204	4.547	0.0001

Results are averaged over the levels of: *Month*, *Elevation*
P value adjustment: bonferroni method for 15 tests

Table 5.4: Pairwise comparisons for the interaction Bias_correction_years × Method *in the PSS ANOVA model*

contrast	estimate	SE	df	t.ratio	p.value
Bias_correction_years = 1980-1989					
EQM_krig - EQM_IDW	-0.0005	0.0030	210	-0.182	1.0000
EQM_grid - EQM_IDW	-0.0015	0.0030	210	-0.490	1.0000
EQM_grid - EQM_krig	-0.0009	0.0030	210	-0.308	1.0000
LT_grid - EQM_IDW	-0.0182	0.0030	210	-6.044	<.0001
LT_grid - EQM_krig	-0.0176	0.0030	210	-5.862	<.0001
LT_grid - EQM_grid	-0.0167	0.0030	210	-5.554	<.0001
LTQM_grid_C - EQM_IDW	0.0101	0.0030	210	3.375	0.0132
LTQM_grid_C - EQM_krig	0.0107	0.0030	210	3.557	0.0070
LTQM_grid_C - EQM_grid	0.0116	0.0030	210	3.864	0.0022
LTQM_grid_C - LT_grid	0.0283	0.0030	210	9.418	<.0001
LTQM_grid_V - EQM_IDW	0.0098	0.0030	210	3.263	0.0193
LTQM_grid_V - EQM_krig	0.0103	0.0030	210	3.445	0.0104
LTQM_grid_V - EQM_grid	0.0113	0.0030	210	3.753	0.0034
LTQM_grid_V - LT_grid	0.0280	0.0030	210	9.306	<.0001
LTQM_grid_V - LTQM_grid_C	-0.0003	0.0030	210	-0.112	1.0000
Bias_correction_years = 1980-2014					
EQM_krig - EQM_IDW	0.0016	0.0030	210	0.524	1.0000
EQM_grid - EQM_IDW	-0.0020	0.0030	210	-0.672	1.0000
EQM_grid - EQM_krig	-0.0036	0.0030	210	-1.196	1.0000
LT_grid - EQM_IDW	-0.0564	0.0030	210	-18.781	<.0001
LT_grid - EQM_krig	-0.0580	0.0030	210	-19.305	<.0001
LT_grid - EQM_grid	-0.0544	0.0030	210	-18.109	<.0001
LTQM_grid_C - EQM_IDW	-0.0165	0.0030	210	-5.487	<.0001
LTQM_grid_C - EQM_krig	-0.0181	0.0030	210	-6.012	<.0001
LTQM_grid_C - EQM_grid	-0.0145	0.0030	210	-4.816	<.0001
LTQM_grid_C - LT_grid	0.0399	0.0030	210	13.293	<.0001
LTQM_grid_V - EQM_IDW	-0.0174	0.0030	210	-5.804	<.0001
LTQM_grid_V - EQM_krig	-0.0190	0.0030	210	-6.328	<.0001
LTQM_grid_V - EQM_grid	-0.0154	0.0030	210	-5.132	<.0001
LTQM_grid_V - LT_grid	0.0390	0.0030	210	12.977	<.0001
LTQM_grid_V - LTQM_grid_C	-0.0010	0.0030	210	-0.317	1.0000

Results are averaged over the levels of: *Month*
P value adjustment: bonferroni method for 15 tests

Table 5.5: Pairwise comparisons for the interaction Elevation × Method *in the RMSE ANOVA model*

contrast	estimate	SE	df	t.ratio	p.value
Elevation = NO					
EQM_krig - EQM_IDW	-0.0016	0.0208	204	-0.079	1.0000
EQM_grid - EQM_IDW	-0.0220	0.0208	204	-1.061	1.0000
EQM_grid - EQM_krig	-0.0204	0.0208	204	-0.982	1.0000
LT_grid - EQM_IDW	-0.1352	0.0208	204	-6.515	<.0001
LT_grid - EQM_krig	-0.1336	0.0208	204	-6.436	<.0001
LT_grid - EQM_grid	-0.1132	0.0208	204	-5.454	<.0001
LTQM_grid_C - EQM_IDW	-0.0570	0.0208	204	-2.747	0.0982
LTQM_grid_C - EQM_krig	-0.0554	0.0208	204	-2.668	0.1237
LTQM_grid_C - EQM_grid	-0.0350	0.0208	204	-1.686	1.0000
LTQM_grid_C - LT_grid	0.0782	0.0208	204	3.768	0.0032
LTQM_grid_V - EQM_IDW	0.0604	0.0208	204	2.913	0.0597
LTQM_grid_V - EQM_krig	0.0621	0.0208	204	2.992	0.0467
LTQM_grid_V - EQM_grid	0.0825	0.0208	204	3.974	0.0015
LTQM_grid_V - LT_grid	0.1957	0.0208	204	9.428	<.0001
LTQM_grid_V - LTQM_grid_C	0.1175	0.0208	204	5.660	<.0001
Elevation = YES					
EQM_krig - EQM_IDW	0.0056	0.0208	204	0.272	1.0000
EQM_grid - EQM_IDW	-0.0359	0.0208	204	-1.730	1.0000
EQM_grid - EQM_krig	-0.0416	0.0208	204	-2.002	0.6990
LT_grid - EQM_IDW	-0.0994	0.0208	204	-4.791	<.0001
LT_grid - EQM_krig	-0.1051	0.0208	204	-5.063	<.0001
LT_grid - EQM_grid	-0.0635	0.0208	204	-3.061	0.0376
LTQM_grid_C - EQM_IDW	0.0172	0.0208	204	0.829	1.0000
LTQM_grid_C - EQM_krig	0.0116	0.0208	204	0.557	1.0000
LTQM_grid_C - EQM_grid	0.0531	0.0208	204	2.559	0.1685
LTQM_grid_C - LT_grid	0.1166	0.0208	204	5.619	<.0001
LTQM_grid_V - EQM_IDW	0.0633	0.0208	204	3.052	0.0386
LTQM_grid_V - EQM_krig	0.0577	0.0208	204	2.780	0.0891
LTQM_grid_V - EQM_grid	0.0992	0.0208	204	4.782	<.0001
LTQM_grid_V - LT_grid	0.1628	0.0208	204	7.843	<.0001
LTQM_grid_V - LTQM_grid_C	0.0461	0.0208	204	2.223	0.4092

Results are averaged over the levels of: *Month*, *Bias_correction_years*
P value adjustment: bonferroni method for 15 tests

Physics settings for the WRF model

Table 5.6: Physics settings and details for the WRF model

Setting	Details
Microphysics	WRF Single–moment 6–class Scheme (Hong, S.–Y., and J.–O. J. Lim, 2006: The WRF single–moment 6–class microphysics scheme (WSM6). J. Korean Meteor. Soc., 42, 129–151.)
Radiation	RRTMG Shortwave and Longwave Schemes (Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long–lived greenhouse gases: Calculations with the AER radiative transfer models. J. Geophys. Res., 113, D13103. doi:10.1029/2008JD009944)
Boundary layer	Mellor–Yamada–Janjic Scheme (MYJ) (Janjic, Zavisa I., 1994: The Step–Mountain Eta Coordinate Model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. Mon. Wea. Rev., 122, 927–945. doi:10.1175/1520-0493(1994)122%3c0927:TSMECM%3e2.0.CO;2)
Cumulus convection	New Simplified Arakawa–Schubert Scheme (for Basic WRF) (Han, Jongil and Hua–Lu Pan, 2011: Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. Wea. Forecasting, 26, 520–533. doi:10.1175/WAF-D-10-05038.1)
Land surface physics	Unified Noah Land Surface Model (Tewari, M., F. Chen, W. Wang, J. Dudhia, M. A. LeMone, K. Mitchell, M. Ek, G. Gayno, J. Wegiel, and R. H. Cuenca, 2004: Implementation and verification of the unified NOAH land surface model in the WRF model. 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, pp. 11–15.)
Surface layer physics	Eta Similarity Scheme (Janjic, Z. I., 1994: The step-mountain Eta coordinate model: further developments of the convection, viscous sublayer and turbulence closure schemes. Mon. Wea. Rev., 122, 927–945. doi:10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2)

Statistical methods

Bayesian kriging

For LT_grid, LTQM_grid_C, and LTQM_grid_V, estimated slope and intercept parameters from transfer functions were kriged to the fine-scale grid using Bayesian kriging. One assumption of Bayesian spatial hierarchical models is that random spatial variates can be modeled by unique Gaussian spatial processes, $Y(s)$, with mean $\mu(s) = E(Y(s))$, and where the measurement locations $\{s_1...s_n\}$ are, in this study, WRF center grid points. In Gaussian spatial processes, observations $Y = \{s_1...s_n\}$ are assumed to follow a multivariate normal distribution [209]:

$$Y|\mu, \theta \sim N_n(\mu\mathbf{1}, \sigma(\theta)),$$

where N_n connotes the N dimensional normal distribution, μ is the constant mean, $\sigma(\theta)_{ij}$ gives the covariance between $Y(s_i)$ and $Y(s_j)$, and $\theta = (\tau^2, \sigma^2, \phi)^T$ is a vector of spatial parameters upon which the covariance matrix depends. For methods LT_-grid, LTQM_grid_C, and LTQM_grid_V, the response variables were monthly slope and intercept estimated parameters. For each month, estimated slopes and intercepts were kriged from station locations to the fine-scale grid. We used a Bayesian spatial hierarchical model of the form:

$$\boldsymbol{Y}(s) = \mu(\mathbf{s}) + w(\mathbf{s}) + \epsilon(\mathbf{s}), \tag{5.1}$$

where $Y(s)$ is the response at location s having a mean structure $\mu(\mathbf{s}) = \mathbf{x}^T(\mathbf{s})\boldsymbol{\beta}$.

The implementation of a full Bayesian spatial model is computationally intensive, due to the inversion of large $(n \times n)$ covariance matrices [209]. To decrease computation time, we instead used a nearest-neighbor Gaussian process model (NNGP), which is computationally more efficient than the full Gaussian process model in (5.1). In NNGP models, the spatial process is estimated based on its m nearest neighbors, but inference is nearly identical to than the model in (5.1) [210]. The `spNNGP` function from the `spNNGP` package in R constructs an NNGP model [210]. In this function, Markov chain Monte Carlo (MCMC) sampling approximates the posterior distribution of the parameter vector θ by fitting the marginalized model $f(y|\theta)p(\theta)$, which integrates over the spatial effects vector W and regression coefficients. The `spNNGP` function allows σ^2 and the ratio τ^2/σ^2 to vary, making it a flexible model [209]. Predictions were made by passing the resulting model fit from `spNNGP` to the `spPredict` function [210], which carries out Bayesian kriging.

Based on inspection of empirical variograms, we used the exponential covariance function for fitting all models.

$$C(t) = \begin{Bmatrix} \tau^2 + \sigma^2 & \text{if} \quad d = 0 \\ \sigma^2 exp(-\phi d) & \text{if} \quad d > 0 \end{Bmatrix} \tag{5.2}$$

In (5.2), $||h|| = d$, and ϕ, τ^2, and σ^2 are the effective range, nugget effect, and partial sill, respectively. The exponential covariance function reaches 0 asymptotically, so the effective range, rather than the range, must be used. The effective range d_0, can be obtained by setting $\exp(-\phi d) = 0.05$, which yields $d_0 = \frac{3}{\phi}$.

We used non-informative priors for the intercept (β_0), partial sill (σ^2), and nugget (τ^2), and we used an informative prior for the effective range (ϕ). We inspected

variograms of residuals from a simple linear model fit to determine appropriate initial values for the effective range, ϕ. For Bayesian kriging in EQM_grid, the following priors were used:

$$\begin{aligned}\beta_0 &\sim N(0, 100)\\ \phi &\sim Unif(\frac{3}{D_{max}}, \frac{3}{10})\\ \sigma^2 &\sim InvGamma(2, 2)\\ \tau^2 &\sim InvGamma(2, 0.1),\end{aligned}$$

where D_{max} was the maximum distance between any two GHCND station locations. The priors for all Bayesian kriging implemented in methods LT_grid, LTQM_grid_V, and LTQM_grid_V were the same as those for EQM_grid, except we used an $InvGamma(2, 0.02)$ prior for τ^2. All daily NNGP kriging models were run with 5000 MCMC samples with a burn-in of 1250 iterations and $m = 15$ nearest neighbors.

Kriging

In methods EQM_grid and EQM_krig daily GHCND station values and bias-corrected WRF values at station locations were interpolated to the 1km grid with non-Bayesian kriging. Non-Bayesian kriging can also be understood in the context of Gaussian processes. As in (5.1), spatial variates $Y = s_1...s_n$ are assumed to follow a multivariate normal distribution [112]. A general expression for the spatial model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \text{ where } \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{5.3}$$

where the covariance matrix, assuming a nugget effect τ^2 is $\Sigma = \sigma^2 H(\phi) + \tau^2 I$, and $H(\phi)_{ij} = \rho(\phi; d_{ij})$ for a valid correlation function ρ. The function $h(y)$ that minimizes the mean square error, $E\left[(Y(s_0) - h(y))^2 | y\right]$, is called the kriging predictor.

It can be shown (e.g. [112]) that the kriging predictor at a new location, $Y^*(s_0)$, takes the form:

$$Y^*(s_0) = \sum_{i=1}^{N} \lambda_i Y(s_i),$$

where s_0 is a new location at which a prediction is to be made, and λ_i are weights chosen such that they satisfy the conditions of unbiasedness and minimize the kriging variance [112]. Unlike Bayesian kriging, covariance parameters must be estimated from the data.

Based on inspection of empirical variograms, we used the exponential covariance function (5.2) for all model fits ($\phi = 150$ km, $\sigma^2 = 15$, and $\tau^2 = 0.2$).

Inverse distance weighting

Inverse distance weighting (IDW) is a deterministic interpolation technique, so the size of prediction errors cannot be quantified [136]. Interpolated values are based on a weighted average of n nearest-neighbor observations. In IDW, observed values close to prediction locations are assumed to be more influential in the prediction compared to observed values far from prediction locations. As the power, p and the number of nearest neighbors n increases, the smoothness of the interpolated surface increases. IDW is an exact interpolator, which means that if a prediction location, s_0 corresponds to an observed location s_i, the predicted value at s_0 will be identical to the value at location s_i. The general equations for IDW are as follows:

$$Y(s_0) = \sum_{i=1}^{n} w_i(s_0) Y(s_i),$$
$$w_i(s_0) = \frac{\tilde{w}_i(s_0)}{\sum_{\ell=1}^{n} w_\ell(s_0)},$$
$$\tilde{w}_i(s_0) = \frac{1}{d(s_i, s_0)^p}.$$

The IDW interpolated value at location s_0 is $Y(s_0)$, $d(s_i, s_0)$ is the distance between observed location s_i and prediction location s_0, n is the number of nearest-neighbor observed locations that contribute to the interpolated value $Y(s_0)$, and p is the power parameter.

Topographic downscaling

Topographic downscaling is a variation on IDW that is often used for high resolution downscaling [31]. Topographic downscaling consists of three main steps:

1. Construction of a historical, empirical lapse rate (relationship between TMAX and elevation using simple linear regression);

2. Adjustment of WRF data to reference elevation (200m) using estimated lapse rate parameters and interpolation of adjusted WRF data to desired locations;

3. Back-transformation of interpolated values using estimated lapse rate parameters.

Following methods by [31] and [12], we utilized historical (1970-1999) GHCND station records to calculate historical, elevational lapse rates for TMAX, using stations

with at least 70% complete records. We estimated the elevational lapse rates for TMAX with a linear regression of the form (5.4):

$$T_{sta} = T_0 - \beta\phi_{sta} - \gamma z_{sta}, \tag{5.4}$$

where T_{sta} is the long-term average station TMAX, T_0 is the intercept, β is the coefficient for GHNCD station latitude (ϕ_{sta}), and γ is the coefficient for station elevation (z_{sta}).

Estimates for β and γ were -1.43 and -0.0059, respectively (Figure 5.1). The estimate of the elevation coefficient, β, refers to an elevational lapse rate of 5.9°Ckm^{-1}, which corresponds closely to that found by [31], as well as the standard elevational lapse rate (6.0°Ckm^{-1}) [133].

WRF TMAX simulations were translated to reference elevation with (5.5)

$$T_{WRF,ref} = T_{WRF} - \gamma(z_{ref} - z_{WRF}), \tag{5.5}$$

where $T_{model,ref}$ is the value of TMAX (° C) at reference elevation, T_{WRF} is the WRF TMAX value (°C), γ is the estimated lapse rate (°Cm^{-1}) from 5.4, z_{ref} is the reference elevation (m), and z_{WRF} is WRF geopotential height (m).

Next, the transformed WRF data were interpolated to GHNCD station locations using IDW. We used a weight of 2 and 9 nearest neighbors for all IDW interpolation, following methods by [31]. Interpolated WRF data were back-transformed to reflect the effect of elevation (5.6)

$$T_{sta,interp} = T_{ref,interp} - \gamma(z_{sta} - z_{ref}). \tag{5.6}$$

In (5.6) $T_{sta,interp}$ is the elevation-adjusted value for TMAX, $T_{ref,interp}$ is the interpolated WRF value at a GHCND station location at reference elevation and z_{sta} and z_{ref} are the GHCND station and reference elevations. After back-transforming interpolated values at GHCND station locations, we applied empirical quantile mapping (EQM) at each station location.

$$X_{corr,t} = \text{ecdf}^{-1}_{obs,m}(\text{ecdf}_{raw,m}(X_{raw,t})), \tag{5.7}$$

In (5.7), $X_{corr,t}$ is the corrected daily value for TMAX on day t, $\text{ecdf}^{-1}_{obs,m}$ is the inverse ecdf of GHCND station data for month m, and $\text{ecdf}_{raw,m}$ is the ecdf of the WRF data for month m, and $X_{raw,t}$ is the uncorrected WRF TMAX value on day t. Next, bias-corrected WRF data at GHCND station locations were translated to reference elevation with (5.8)

$$T_{EQM,ref} = T_{EQM} - \gamma(z_{ref} - z_{sta}), \tag{5.8}$$

where $T_{EQM,ref}$ is the bias corrected, interpolated value for TMAX (° C) at reference elevation, T_{EQM} is the bias corrected, WRF interpolation at a GHCND station location (°C), and γ, z_{ref} and z_{sta} are as defined in (5.6). Finally, the reference-adjusted, bias- corrected WRF interpolations at GHCND station locations were again interpolated to a 1km grid using IDW:

$$Y(s_0) = \sum_{i=1}^{n} w_i(s_0) Y(s_i),$$
$$w_i(s_0) = \frac{\tilde{w}_i(s_0)}{\sum_{\ell=1}^{n} w_\ell(s_0)},$$
$$\tilde{w}_i(s_0) = \frac{1}{d(s_i, s_0)^p}.$$

In this context, $Y(s_0)$ is the IDW interpolated TMAX value at fine-scale grid cell s_0, $Y(s_i)$ is the value at station location s_i, $d(s_0, s_i)$ is the distance between GHCND station location s_i and the center of fine-scale grid cell s_0, and n and p were set to 9 and 2, respectively. Finally, the high-resolution values were translated to actual elevation with (5.9)

$$T_{fine,\,interp} = T_{ref,\,interp} - \gamma(z_{fine} - z_{ref}). \tag{5.9}$$

In (5.9), $T_{fine,\,interp}$ is the final downscaled value on the fine-scale grid, $T_{ref,\,interp}$ is the interpolated temperature value at reference elevation, z_{fine} is the elevation at the fine-scale grid, and γ and z_{ref} are as defined in (5.6).

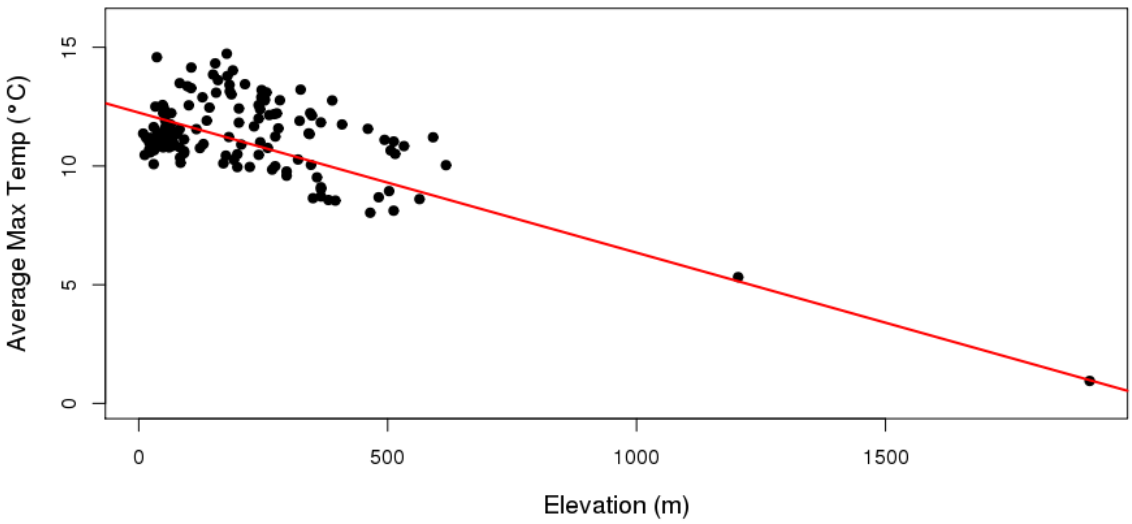

Figure 5.1: Elevational lapse rate adjustment for TMAX. Note: the elevational lapse rate did not change appreciably with omission of the two high elevation stations.

Evaluation of downscaling and bias-correction methods using an alternative performance metric: sorted RMSE

An alternative metric for assessing distributional similarity of observed and simulated data is sorted RMSE. Since PSS is more widely in the climate literature than sorted RMSE, we reported results for PSS in the main manuscript. Sorted RMSE (SRMSE) was calculated in the same way as RMSE, except that both bias-corrected and observed daily TMAX values were sorted from least to greatest prior to calculations. The full ANOVA model included the same variables as the models for RMSE and

PSS (Table 5.8). The final ANOVA model included the main effects *Month*, *Bias_correction_years*, and *Method* as well as the interaction terms *Month* × *Method* and *Bias_correction_years* × *Method* (Table 5.7).

The results for sorted RMSE were very similar to those of PSS. Generally, SRMSE values improved when bias-correction was based on the 1980-2014 GHCND dataset compared to the 1980-1989 GHCND dataset. LT_grid performed worst overall regardless of whether the 1980-1989 or 1980-2014 GHCND dataset was used for bias-correction (Figures 5.2 and 5.3). Similar to PSS, SRMSE exhibited less monthly variation.

The interaction of *Month* × *Method* was significant, and the interaction was most apparent for LT_grid, and estimated mean marginal SRMSE of LT_grid was significantly greater than that of all other methods in months 1-5 (Figure 5.4).

The interaction *Method* × *Bias_correction_years* was significant, and the interaction plot (Figure 5.5) shows that while EQM_IDW, EQM_krig, and EQM_grid performed better when bias correction was done with the 1980-2014 GHCND dataset, LTQM_grid_C and LTQM_grid_V performed better when the 1980-1989 GHCND dataset was used for bias-correction. LT_grid performed worst overall regardless of whether the 1980-2014 or 1980-1989 GHCND dataset was used for bias-correction. Estimated mean marginal SRMSE was significantly greater than that of all other methods when either the 1980-2014 or 1980-1989 GHCND dataset was used for bias-correction. Similar to results for PSS, the interaction *Method* × *Elevation* was not significant.

Table 5.7: ANOVA table for final SRMSE model.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Month	11	2.56	0.23	10.01	0.0000
Bias_correction_years	1	6.09	6.09	261.56	0.0000
Method	5	8.30	1.66	71.37	0.0000
Month × Method	55	5.07	0.09	3.96	0.0000
Method × Bias_correction_years	5	3.82	0.76	32.79	0.0000
Residuals	210	4.89	0.02		

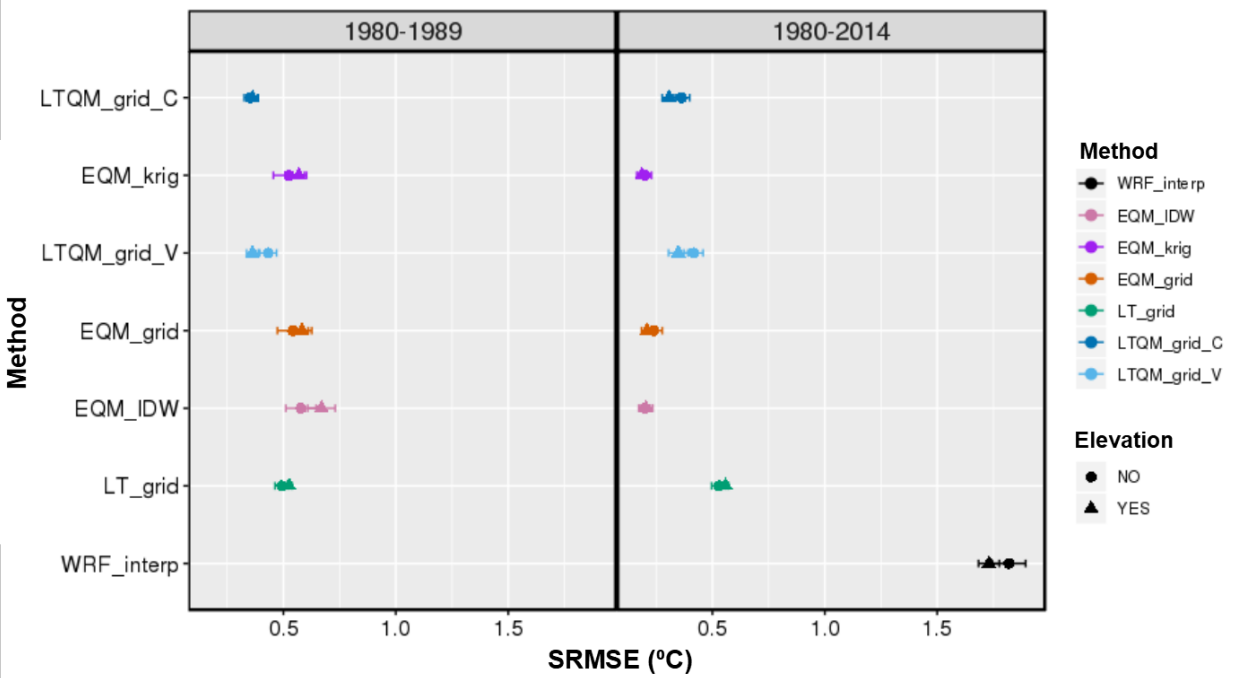

Figure 5.2: Mean SRMSE by Method *and* Bias_correction_years*, where "1980-1989" and "1980-2014" denote the GHCND station datasets used to bias-correct 1990-2014 and 1980-2014 WRF simulations, respectively. Error bars represent standard errors over five spatial cross-validation folds. "WRF_interp" denotes raw WRF simulations interpolated to station locations and are shown to indicate relative improvement of all methods over raw WRF interpolated values.*

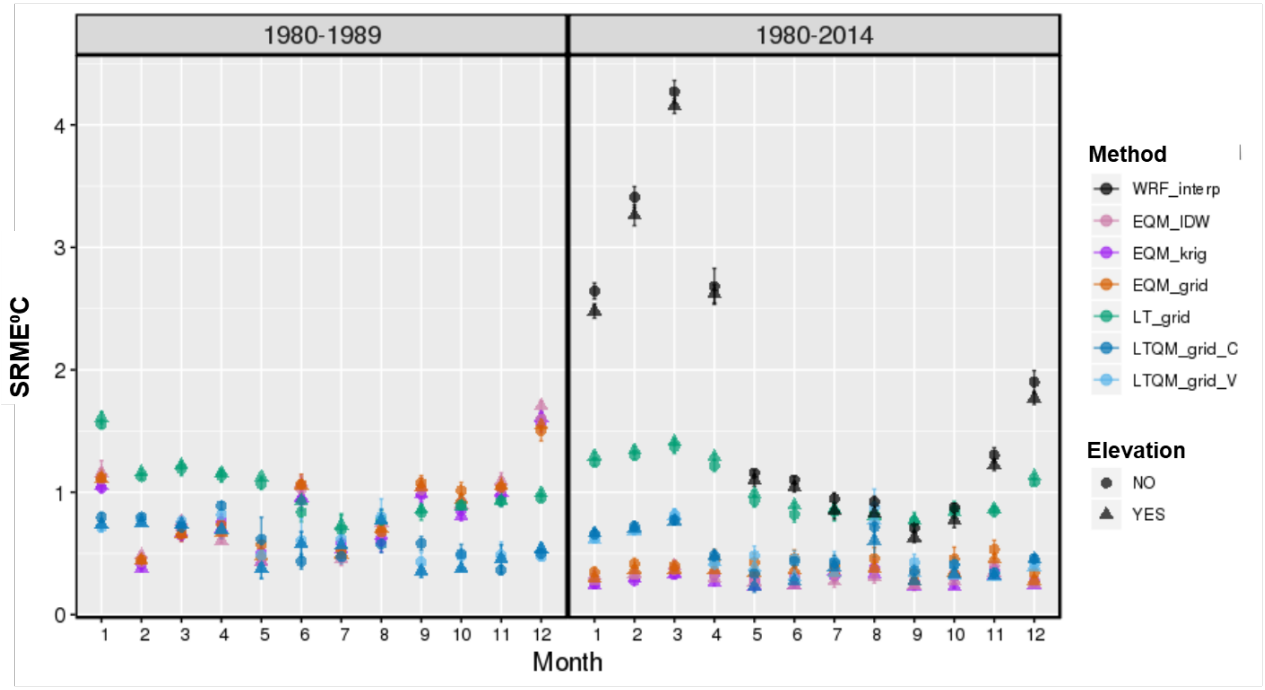

Figure 5.3: Mean SRMSE by Method, Month, *and* Bias_correction_years, *where "1980-1989" and "1980-2014" denote the GHCND station datasets used to bias-correct 1990-2014 and 1980-2014 WRF simulations, respectively. Error bars represent standard errors over five spatial cross-validation folds. "WRF_interp" denotes raw WRF simulations interpolated to station locations and are shown to indicate relative improvement of all methods over raw WRF interpolated values.*

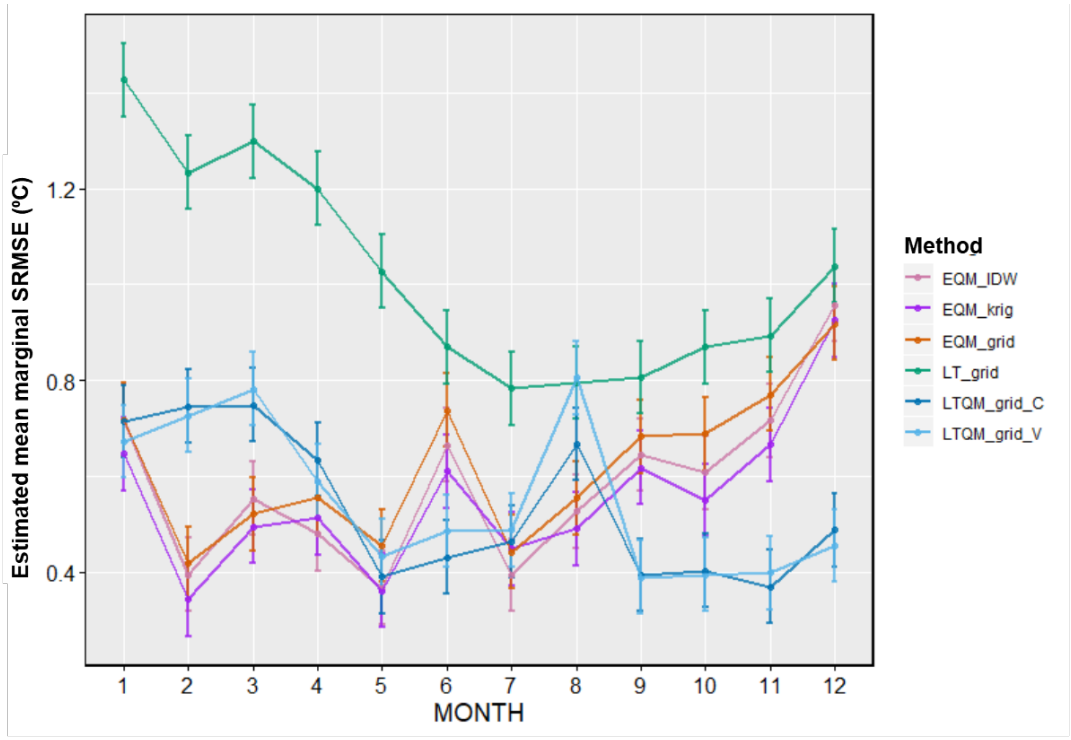

Figure 5.4: Interaction plot showing estimated mean marginal SRMSE for the interaction Method × Month. *Error bars represent 95% confidence intervals.*

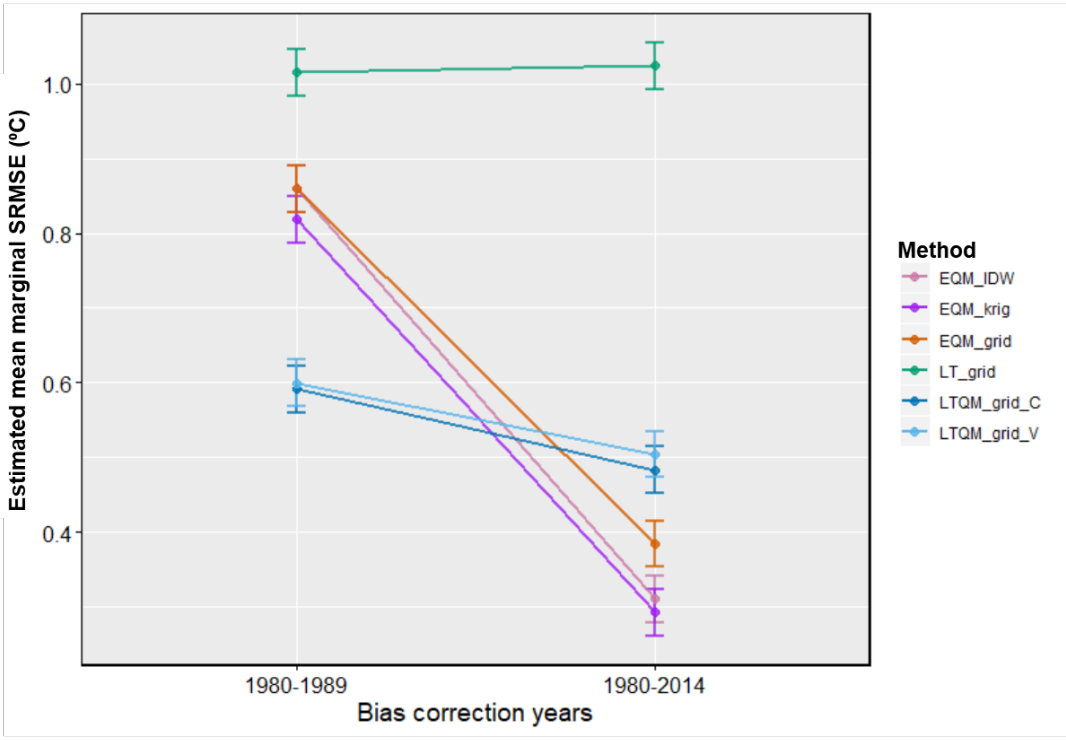

Figure 5.5: Interaction plot showing estimated mean marginal SRMSE for the interaction Method × Bias_correction_years*, where "1980-1989" and "1980-2014" denote the GHCND station datasets used to bias-correct 1990-2014 and 1980-2014 WRF simulations, respectively. Error bars represent 95% confidence intervals. "WRF_interp" denotes raw WRF simulations interpolated to station locations and are shown to indicate relative improvement of all methods over raw WRF interpolated values.*

Table 5.8: ANOVA table for full SRMSE model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Month	11	2.56	0.23	9.65	0.0000
Method	5	8.30	1.66	68.81	0.0000
Bias_correction_years	1	6.09	6.09	252.18	0.0000
Elevation	1	0.03	0.03	1.19	0.2769
Month × Method	55	5.07	0.09	3.82	0.0000
Method × Bias_correction_years	5	3.82	0.76	31.62	0.0000
Method × Elevation	5	0.06	0.01	0.49	0.7867
Bias_correction_years × Elevation	1	0.01	0.01	0.58	0.4490
Method:Bias_correction_years × Elevation	5	0.01	0.00	0.06	0.9980
Residuals	198	4.78	0.02		

References

[12] G. E. Liston and K. Elder, "A meteorological distribution system for high-resolution terrestrial modeling (micromet)," *Journal of Hydrometeorology*, vol. 7, no. 2, pp. 217–234, 2006.

[31] J. M. Winter, B. Beckage, G. Bucini, R. M. Horton, and P. J. Clemins, "Development and evaluation of high-resolution climate simulations over the mountainous northeastern united states," *Journal of Hydrometeorology*, vol. 17, no. 3, pp. 881–896, 2016.

[112] O. Schabenberger and C. A. Gotway, *Statistical methods for spatial data analysis*. Chapman and Hall/CRC, 2017.

[133] R. G. Barry, *Mountain weather and climate*. Psychology Press, 1992.

[136] C. K. Wikle, A. Zammit-Mangion, and N. Cressie, *Spatio-temporal Statistics with R*. CRC Press, 2019.

[209] S. Banerjee, B. P. Carlin, and A. E. Gelfand, *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, 2004.

[210] A. Finley, *Spnngp*, Online, Jul. 2017. [Online]. Available: https://cran.r-project.org/web/packages/spNNGP/spNNGP.pdf.

A2. Chapter 3 supplementary material

Topographic downscaling

Inverse distance weighting (IDW) is a deterministic interpolation technique, so the size of prediction errors cannot be quantified [136]. Interpolated values are based on a weighted average of n nearest-neighbor observations. In IDW, observed values close to prediction locations are assumed to be more influential in the prediction compared to observed values far from prediction locations. As the power, p and the number of nearest neighbors n increases, the smoothness of the interpolated surface increases. IDW is an exact interpolator, which means that if a prediction location, s_0 corresponds to an observed location s_i, the predicted value at s_0 will be identical to the value at location s_i. The general equations for IDW are as follows:

$$
\begin{aligned}
Y(s_0) &= \sum_{i=1}^{n} w_i(s_0) Y(s_i), \\
w_i(s_0) &= \frac{\tilde{w}_i(s_0)}{\sum_{\ell=1}^{n} w_\ell(s_0)}, \\
\tilde{w}_i(s_0) &= \frac{1}{d(s_i, s_0)^p}.
\end{aligned}
$$

The IDW interpolated value at location s_0 is $Y(s_0)$, $d(s_i, s_0)$ is the distance between observed location s_i and prediction location s_0, n is the number of nearest-neighbor observed locations that contribute to the interpolated value $Y(s_0)$, and p is the power parameter.

Topographic downscaling is a variation on IDW that is often used for high- resolution downscaling [31]. Topographic downscaling consists of three main steps:

1. Construction of a historical, empirical lapse rate (relationship between precipitationTMAX and elevation);

2. Adjustment of raw model data to reference elevation (200m) using estimated lapse rate parameters and interpolation of adjusted model data to desired locations using IDW;

3. Back-transformation of interpolated values using estimated lapse rate parameters.

We estimated elevational lapse rates using methods by [31] and [12]. Elevational lapse rates were calculated using historical (1970-1999) GHCND station records. Because precipitation has a nonlinear relationship with elevation [12], we estimated the lapse rate with the nonlinear transformation in (5.10)

$$P_{sta} = P_{ref}\left[\frac{1+\chi(z_{sta}-z_{ref})}{1-\chi(z_{sta}-z_{ref})}\right], \tag{5.10}$$

where P_{sta} is the mean daily precipitation at a GHCND station location, P_{ref} is the reference precipitation (mean daily precipitation at the reference elevation), χ is the elevational lapse rate ($^{\circ}\mathrm{Cm}^{-1}$), and z_{sta} and z_{ref} are the station and reference elevations (m), respectively. The reference elevation was set to the median elevation (200 m), following the methods of [31]. The parameter χ was estimated using nonlinear-weighted least squares with the function `nls` in R. Our result for the elevational lapse rate (0.00025) corresponded to that found by [31].

Next, transformed model data were interpolated to GHNCD station locations using IDW. We used a weight of 2 and 9 nearest neighbors for all IDW interpolation, following methods by [31]. Interpolated daily precipitation values interpolated to the

fine-scale grid were back-transformed using (5.11):

$$P_{fine,\,interp} = P_{ref,\,interp}\left[\frac{1+\chi(z_{fine}-z_{ref})}{1-\chi(z_{fine}-z_{ref})}\right], \tag{5.11}$$

where $P_{fine,\,interp}$, elevation-adjusted high resolution precipitation value, and $P_{ref,\,interp}$ is the interpolated precipitation value at reference elevation.

Bayesian ANOVA details

Table 5.9 shows pairwise comparisons for log-transformed MAE and MAE95 metrics among EQM-, EQM-LIN-corrected, and raw model data. The intercept, β_0 and coefficient β_1 were both given $N(0, 1 \times 10^{-6})$ priors, and a $Unif(0, 100)$ prior was used for the standard deviation, σ. All models were run with 30,000 iterations with a burn-in of 3000 and three Markov-chain-Monte-Carlo (MCMC) chains. Convergence of MCMC chains was assessed ocularly with trace plots. Pairwise comparisons were computed using Tukey's method [211].

Table 5.9: Pairwise comparisons of posterior means of log-transformed MAE and MAE95 metrics among EQM-, EQM-LIN- corrected, and raw model (Mod) data. 'Lower' and 'Upper' denote lower and upper 95% highest posterior density (HPD) intervals, and 'Std Error' refers to the standard error of the difference in means of posterior distributions. Significant comparisons at the 5% significance levels are denoted with (*).

Difference estimate	Estimate	Std Error	Lower	Upper
		MAE95		
EQM - EQM-LIN*	0.926	0.173	0.583	1.26
Mod - EQM-LIN*	1.63	0.172	1.29	1.32
Mod - EQM*	0.702	0.172	0.367	1.04
		MAE		
EQM - EQM-LIN	0.0751	0.110	-0.142	0.290
Mod - EQM-LIN*	0.433	0.110	0.217	0.647
Mod - EQM*	0.359	0.110	0.139	0.573

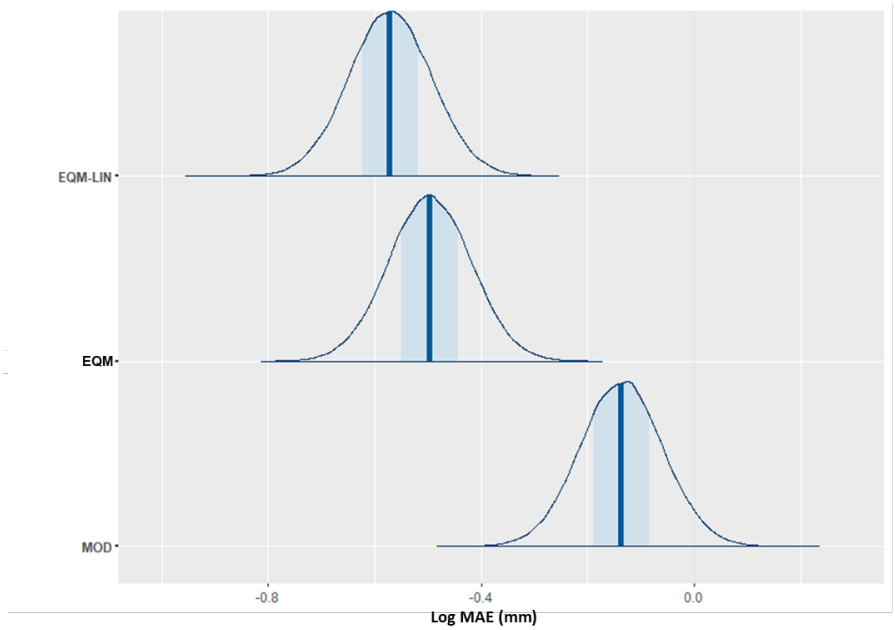

Figure 5.6: Posterior density of log-transformed MAE (mm) for EQM-, EQM-LIN- corrected, and raw model (Mod) data. Plots are computed from posterior draws with all chains merged. Median values are denoted by vertical lines, and shaded areas denote the 50% credible interval around the median.

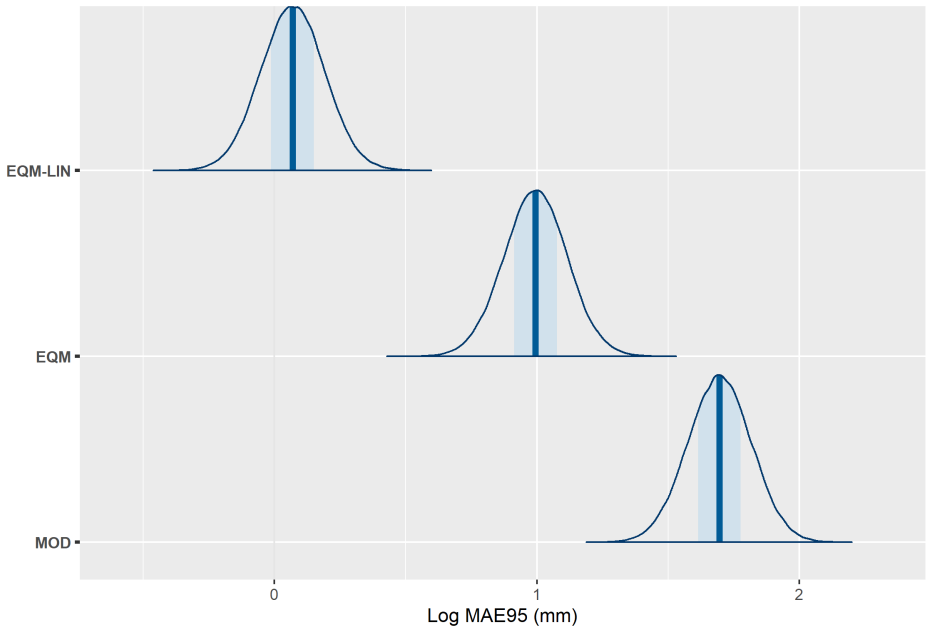

Figure 5.7: Posterior density of log-transformed MAE95 (mm) for EQM-, EQM-LIN- corrected, and raw model (Mod) data. Plots are computed from posterior draws with all chains merged. Median values are denoted by vertical lines, and shaded areas denote the 50% credible interval around the median.

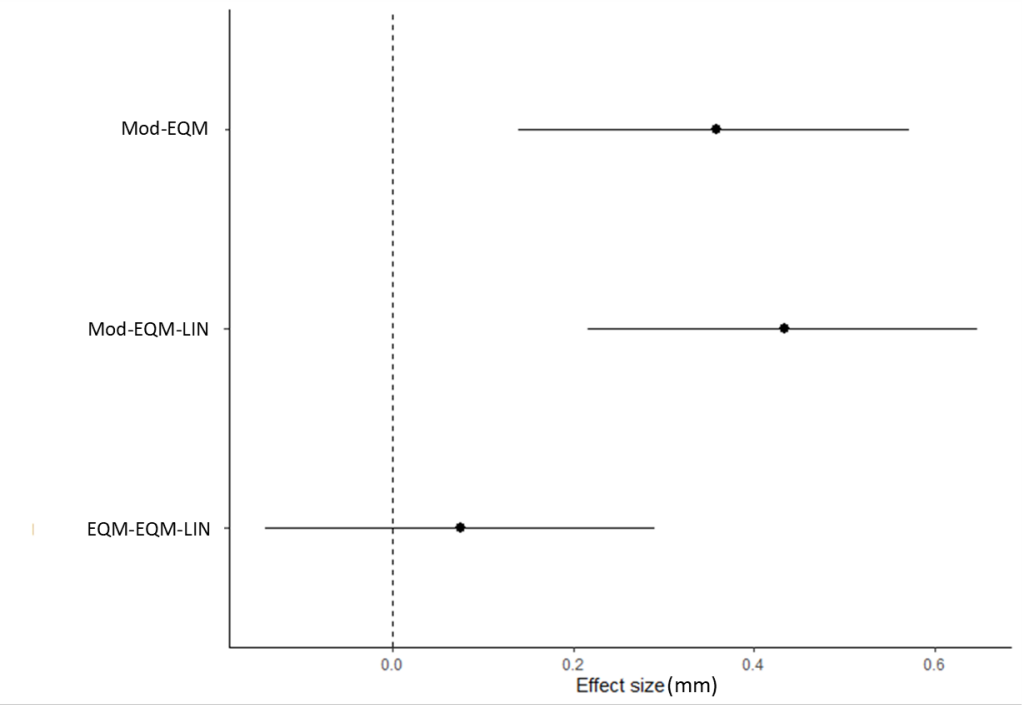

Figure 5.8: Posterior means (black dots) of differences for log-transformed MAE (mm) among EQM-, EQM-LIN- corrected, and raw model (Mod) data. Black lines denote 95% credible intervals. Credible intervals that cross the dotted line are not considered significant at the 5% significance level.

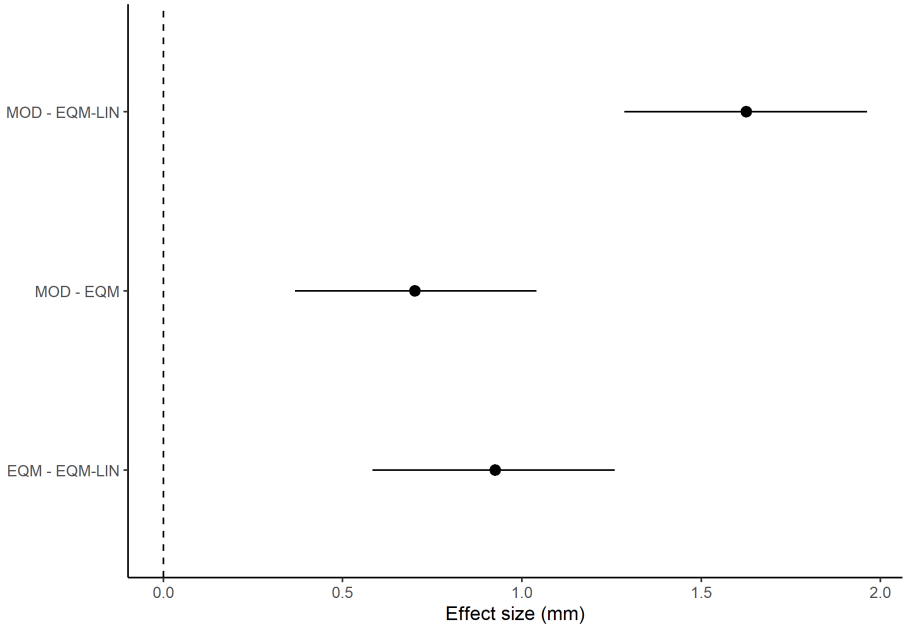

Figure 5.9: Posterior means (black dots) of differences for log-transformed MAE95 (mm) among EQM-, EQM-LIN- corrected, and raw model (Mod) data. Black lines denote 95% credible intervals. Credible intervals that cross the dotted line are not considered significant at the 5% significance level.

Additional bias-correction methods

In addition to EQM-LIN and EQM, we also evaluated distribution mapping (DM) using the gamma distribution (DM-GAMMA) and a hybrid DM approach using EQM for correcting the bulk of model data and the generalized Pareto distribution (GPD) for correcting upper quantiles (EQM-GPD). DM-GAMMA is a common method for bias-correcting model precipitation data [45], [57], [61], and EQM-GPD has been shown to outperform DM-GAMMA in correcting extreme tails [66], [107]. We also included quantile delta mapping (QDM) [47], a trend-preserving method that is less susceptible to overfitting on calibration data and has been shown to effectively correct

precipitation extremes [47]. We refer the reader to [47] for a detailed explanation of QDM. DM-GAMMA was implemented using the `qmap` package [61] in R. EQM-GPD was implemented using a combination of functions from the `qmap` package as well as custom code. Bias-correction via QDM was carried out using the `MBC` package [47] in R.

DM-GAMMA

Bias-correction correction via DM is accomplished through the use of a transfer function (TF) that minimizes the difference between the cumulative distribution functions (CDFs) of model and observed data [53]. The Gamma distribution is frequently used to model wet-day precipitation and is often used for bias-correction via DM [45], [57], [61]. Here, we use the Bernoulli-Gamma distribution, a mixture model. The Bernoulli distribution is models the occurrence of zeros with probability of 1-p, while precipitation values greater than 0 are fit to the Gamma distribution. The probability density function (PDF) of the Bernoulli-Gamma distribution is expressed as:

$$f(x) = \begin{cases} p \cdot \gamma(x) & \text{if } x > 0 \\ 1 - p & \text{if } x \leq 0 \end{cases}$$

The Gamma pdf is specified as

$$\gamma(x|\theta, k) = \frac{e^{-x/\theta} x^{k-1}}{\Gamma(k)\theta^k},$$

where $f(P|\theta, k)$ is the Gamma PDF, P is daily wet-day precipitation, k and θ are the shape and scale parameters, respectively, and Γ is the Gamma function [58]. The

corrected daily precipitation value, $x_{corr,t}$ can be expressed via the TF as:

$$x_{corr,t} = F_{obs}^{-1}(F_{mod}(x_{mod,t}|\theta_{mod}, k_{mod})|\theta_{obs}, k_{obs}), \tag{5.12}$$

where $x_{corr,t}$ is the corrected daily precipitation value, F_{obs}^{-1} is Gamma-Bernoulli quantile function for observed data, F_{mod}^{-1} denotes the Gamma-Bernoulli cumulative distribution function (CDF) for observed data; $x_{mod,t}$ is the model precipitation value on day t, and θ_{mod}, k_{mod}, θ_{obs}, k_{obs} are estimated shape and rate parameters of fitted Gamma CDFs for model and observed data, respectively. Rate and scale parameters were estimated using the method of maximum likelihood. Observed and model precipitation distributions were estimated using 10,000 quantiles.

EQM-GPD

EQM-GPD is a hybrid approach in which daily model precipitation below a specified threshold were corrected using EQM, while TFs for precipitation values above the threshold were constructed by fitting observed and model data to GPD distributions [106]. The GPD distribution function with shape and scale parameters ξ and σ is given by (5.13)

$$f(y) = \begin{cases} \frac{1}{\sigma}(1 - \xi y/\sigma)^{\frac{1}{\xi}-1}, \ \xi \neq 0, \sigma > 0 \\ \frac{1}{\sigma}\exp(-y/\sigma), \ \xi = 0, \sigma > 0 \end{cases}, \tag{5.13}$$

where $0 \leq y \leq \infty$ if $\xi \geq 0$ and $0 \leq y \leq -\sigma/\xi$ if $\xi < 0$ [111]. The TF used to correct daily model precipitation is:

$$P_{corr,t} = \begin{cases} \text{ecdf}_{obs}^{-1}(\text{ecdf}_{mod}(P_{mod,t})), & X_{mod,t} < \text{ecdf}_{mod}^{-1}(\tau_{GPD}) \\ F_{obs}^{-1}(F_{mod}(P_{mod,t}|\xi_{mod}, \sigma_{mod})|\xi_{obs}, \sigma_{obs}), & P_{mod,t} > \text{ecdf}_{mod}^{-1}(\tau_{GPD}). \end{cases} \tag{5.14}$$

In (5.14), $P_{corr,t}$ is the corrected daily precipitation value, F_P and F_P^{-1} denote the CDF and quantile function of the GPD, respectively, $P_{mod,t}$ is the model precipitation value on day t, θ and k are estimated shape and rate parameters, and $0 < \tau_{GPD} < 1$. We chose the threshold to be the 95th quantile of observed or model precipitation, because we found that values above the 95th quantiles of both model and observed data could be adequately fitted to GPDs. Use the of 95th quantile agrees with previous studies [66], [107] in which GPDs were used for bias-correction approaches. Various approaches have been used to select the threshold; [65], [66], and [107] used the 95th quantile, [212] used the 95th or 99th quantile, [101] estimated the threshold using a probability weighted moments estimator, and [213] inspected mean-excess plots. Thus, we define $\tau_{GPD} = 0.95$ as the value at which ecdf_{obs}^{-1} and ecdf_{mod}^{-1} are evaluated. To clarify, although $\tau_{GPD} = 0.95$, the actual quantile values (in mm precipitation) varied for each month and data type (observed and model) as $\text{ecdf}_{obs}^{-1}(\tau_{GPD})$ is not necessarily equal to $\text{ecdf}_{mod}^{-1}(\tau_{GPD})$ for a given month. GPDs were fit to observed and model data using maximum likelihood (ML) with the R package `ExtRemes` [214].

Results for additional bias-correction methods

MAE and MAE95

Bayesian ANOVAs were carried out to compared mean MAE and MAE95 among all bias-correction methods. The analyses were conducted as described for the main manuscript results.

MAE For mean MAE, QDM performed best, followed by EQM-LIN (although results for QDM were not significantly different from those for EQM-LIN). DM-GAMMA performed similarly to EQM, and EQM-GPD provided only a marginal improvement over raw model data (Figure 5.10 a). (Table 5.10 shows all pairwise differences; Figures 5.11, 5.12, 5.13 show posterior distributions, differences in posterior distributions, and effect sizes, respectively for the MAE ANOVA analysis).

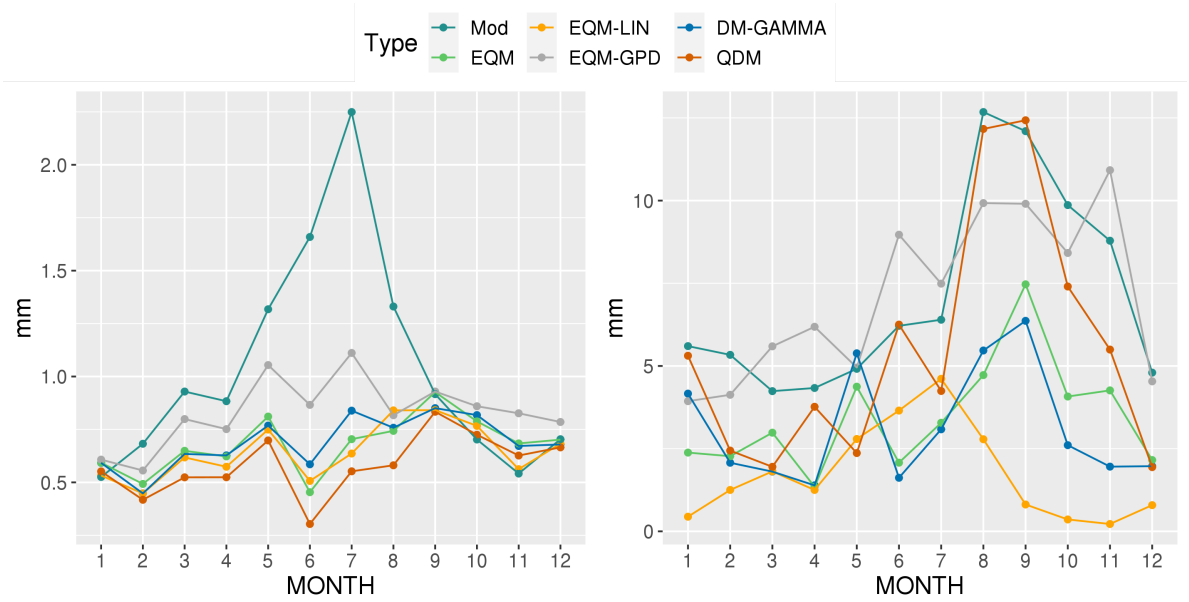

Figure 5.10: Monthly mean MAE (mm) (a) and MAE95 (mm) (b) for raw model (Mod), EQM-, EQM-LIN-, EQM-GPD-, DM-GAMMA, and QDM-corrected data. Please note the difference in y-axes limits for plots a and b.

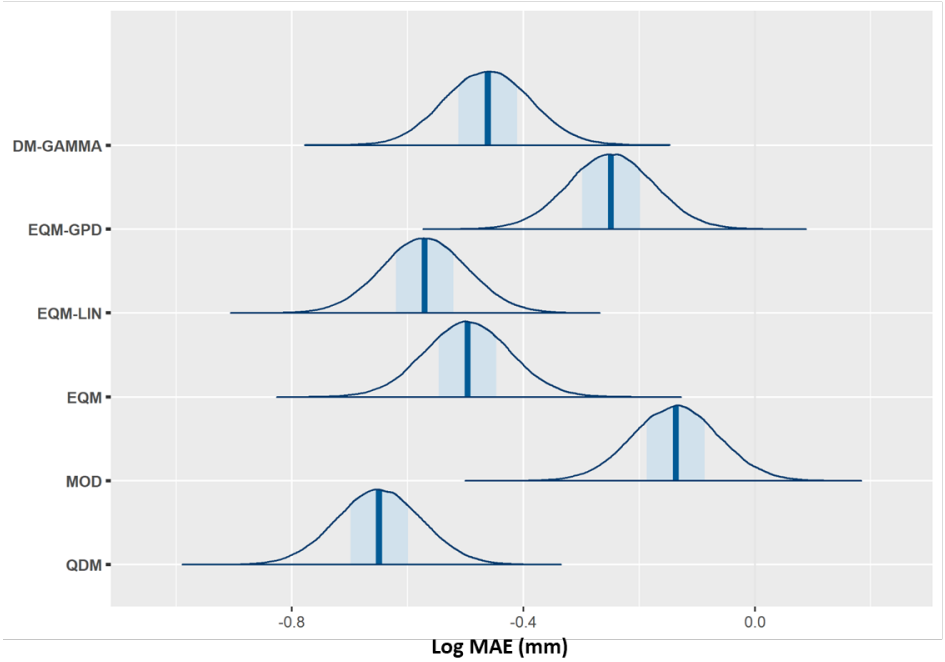

Figure 5.11: Posterior densities of log-transformed MAE (mm) for EQM-, EQM-LIN-, EQM-GPD, DM-GAMMA-, and QDM-corrected, and raw model (Mod) data. Plots are computed from posterior draws with all chains merged. Median values are denoted by vertical lines, and shaded areas denote the 50% credible interval around the median.

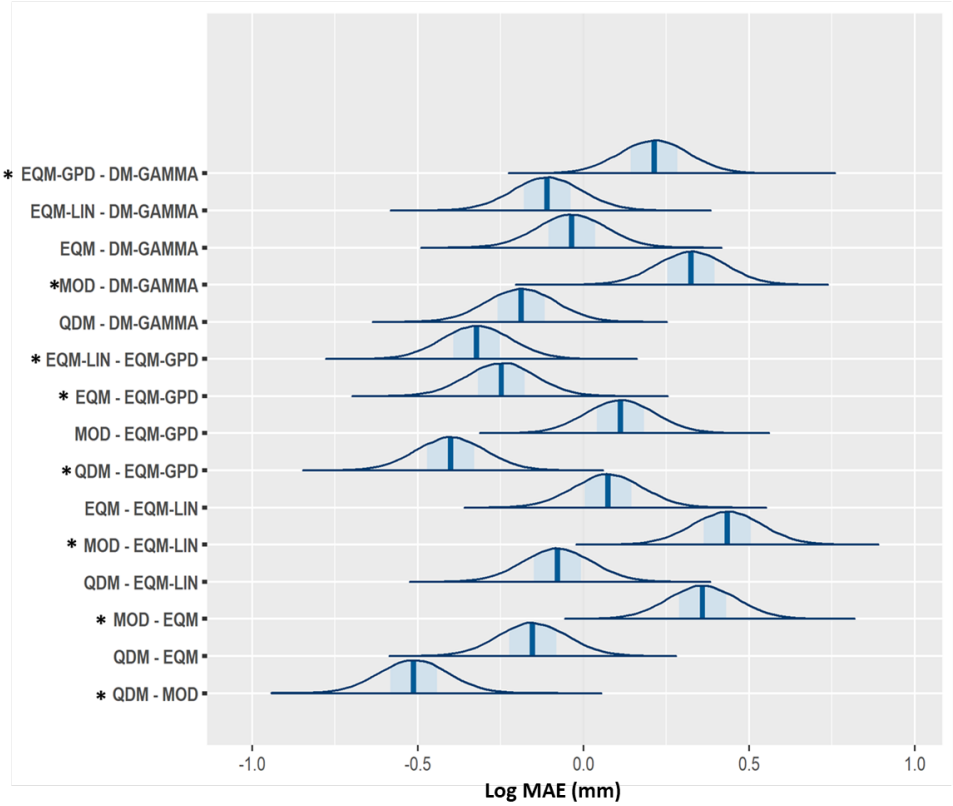

Figure 5.12: Posterior densities of pairwise differences of log-transformed MAE (mm) among EQM-, EQM-LIN-, EQM-GPD, DM-GAMMA-, and QDM-corrected, and raw model (Mod) data. Plots are computed from posterior draws with all chains merged. Median values are denoted by vertical lines, and shaded areas denote the 50% credible interval around the median. Differences that are signficant are denoted with (*).

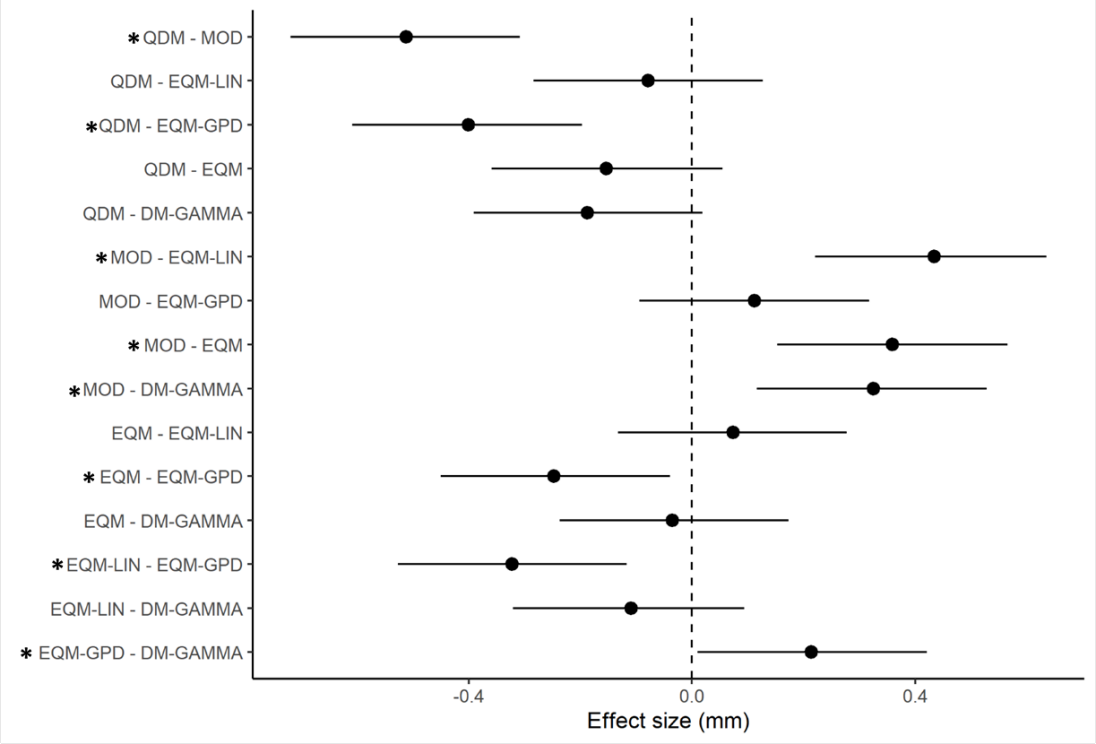

Figure 5.13: Posterior means (black dots) of differences for log-transformed MAE95 (mm) among EQM-, EQM-LIN-, EQM-GPD, DM-GAMMA-, and QDM-corrected, and raw model (Mod) data. Black lines denote 95% credible intervals. HPD Credible intervals that cross the dotted line are not considered significant at the 5% significance level; intervals that do not include 0 are denoted with (*).

Table 5.10: Pairwise comparisons of posterior means of log-transformed MAE among EQM-, EQM-LIN- corrected, raw model (Mod), DM-GAMMA-, EQM-GPD-, and QDM-corrected data. 'Lower' and 'Upper' denote lower and upper 95% highest posterior density (HPD) intervals, and 'Std Error' refers to the standard error of the difference in means of posterior distributions. HPD intervals that do not include 0 are denoted with (*).

Comparison	Difference estimate	Std Error	Lower	Upper
EQM-GPD - DM-GAMMA*	0.21	0.11	0.01	0.42
EQM-LIN - DM-GAMMA	-0.11	0.11	-0.32	0.09
EQM - DM-GAMMA	-0.04	0.10	-0.24	0.17
MOD - DM-GAMMA*	0.33	0.11	0.12	0.53
QDM - DM-GAMMA	-0.19	0.11	-0.39	0.02
EQM-LIN - EQM-GPD*	-0.32	0.10	-0.53	-0.12
EQM - EQM-GPD	-0.25	0.11	-0.45	-0.04
MOD - EQM-GPD	0.11	0.11	-0.09	0.32
QDM - EQM-GPD*	-0.40	0.11	-0.61	-0.20
EQM - EQM-LIN	0.07	0.10	-0.13	0.28
MOD - EQM-LIN*	0.43	0.11	0.22	0.64
QDM - EQM-LIN	-0.08	0.11	-0.28	0.13
MOD - EQM*	0.36	0.11	0.15	0.57
QDM - EQM	-0.15	0.11	-0.36	0.05
QDM - MOD*	-0.51	0.11	-0.72	-0.31

MAE95 For mean MAE95, EQM-LIN performed significantly better than any other method. Like results for MAE, DM-GAMMA performed similarly to EQM. EQM-GPD performed worse and did not provide improvement over raw model data and, in fact, increased MAE95 values for months 2,3,4,6,7 and 11 (Figure 5.10 b). While QDM provided an overall significant improvement in MAE95 over raw model data, the effect size was small (Figure 5.16). QDM provided no improvement over raw model data during months in which extreme precipitation is already substantially overestimated in raw model data (months 8, and 9) (Figure 5.10 b). (Table 5.11 shows all pairwise differences; Figures 5.14, 5.15, 5.16 show posterior distributions, differences in posterior distributions, and effect sizes, respectively, for the MAE95 ANOVA analysis).

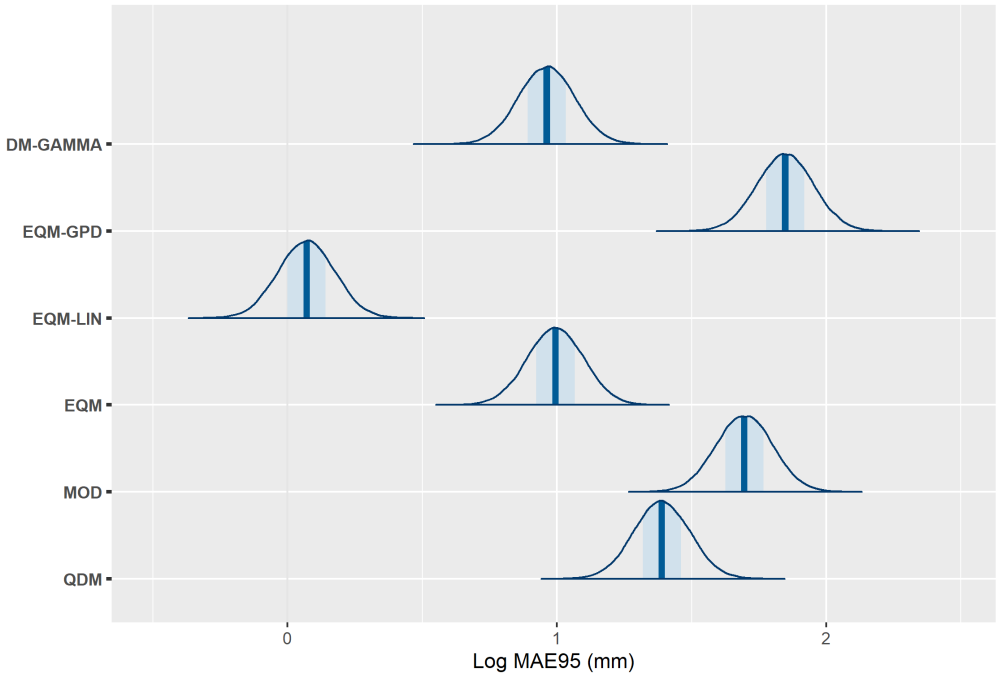

Figure 5.14: Posterior densities of log-transformed MAE95 (mm) for EQM-, EQM-LIN-, EQM-GPD, DM-GAMMA-, and QDM-corrected, and raw model (Mod) data. Plots are computed from posterior draws with all chains merged. Median values are denoted by vertical lines, and shaded areas denote the 50% credible interval around the median. Differences that are signficant are denoted with (*).

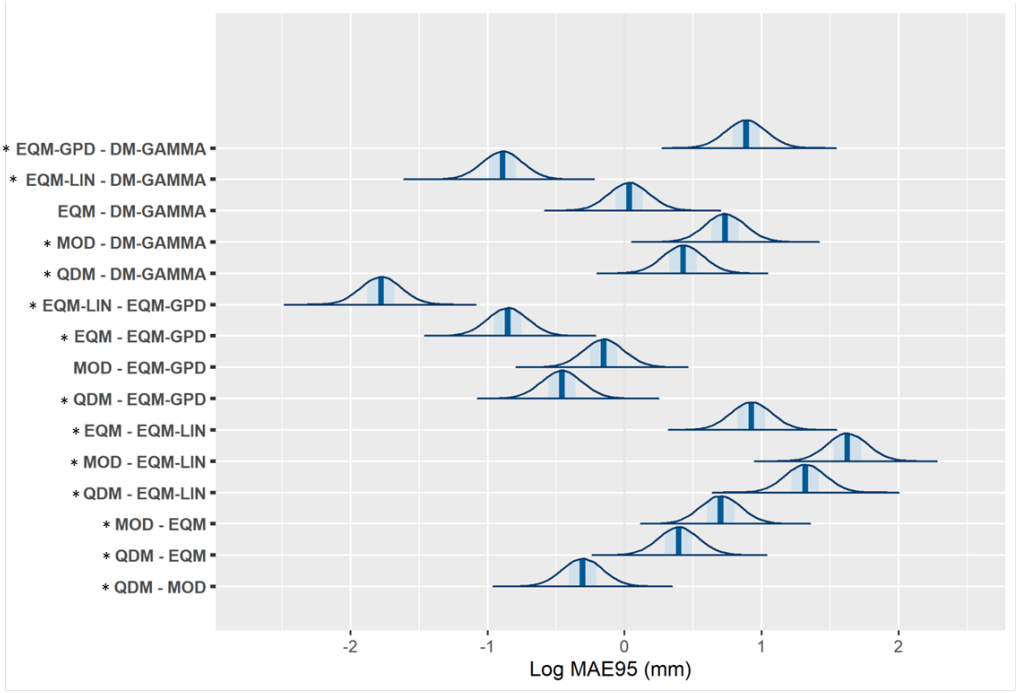

Figure 5.15: Posterior densities of pairwise differences of log-transformed MAE95 (mm) among EQM-, EQM-LIN-, EQM-GPD, DM-GAMMA-, and QDM-corrected, and raw model (Mod) data. Plots are computed from posterior draws with all chains merged. Median values are denoted by vertical lines, and shaded areas denote the 50% credible interval around the median.

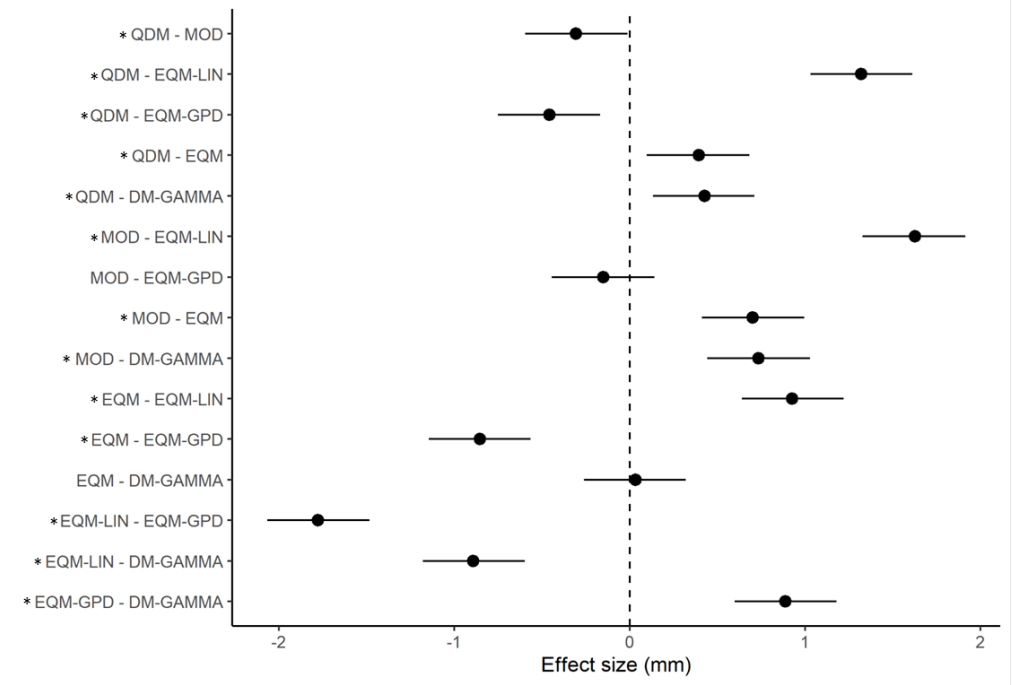

Figure 5.16: Posterior means (black dots) of differences for log-transformed MAE95 (mm) among EQM-, EQM-LIN-, EQM-GPD, DM-GAMMA-, and QDM-corrected, and raw model (Mod) data. Black lines denote 95% credible intervals. HPD credible intervals that cross the dotted line are not considered significant at the 5% significance level; intervals that do not include 0 are denoted with (*).

Table 5.11: Pairwise comparisons of posterior means of log-transformed MAE95 among EQM-, EQM-LIN- corrected, raw model (Mod), DM-GAMMA-, EQM-GPD-, and QDM-corrected data. 'Lower' and 'Upper' denote lower and upper 95% highest posterior density (HPD) intervals, and 'Std Error' refers to the standard error of the difference in means of posterior distributions. HPD intervals that do not include 0 are denoted with (*).

Comparison	Difference estimate	Std error	Lower	Upper
EQM-GPD - DM-GAMMA*	0.89	0.15	0.60	1.18
EQM-LIN - DM-GAMMA*	-0.89	0.15	-1.18	-0.60
EQM - DM-GAMMA	0.03	0.15	-0.26	0.32
MOD - DM-GAMMA*	0.73	0.15	0.44	1.03
QDM - DM-GAMMA*	0.43	0.15	0.13	0.71
EQM-LIN - EQM-GPD*	-1.78	0.15	-2.07	-1.48
EQM - EQM-GPD*	-0.85	0.15	-1.15	-0.56
MOD - EQM-GPD	-0.15	0.15	-0.45	0.14
QDM - EQM-GPD*	-0.46	0.15	-0.75	-0.17
EQM - EQM-LIN*	0.92	0.15	0.64	1.22
MOD - EQM-LIN*	1.62	0.15	1.33	1.91
QDM - EQM-LIN*	1.32	0.15	1.03	1.61
MOD - EQM*	0.70	0.15	0.41	0.99
QDM - EQM*	0.39	0.15	0.10	0.68
QDM - MOD*	-0.31	0.15	-0.60	-0.01

ETCCDI indices

QDM and DM-GAMMA performed similarly to EQM and EQM-LIN with respect to 'D' and 'S' indices (Figures 5.17, 5.18). QDM and DM-GAMMA resulted in larger underestimations of WetDays compared to EQM and EQM-LIN; however, distributions of TotalP calculated from DM-GAMMA- and QDM-corrected-data were similar to those of observed data (Figure 5.19). SPI was overestimated by raw model and all bias-corrected data. EQM-GPD performed worst overall, and bias-correction via EQM-GPD adversely affected results for S90 and S95 compared to raw model data (e.g. distributional discrepancies of S90 and S95 compared to observed data were increased) (Figure 5.18). EQM-GPD provided some improvement over raw model

data for TotalP, but the distribution of TotalP was still significantly different from that of observed data (Figure 5.19). Overall, EQM-LIN and EQM resulted in larger improvements to ETCCDI indices than any other bias-correction method tested in this study.

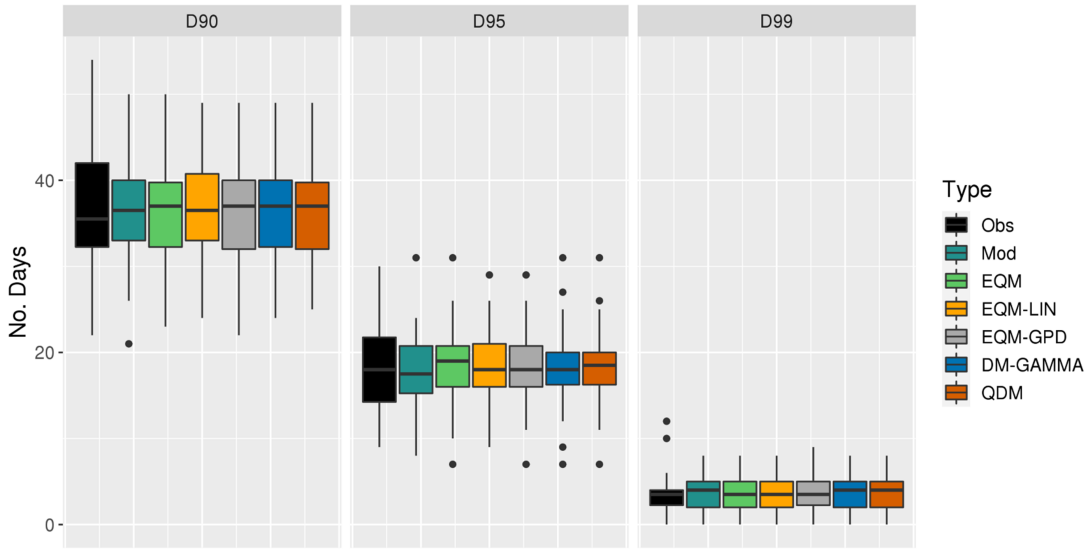

Figure 5.17: Boxplots of a) D90, D95, and D99 for observed (Obs), model (Mod), EQM-, EQM-LIN-, EQM-GPD-, DM-GAMMA, and QDM-corrected data. Boxplots reflect 30 annual values for each data type and ETCCDI index. Significance of KS-tests of distributional similarity of Mod, EQM, EQM-LIN, EQM-GPD, DM-GAMMA, or QDM, compared to Obs at $\alpha = 0.05$*, adjusted with the Holm-Bonferroni method, are indicated with (*); dots denote outliers.*

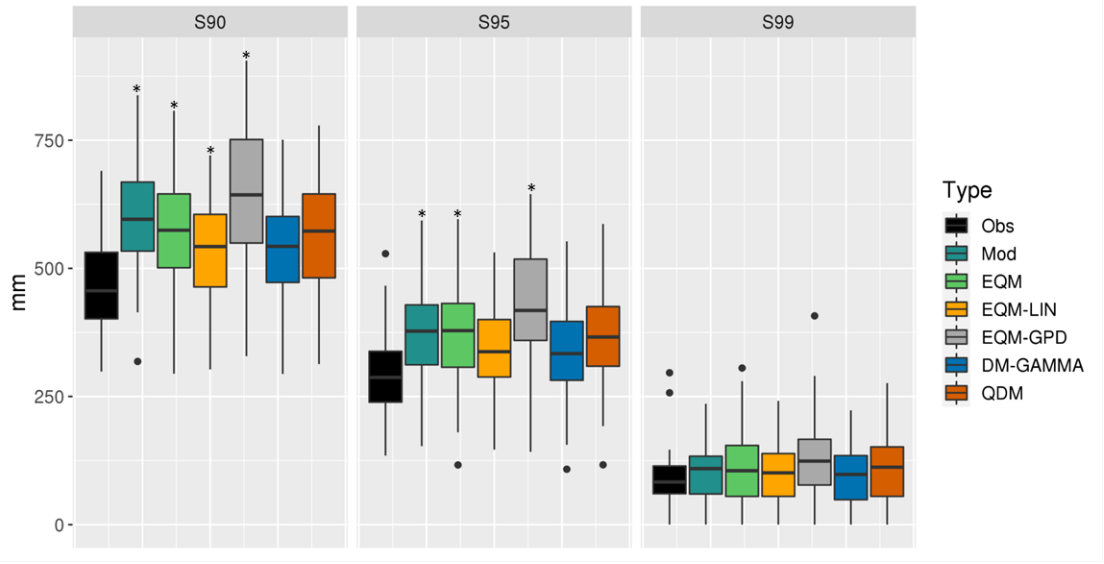

Figure 5.18: D90, D95, and D99 for observed (Obs), model (Mod), EQM-, EQM-LIN-, EQM-GPD-, DM-GAMMA, and QDM-corrected data. Boxplots reflect 30 annual values for each data type and ETCCDI index. Significance of KS-tests of distributional similarity of Mod, EQM, EQM-LIN, EQM-GPD, DM-GAMMA, or QDM, compared to Obs at $\alpha = 0.05$, adjusted with the Holm-Bonferroni method, are indicated with (*); dots denote outliers.

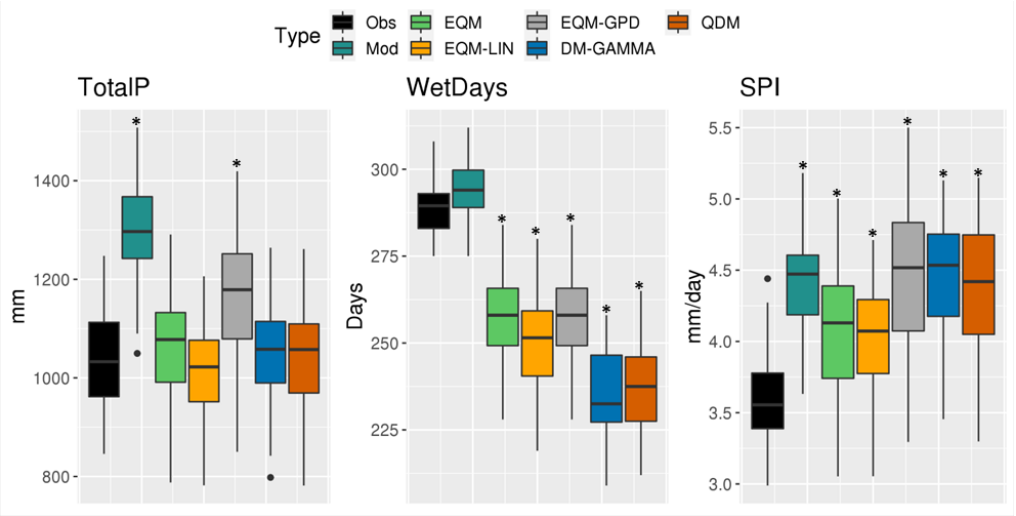

Figure 5.19: Boxplots of TotalP, WetDays, and SPI for observed (Obs), model (Mod), EQM-, EQM-LIN-, EQM-GPD-, DM-GAMMA, and QDM-corrected data. Boxplots reflect 30 annual values for each data type and ETCCDI index. Significance of KS-tests of distributional similarity of Mod, EQM, EQM-LIN, EQM-GPD, DM-GAMMA, or QDM, compared to Obs at $\alpha = 0.05$, adjusted with the Holm-Bonferroni method, are indicated with (*); dots denote outliers.

A visual examination of WRF wet bias with respect to observed data

The excess of low-precipitation days in raw model data can be visualized with histograms. Histograms in Figures 5.20 and 5.21 show the percent contribution to total monthly precipitation within 2mm bin widths for the entire distribution of raw model and observed data by month. In nearly all months, the contribution of low-precipitation amounts to total precipitation is greater for raw model compared to observed data. The difference is most striking for summer and early fall months.

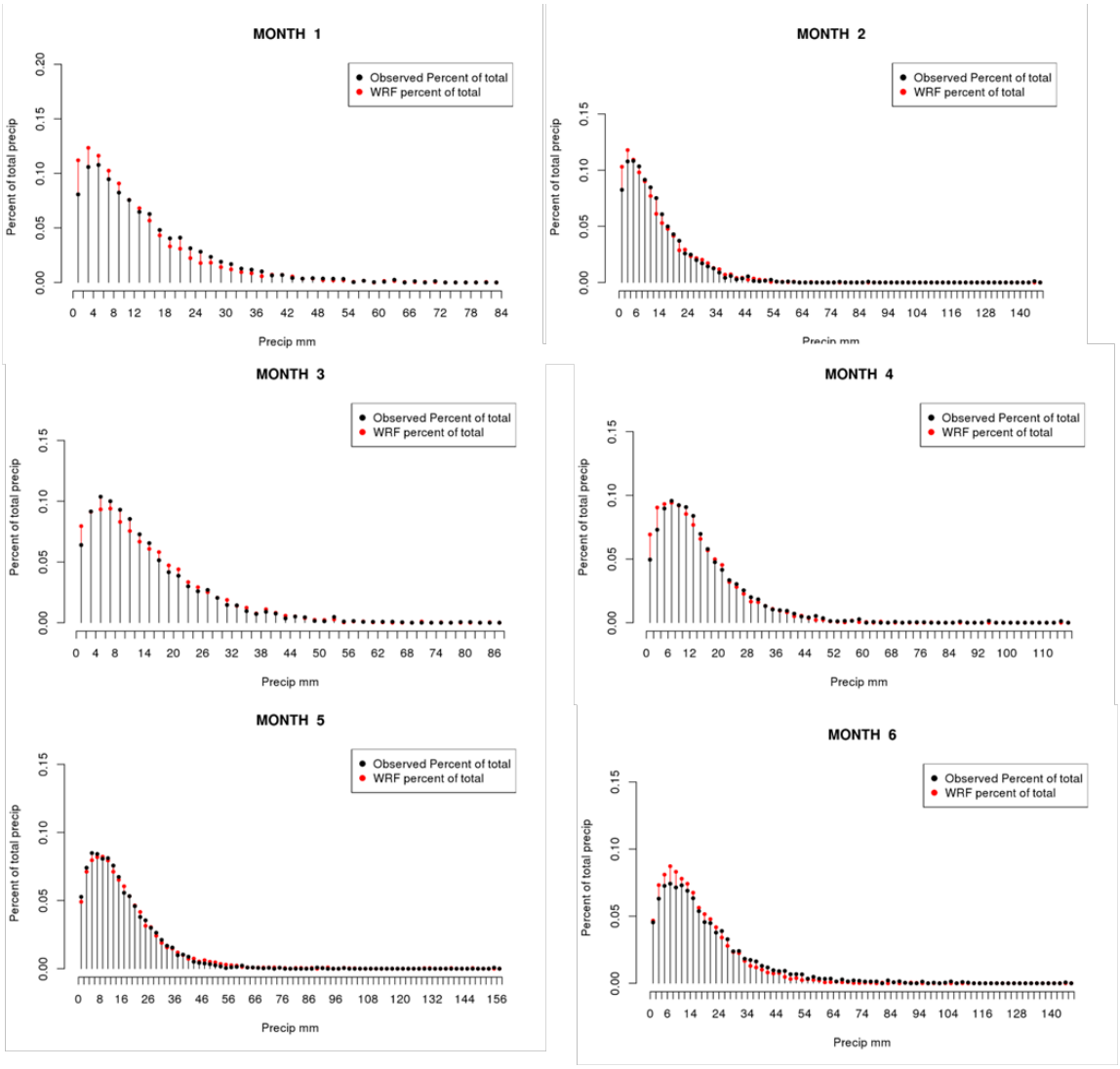

Figure 5.20: Proportion contribution to total daily precipitation by 2mm bin widths for months 1-6 over the calibration period 1976-2005. Raw model data are denoted by black lines, and observed data are denoted by red lines.

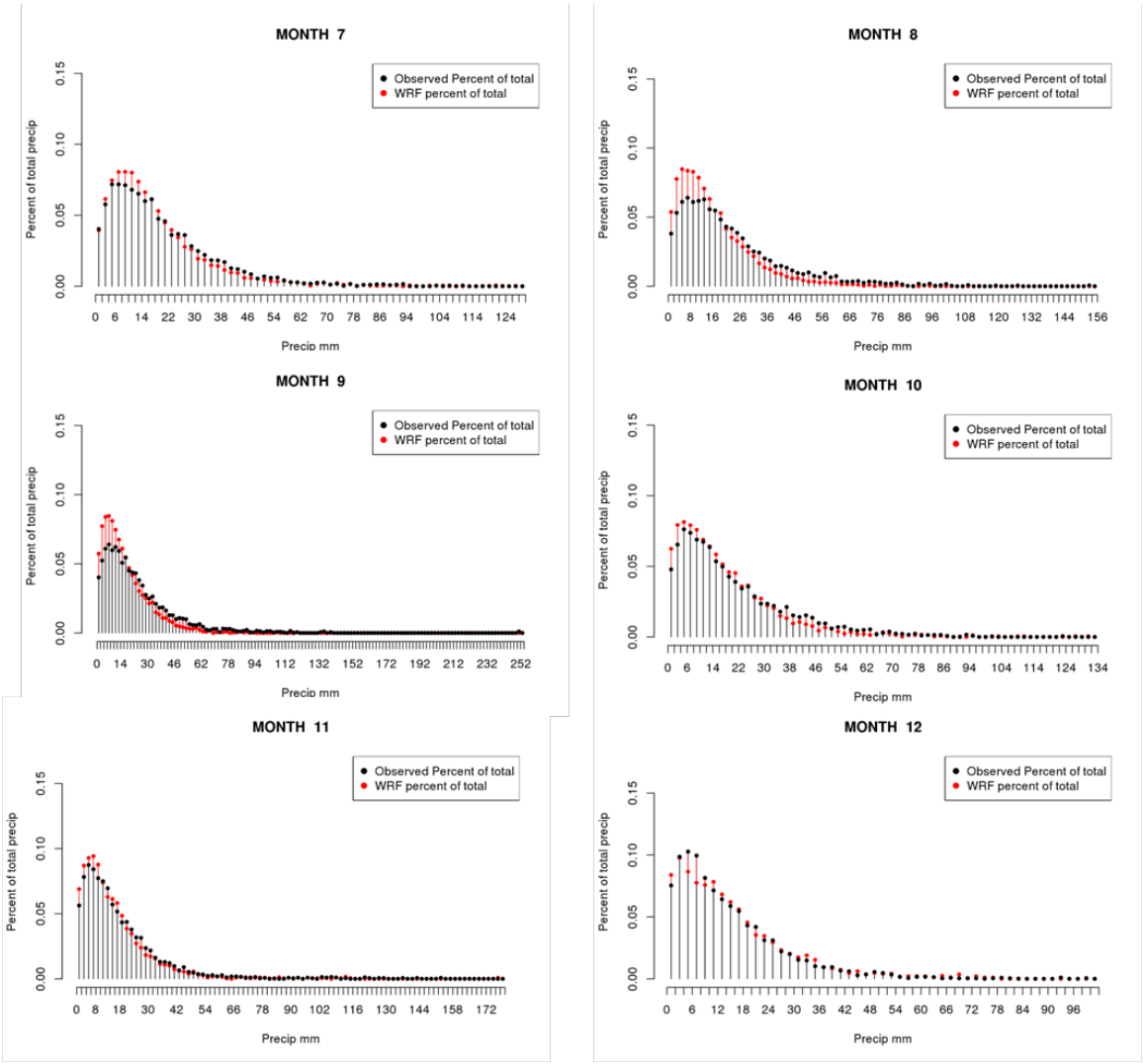

Figure 5.21: Proportion contribution to total daily precipitation by 2mm bin widths for months 7-12 over the calibration period 1976-2005. Raw model data are denoted by black lines, and observed data are denoted by red lines.

WRF physics settings

Table 5.12: Physics settings and details for the WRF model

Setting	Details
Microphysics	WRF Single–moment 6–class Scheme (Hong, S.–Y., and J.–O. J. Lim, 2006: The WRF single–moment 6–class microphysics scheme (WSM6). J. Korean Meteor. Soc., 42, 129–151.)
Radiation	RRTMG Shortwave and Longwave Schemes (Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long–lived greenhouse gases: Calculations with the AER radiative transfer models. J. Geophys. Res., 113, D13103. doi:10.1029/2008JD009944)
Boundary layer	Mellor–Yamada–Janjic Scheme (MYJ) (Janjic, Zavisa I., 1994: The Step–Mountain Eta Coordinate Model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. Mon. Wea. Rev., 122, 927–945. doi:10.1175/1520-0493(1994)122%3c0927:TSMECM%3e2.0.CO;2)
Cumulus convection	New Simplified Arakawa–Schubert Scheme (for Basic WRF) (Han, Jongil and Hua–Lu Pan, 2011: Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. Wea. Forecasting, 26, 520–533. doi:10.1175/WAF-D-10-05038.1)
Land surface physics	Unified Noah Land Surface Model (Tewari, M., F. Chen, W. Wang, J. Dudhia, M. A. LeMone, K. Mitchell, M. Ek, G. Gayno, J. Wegiel, and R. H. Cuenca, 2004: Implementation and verification of the unified NOAH land surface model in the WRF model. 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, pp. 11–15.)
Surface layer physics	Eta Similarity Scheme (Janjic, Z. I., 1994: The step-mountain Eta coordinate model: further developments of the convection, viscous sublayer and turbulence closure schemes. Mon. Wea. Rev., 122, 927–945. doi:10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2)

References

[12] G. E. Liston and K. Elder, "A meteorological distribution system for high-resolution terrestrial modeling (micromet)," *Journal of Hydrometeorology*, vol. 7, no. 2, pp. 217–234, 2006.

[31] J. M. Winter, B. Beckage, G. Bucini, R. M. Horton, and P. J. Clemins, "Development and evaluation of high-resolution climate simulations over the mountainous northeastern united states," *Journal of Hydrometeorology*, vol. 17, no. 3, pp. 881–896, 2016.

[45] C. Piani, J. Haerter, and E. Coppola, "Statistical bias correction for daily precipitation in regional climate models over europe," *Theoretical and Applied Climatology*, vol. 99, no. 1-2, pp. 187–192, 2010.

[47] A. J. Cannon, S. R. Sobie, and T. Q. Murdock, "Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?" *Journal of Climate*, vol. 28, no. 17, pp. 6938–6959, 2015.

[53] M. Jakob Themeßl, A. Gobiet, and A. Leuprecht, "Empirical-statistical downscaling and error correction of daily precipitation from regional climate models," *International Journal of Climatology*, vol. 31, no. 10, pp. 1530–1544, 2011.

[57] R. W. Katz, "Extreme value theory for precipitation: Sensitivity analysis for climate change," *Advances in Water Resources*, vol. 23, no. 2, pp. 133–139, 1999.

[58] D. S. Wilks, *Statistical methods in the atmospheric sciences*. Academic press, 2011, vol. 100.

[61] L. Gudmundsson, J. Bremnes, J. Haugen, and T. Engen-Skaugen, "Downscaling rcm precipitation to the station scale using statistical transformations–a comparison of methods," *Hydrology and Earth System Sciences*, vol. 16, no. 9, pp. 3383–3390, 2012.

[65] P. Reiter, O. Gutjahr, L. Schefczyk, G. Heinemann, and M. Casper, "Does applying quantile mapping to subsamples improve the bias correction of daily precipitation?" *International Journal of Climatology*, vol. 38, no. 4, pp. 1623–1633, 2018.

[66] O. Gutjahr and G. Heinemann, "Comparing precipitation bias correction methods for high-resolution regional climate simulations using cosmo-clm," *Theoretical and applied climatology*, vol. 114, no. 3, pp. 511–529, 2013.

[101] A. Mamalakis, A. Langousis, R. Deidda, and M. Marrocu, "A parametric approach for simultaneous bias correction and high-resolution downscaling of climate model rainfall," *Water Resources Research*, vol. 53, no. 3, pp. 2149–2170, 2017.

[106] S. Tani and A. Gobiet, "Quantile mapping for improving precipitation extremes from regional climate models," *Journal of Agrometeorology*, vol. 21, no. 4, pp. 434–443, 2019.

[107] W. Yang, J. Andréasson, L. Phil Graham, J. Olsson, J. Rosberg, and F. Wetterhall, "Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies," *Hydrology Research*, vol. 41, no. 3-4, pp. 211–229, 2010.

[111] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels, *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.

[136] C. K. Wikle, A. Zammit-Mangion, and N. Cressie, *Spatio-temporal Statistics with R*. CRC Press, 2019.

[211] J. W. Tukey, "Comparing individual means in the analysis of variance," *Biometrics*, pp. 99–114, 1949.

[212] D.-I. Kim, H.-H. Kwon, and D. Han, "Bias correction of daily precipitation over south korea from the long-term reanalysis using a composite gamma-pareto distribution approach," *Hydrology Research*, vol. 50, no. 4, pp. 1138–1161, 2019.

[213] Y. Li, W. Cai, and E. Campbell, "Statistical modeling of extreme rainfall in southwest western australia," *Journal of climate*, vol. 18, no. 6, pp. 852–863, 2005.

[214] E. Gilleland and R. W. Katz, "Extremes 2.0: An extreme value analysis package in r," *Journal of Statistical Software*, vol. 72, no. 1, pp. 1–39, 2016.

A3. Chapter 4 supplementary material

List of additional tables and figures

1. Figure S1. The input data to the HetGP model [199] and modified time covariate to enforce periodicity for fitted seasonal means and standard deviations.

2. Figure S2. Elevational lapse rate adjustment for TMAX.

3. Figure S3. Cross-validated results of MAE(°C) for uncorrected model (MOD), and EQM- and DMTA-corrected data (without temporal adjustment) relative to observed data during the historical period (1976-2005) over day of year (DOY).

4. Figure S4. EQM transfer function (TF) for month 12.

5. Figure S5. Boxplots showing the distribution of daily, spatially-explicit, uncorrected, bias-corrected model, and observed data (°C) for month 12 during the historical time period (1976-2005).

6. Figure S6. Boxplots showing the distribution of daily, spatially-explicit, uncorrected and bias-corrected model data (°C) for month 12 during a future time period (2090-2099). Data at the 78 GHCND locations are represented in this plot.

7. Table S1. Mean daily TMAX (°C) for observed and model data by month during 1976-2005.

8. Table S2. Monthly standard deviations (SDs) (°C) for observed and model data by month during 1976-2005.

HetGP details: enforcing periodicity

The input data to the HetGP model [199] was manipulated such that resulting model predictions of the seasonal means and standard deviations for were periodic. The spatially averaged time series $T_{(\cdot)}$ was input three times, and an additional 50 days were added on at the beginning and end of the input vector. The time covariate was modified as shown in Figure S1. Predictions were only made over the middle portion of the input vector (timestep = 366 to 730), and these predictions were periodic.

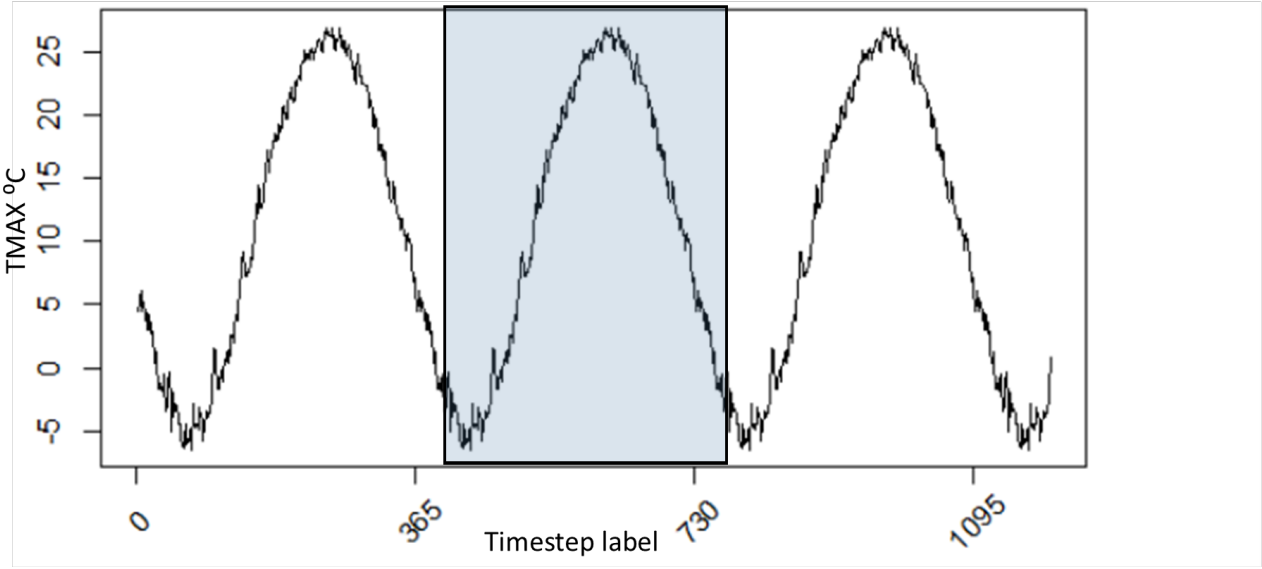

Figure S1. The input data to the HetGP model [199] and modified time covariate to enforce periodicity.

Implementing EQM with temporal adjustment via a process convolution approach

To implement EQM *with* temporal adjustment, data processing and temporal adjustment steps in sections 4.3.3-4.3.5 were followed. The resulting temporally-adjusted time series, $\tilde{\mathbf{W}}_{Mod}$, was back-transformed, except only estimate trends derived from *model* data (rather than observed data) were used. Let $\tilde{\mathbf{T}}^*_{Mod}$ be the resulting back-transformed, (spatially averaged) time series; then, spatially-explicit model data were adjusted as follows:

$$\begin{aligned} T_{Mod^*_{i,t}} &= b \times T_{Mod_{i,t}} + a, \\ a &= \tilde{T}^*_{Mod,t} - T_{Mod,t} \times \frac{\alpha^*_{Mod,t}}{\alpha_{Mod,t}}, \\ b &= \frac{\alpha^*_{Mod,t}}{\alpha_{Mod,t}}. \end{aligned} \tag{5.15}$$

In (5.15) $T^*_{Mod_{i,t}}$ is the temporally-adjusted daily model value at location i and day t. $T_{Mod_{i,t}}$ is the (unadjusted) model value at location i and day t; $\tilde{T}^*_{Mod,t}$ is the value of $\tilde{\mathbf{T}}^*_{Mod}$ on day t, $\mathbf{T}_{Mod,t}$ is the value of $\mathbf{T}_{Mod}$ on day t. Finally, $\boldsymbol{\alpha}^*_{Mod,t}$ and $\boldsymbol{\alpha}_{Mod,t}$ are seasonal SDs derived from temporally-adjusted and unadjusted model data, respectively. This linear transformation ensures that daily means and SDs of spatially-explicit model data match those of $\tilde{\mathbf{T}}^*_{Mod}$, for all $t \in T$. Finally, EQM was carried out on a monthly basis using temporally-adjusted, spatially explicit model (from 5.15) and observed data. EQM with temporal adjustment was carried out in this manner

because EQM is typically carried out using spatially explicit model data. Therefore, the temporal dependence of spatially-averaged model data was corrected, and the linear transformation in (5.15) to was applied to spatially explicit model values to reflect the temporal adjustment. Then, EQM was carried out using (adjusted), spatially-explicit model data.

Cross-validation

During cross-validation, estimates of $\sigma^2_{x_{\ell,Obs}}$ and $\sigma^2_{w,Obs}$ corresponding to each observed training set were used to adjust temporal dependence of all model data in each fold following steps in sections 4.3.3-4.3.4.

During the back-transformation, only trends derived from *model* data ($\boldsymbol{\alpha}_{Mod}$, $\boldsymbol{\delta}_{Mod}, and c_{Mod}$) were used. Thus, after the back-transformation, only the temporal dependence of model data is adjusted. Next, the linear transformation in (5.15) was applied to all spatially-explicit model values. After this step, spatially-explicit model data reflect the adjustment of temporal adjustment. Next, these spatially explicit model values were split into training and testing sets. Finally, EQM was carried out on a monthly basis using the adjusted, spatially-explicit, daily model data as well as observed data in training sets, and bias-correction was applied to spatially-explicit model values in test sets.

Extended discussion of effects of bias-correction on extreme quantiles

We provide an example of how DMTA and EQM perform for calibration and future simulations (2090-2099) for month 12. During the calibration period, the EQM TF for month 12 is somewhat parallel to the 1:1 line until the 99^{th} model quantile (17°C), when it increases sharply (Figure S4). While model quantiles below the sharp increase are increased by 1.7°C to 3°C, quantiles falling within the sharply increasing tail are increased by nearly 15°C. This sharp increase is due to the differences in maximum values of model and observed during the calibration period. The shape of the TF at tails is highly uncertain, as data in tails are, by definition, scarce. We also note that after the TF increases sharply, it becomes linear again (blue dashed line in Figure S4) and represents the extrapolation proposed by [75]. During the calibration period, the tail of the TF has little impact on the correction, and generally, model data corrected by EQM and DMTA are very similar (Figure S5). However, as temperatures rise in the future, it becomes increasingly likely that more model values will be transformed by the tail of the EQM TF, and consequently, those values will be increased substantially. Figure S6 shows daily boxplots for uncorrected, EQM-, and DMTA-corrected model data during the years 2090-2099. Because model values are more frequently transformed by the tail of the EQM TF, proportionately more model values are being increased substantially, resulting in potential artifacts. These inflated values are apparent in the boxplots for days 6, 9, 10, 11, 12, 21, 22, 23, and 24 in Figure S6. In contrast, because the correction of seasonal means and SDs applied in DMTA changes

smoothly over day of year, the correction remains much more consistent.

Inverse distance weighting

Inverse distance weighting (IDW) is a deterministic interpolation technique, so the size of prediction errors cannot be quantified [136]. Interpolated values are based on a weighted average of n nearest-neighbor observations. In IDW, observed values close to prediction locations are assumed to be more influential in the prediction compared to observed values far from prediction locations. As the power p and the number of nearest neighbors n increases, the smoothness of the interpolated surface increases. IDW is an exact interpolator, which means that if a prediction location s_0 corresponds to an observed location s_i, the predicted value at s_0 will be identical to the value at location s_i. The general equations for IDW are as follows:

$$
\begin{aligned}
Y(s_0) &= \sum_{i=1}^{n} w_i(s_0) Y(s_i), \\
w_i(s_0) &= \frac{\tilde{w}_i(s_0)}{\sum_{\ell=1}^{n} w_\ell(s_0)}, \\
\tilde{w}_i(s_0) &= \frac{1}{d(s_i, s_0)^p}.
\end{aligned}
$$

The IDW interpolated value at location s_0 is $Y(s_0)$, $d(s_i, s_0)$ is the distance between observed location s_i and prediction location s_0, n is the number of nearest-neighbor observed locations that contribute to the interpolated value $Y(s_0)$, and p is the power parameter.

Topographic downscaling

Topographic downscaling is a variation on IDW that is often used for high resolution downscaling [31]. Topographic downscaling consists of three main steps:

1. Construction of a historical, empirical lapse rate (relationship between TMAX and elevation using simple linear regression);

2. Adjustment of WRF data to reference elevation (200m) using estimated lapse rate parameters and interpolation of adjusted WRF data to desired locations;

3. Back-transformation of interpolated values using estimated lapse rate parameters.

Following methods by [31] and [12], we utilized historical (1970-1999) GHCND station records to calculate historical, elevational lapse rates for TMAX, using stations with at least 70% complete records. We estimated the elevational lapse rates for TMAX with a linear regression of the form (5.16):

$$T_{sta} = T_0 - \beta\phi_{sta} - \gamma z_{sta}, \tag{5.16}$$

where T_{sta} is the long-term average station TMAX, T_0 is the intercept, β is the coefficient for GHNCD station latitude (ϕ_{sta}), and γ is the coefficient for station elevation (z_{sta}).

Estimates for β and γ were -1.43 and -0.0059, respectively (Figure S1). The estimate of the elevation coefficient, β, refers to an elevational lapse rate of $5.9°\text{Ckm}^{-1}$, which corresponds closely to that found by [31], as well as the standard elevational lapse rate ($6.0°\text{Ckm}^{-1}$) [133].

WRF TMAX simulations were translated to reference elevation with (5.17)

$$T_{WRF,ref} = T_{WRF} - \gamma(z_{ref} - z_{WRF}), \tag{5.17}$$

where $T_{model,ref}$ is the value of TMAX (° C) at reference elevation, T_{WRF} is the WRF TMAX value (°C), γ is the estimated lapse rate ($^{\circ}\mathrm{Cm}^{-1}$) from 5.16, z_{ref} is the reference elevation (m), and z_{WRF} is WRF geopotential height (m).

Next, the transformed WRF data were interpolated to GHNCD station locations using IDW. We used a weight of 2 and 9 nearest neighbors for all IDW interpolation, following methods by [31]. Interpolated WRF data were back-transformed to reflect the effect of elevation (5.18)

$$T_{sta,\,interp} = T_{ref,\,interp} - \gamma(z_{sta} - z_{ref}). \tag{5.18}$$

In (5.18) $T_{sta,interp}$ is the elevation-adjusted value for TMAX, $T_{ref,interp}$ is the interpolated WRF value at a GHCND station location at reference elevation and z_{sta} and z_{ref} are the GHCND station and reference elevations. After back-transforming interpolated values at GHCND station locations, we applied empirical quantile mapping (EQM) at each station location.

$$X_{corr,t} = \mathrm{ecdf}^{-1}_{obs,m}(\mathrm{ecdf}_{raw,m}(X_{raw,t})), \tag{5.19}$$

In (5.19), $X_{corr,t}$ is the corrected daily value for TMAX on day t, $\mathrm{ecdf}^{-1}_{obs,m}$ is the inverse ecdf of GHCND station data for month m, and $\mathrm{ecdf}_{raw,m}$ is the ecdf of the WRF data for month m, and $X_{raw,t}$ is the uncorrected WRF TMAX value on day t. Next, bias-corrected WRF data at GHCND station locations were translated to

reference elevation with (5.20)

$$T_{EQM,ref} = T_{EQM} - \gamma(z_{ref} - z_{sta}), \tag{5.20}$$

where $T_{EQM,ref}$ is the bias corrected, interpolated value for TMAX (° C) at reference elevation, T_{EQM} is the bias corrected, WRF interpolation at a GHCND station location (°C), and γ, z_{ref} and z_{sta} are as defined in (5.18). Finally, the reference-adjusted, bias- corrected WRF interpolations at GHCND station locations were again interpolated to a 1km grid using IDW:

$$\begin{aligned} Y(s_0) &= \sum_{i=1}^{n} w_i(s_0) Y(s_i), \\ w_i(s_0) &= \frac{\tilde{w}_i(s_0)}{\sum_{\ell=1}^{n} w_\ell(s_0)}, \\ \tilde{w}_i(s_0) &= \frac{1}{d(s_i, s_0)^p}. \end{aligned}$$

In this context, $Y(s_0)$ is the IDW interpolated TMAX value at fine-scale grid cell s_0, $Y(s_i)$ is the value at station location s_i, $d(s_0, s_i)$ is the distance between GHCND station location s_i and the center of fine-scale grid cell s_0, and n and p were set to 9 and 2, respectively. Finally, the high-resolution values were translated to actual elevation with (5.21)

$$T_{fine,interp} = T_{ref,interp} - \gamma(z_{fine} - z_{ref}). \tag{5.21}$$

In (5.21), $T_{fine,interp}$ is the final downscaled value on the fine-scale grid, $T_{ref,interp}$ is the interpolated temperature value at reference elevation, z_{fine} is the elevation at

the fine-scale grid, and γ and z_{ref} are as defined in (5.6).

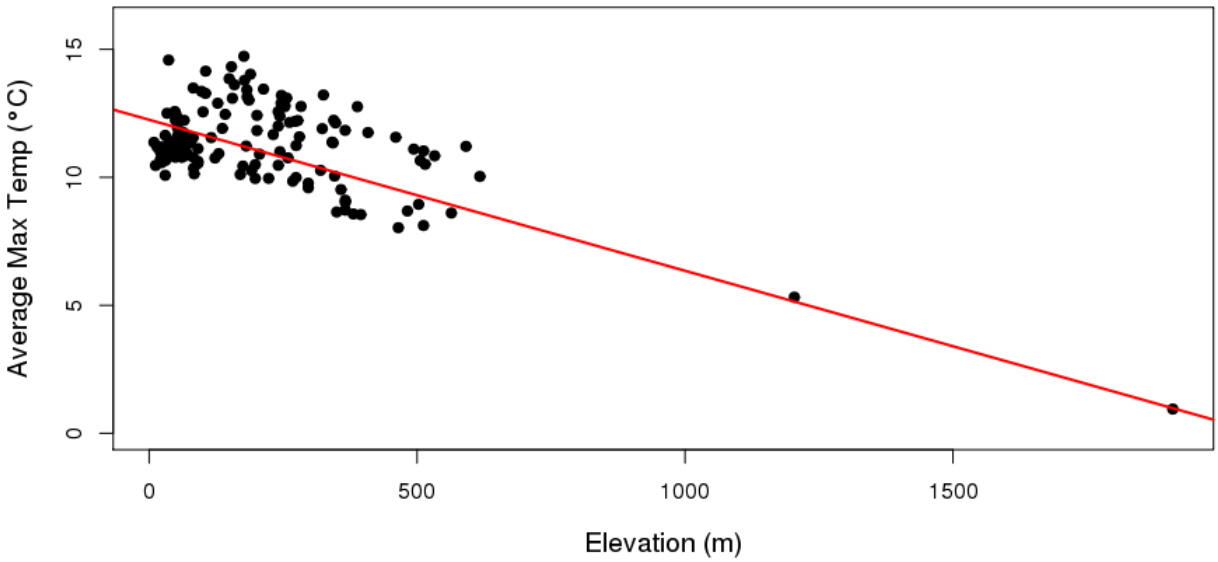

Figure S2. Elevational lapse rate adjustment for TMAX. Note: the elevational lapse rate did not change appreciably with omission of the two high elevation stations.

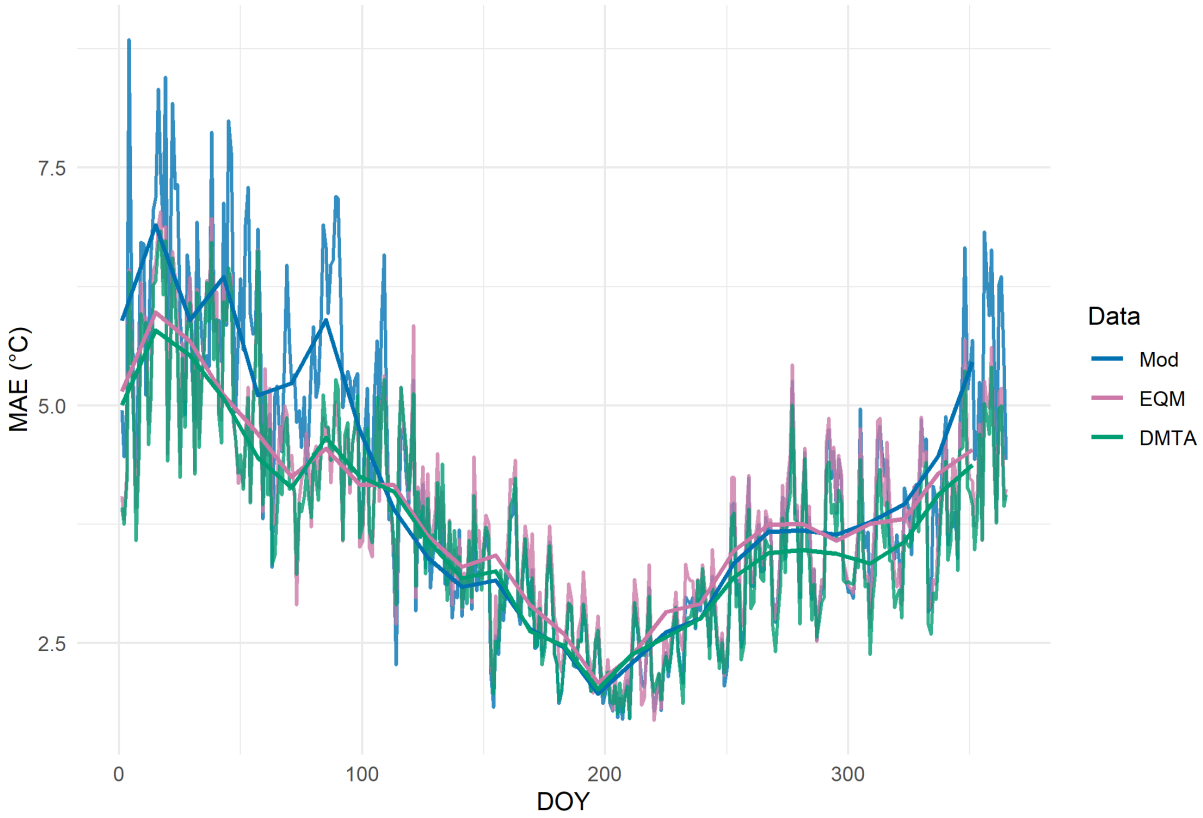

*Figure S3. Cross-validated results of MAE(°C) for uncorrected model (MOD), and EQM- and DMTA-corrected data (*without *correction of temporal dependence) relative to observed data during the historical period (1976-2005) over day of year (DOY). Averages over 14-day increments of DOY are overlaid to aid interpretation of the plot. Data are averaged over the 78 GHCND station locations and DOY.*

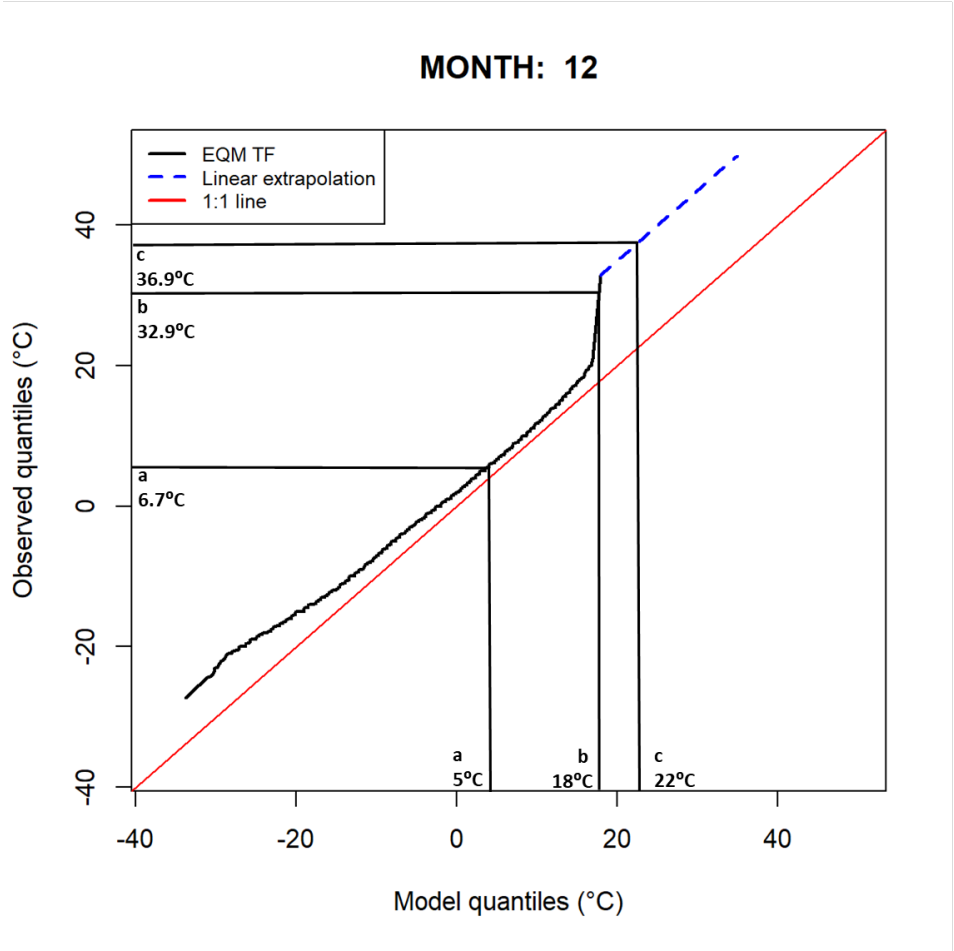

Figure S4. EQM transfer function (TF) (black line) for month 12 constructed using 1000 quantiles of observed and model data during the historical period (1976-2005). The shape of the TF beyond the range of data values is shown by the blue dashed line. The lines at a, b, and c show examples of how much model values are changed via the TF. At a), the TF increases a model value of 5° C to 6.7° C. At b) a model value increases from 18° C to 32.9° C, and at c) a model value of 22° C increases to 36.9° C. The correction at b) results in much larger increases in model values between 17.9 and 32.8° C, as the shape of the TF increases sharply. Furthermore, the correction at b) is unreliable, as that portion of the TF is based on interpolation of the maximum quantiles of model and observed data.

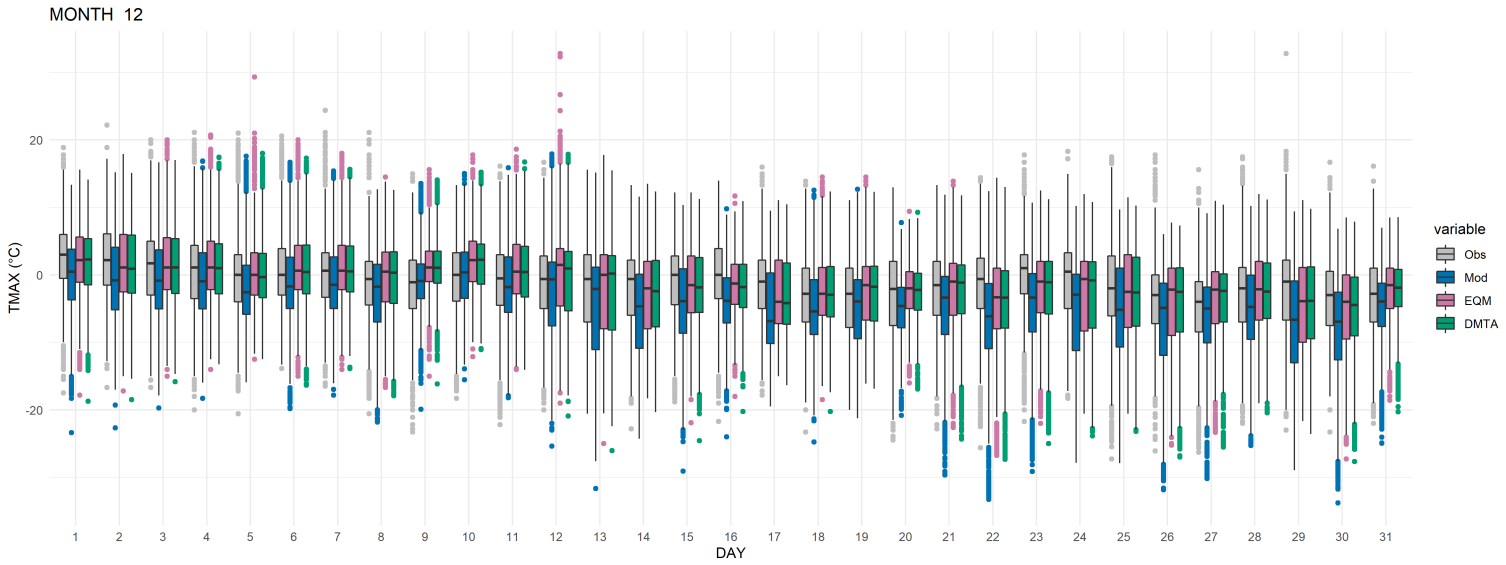

Figure S5. Boxplots showing the distribution of daily, spatially-explicit, uncorrected, bias-corrected model, and observed data (°C) for month 12 during the historical time period (1976-2005). Data at the 78 GHCND locations are represented in this plot.

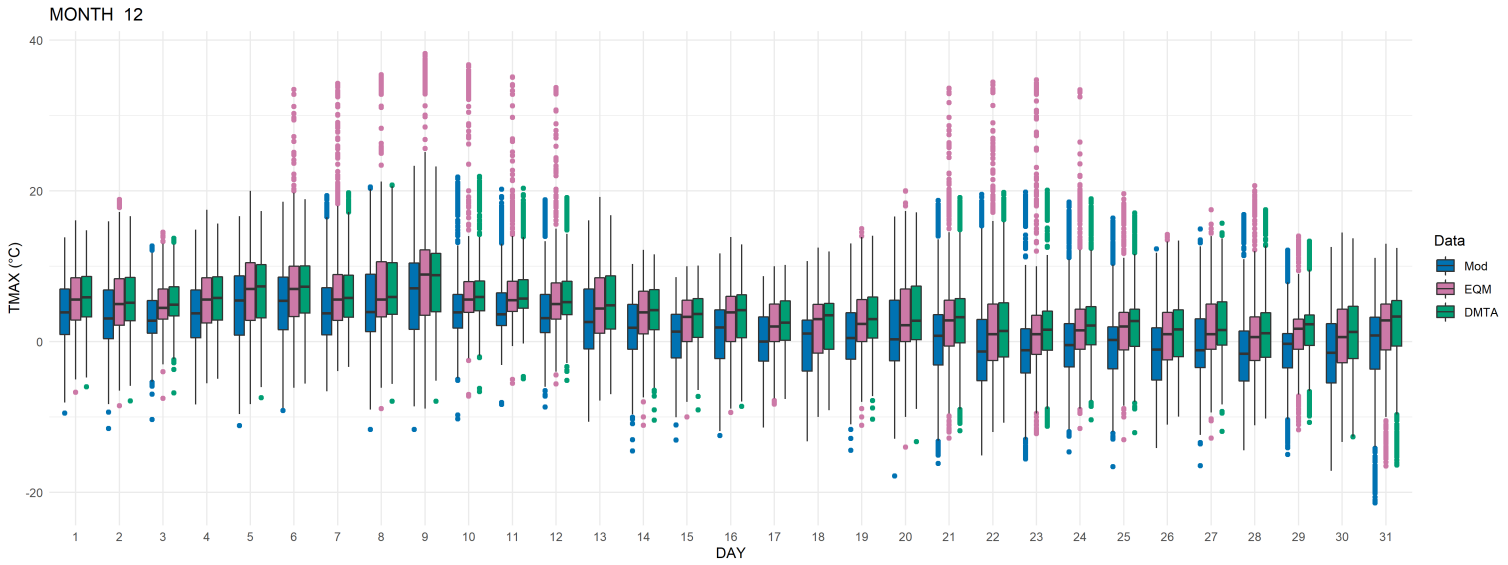

Figure S6. Boxplots showing the distribution of daily, spatially-explicit, uncorrected and bias-corrected model data (°C) for month 12 during a future time period (2090-2099). Data at the 78 GHCND locations are represented in this plot. Note the high frequency of TMAX values greater than 20 °C in EQM-corrected data.

Table S1. Mean daily TMAX (°C) for observed and model data by month during 1976-2005. Means are calculated using model data downscaled to the 78 GHCND station locations and GHCND station data.

MONTH	Mean	
	Obs	Mod
1	-4.84	-8.18
2	-2.45	-5.81
3	3.14	-0.68
4	11.22	9.06
5	18.87	18.42
6	23.54	23.49
7	25.89	25.59
8	24.80	24.56
9	20.01	20.13
10	12.81	12.74
11	5.78	4.47
12	-1.33	-3.82

Table S2. Monthly standard deviations (SDs) (°C) for observed and model data by month during 1976-2005. Standard deviations are calculated using model data downscaled to the 78 GHCND station locations and GHCND station data.

MONTH	SD	
	Obs	Mod
1	7.15	8.05
2	6.66	7.56
3	6.63	6.32
4	6.43	6.86
5	5.66	4.72
6	4.69	3.89
7	3.71	3.25
8	3.91	3.47
9	4.80	4.60
10	5.50	5.36
11	5.85	5.82
12	6.38	7.05

References

[12] G. E. Liston and K. Elder, "A meteorological distribution system for high-resolution terrestrial modeling (micromet)," *Journal of Hydrometeorology*, vol. 7, no. 2, pp. 217–234, 2006.

[31] J. M. Winter, B. Beckage, G. Bucini, R. M. Horton, and P. J. Clemins, "Development and evaluation of high-resolution climate simulations over the mountainous northeastern united states," *Journal of Hydrometeorology*, vol. 17, no. 3, pp. 881–896, 2016.

[75] J. Boé, L. Terray, F. Habets, and E. Martin, "Statistical and dynamical downscaling of the seine basin climate for hydro-meteorological studies," *International Journal of Climatology: A Journal of the Royal Meteorological Society*, vol. 27, no. 12, pp. 1643–1655, 2007.

[133] R. G. Barry, *Mountain weather and climate*. Psychology Press, 1992.

[136] C. K. Wikle, A. Zammit-Mangion, and N. Cressie, *Spatio-temporal Statistics with R*. CRC Press, 2019.

[199] M. Binois and R. B. Gramacy, "Hetgp: Heteroskedastic gaussian process modeling and sequential design in r," *Journal of Statistical Software*, vol. 98, pp. 1–44, 2021.

A4. An efficient spatiotemporal model using Kronecker products

Introduction

In Chapter 4, a method for temporally-coherent bias correction was presented. An additional spatio-temporal correction could be added to capture model biases over the study area and day of year. Here, I present an efficient spatio-temporal model that could be used for such a problem. The spatiotemporal model increases computational efficiency via the assumption of a separable space-time covariance function that can be written as a Kronecker product. I present the model, prove efficient computations for deriving the posterior conditional mean, and describe parameter estimation. While the modeling approach presented here is applied to modeling space-time residuals as part of a bias-correction method, the model could be used in a variety applications.

Here, model data refer to daily model simulations of maximum temperature (TMAX) generated by a regional climate model (RCM). Observed data refer to climate station observations from 78 locations. In this model, it is assumed that the covariance function is separable in space and time and that the temporal covariance function is periodic over a 365-day period. Prior to modeling, all model and station data over a 30 year historical period (1976-2005) were averaged over day of year, yielding $78 \times 365 = 2840$ total values in both the model and observed time series.

Model construction

Suppose the vector y_{it} represents the difference between model data and observed data for the i^{th} spatial location $i = 1 \dots n$ and day t $t = 1 \dots 365$. For any day t , y_t can be modeled as:

$$y_t = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{nt} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \Sigma_{n \times n} \right)$$

Assume $\Sigma_{n \times n}$ is known, and $\Sigma_{n \times n_{i,j}} = C(s_i, s_j; \theta)$, where $C(\theta)$ is the Gaussian covariance function with parameter vector θ and s denotes a spatial location. The parameter vector θ is estimated using maximum likelihood, which will be discussed later.

Now, consider approximately one month of data (we assume 30 days in one month here for simplicity). Suppose $Z_n = (y_1 \dots y_{30})$, where $y_i, i = 1 \dots 30$ is a vector of length n, where n is the number of stations.

$$Z_n = \begin{cases} \mathbf{y_1} \\ \vdots \quad \sim N\left(\mathbf{0_{n \times 30}}, \Sigma_t \otimes \Sigma_{n \times n}\right) \\ \mathbf{y_{30}} \end{cases} \tag{5.22}$$

where Σ_t is a correlation matrix of dimension 30.

The ultimate goal is to make predictions at locations over a fine-scale (1km) grid. Let Z_{grid} denote locations at which predictions are to be made. The vector Z_{total} can be partitioned as $(Z_{grid}, Z_n)^T$, where Z_n is defined as above. The distribution of Z_{total} can then be written as

$$\begin{pmatrix} Z_{grid} \\ Z_n \end{pmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Sigma_t \otimes \Sigma_{grid} & \Sigma_t \otimes \Sigma(X_{grid}, X_n) \\ \Sigma_t \otimes \Sigma(X_n, X_{grid}) & \Sigma_t \otimes \Sigma_{n \times n} \end{bmatrix} \right) \tag{5.23}$$

The sizes of components of the covariance matrix are

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} 30m \times 30m & 30m \times 30n \\ 30n \times 30m & 30n \times 30n. \end{bmatrix}$$

The conditional distribution $Z_{grid}|Z_n = z_n$ is also a Gaussian distribution with mean vector $\Sigma_{12}\Sigma_{22}^{-1}z_n$ and covariance matrix $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$:

$$\begin{aligned} Z_{grid}|Z_n = z_n &\sim N(\mu, V) \\ \mu &= \Sigma_t \otimes \Sigma(X_{grid}, X_n) \times (\Sigma_t \otimes \Sigma_{n\times n})^{-1} z_n \\ V &= \Sigma_t \otimes \Sigma_{grid} - \Sigma_t \otimes \Sigma(X_{grid}, X_n) \times (\Sigma_t \otimes \Sigma_{n\times n})^{-1} \times \Sigma_t \otimes \Sigma(X_n, X_{grid}), \end{aligned}$$

where z_n is a $30n$ vector consisting of n station measurements for a period of 30 days.

Alternatively, the distribution of Z_{total} may be written as

$$\begin{pmatrix} Z_n \\ Z_{grid} \end{pmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Sigma_t \otimes \Sigma_{n\times n} & \Sigma_t \otimes \Sigma(X_n, X_{grid}) \\ \Sigma_t \otimes \Sigma(X_{grid}, X_n) & \Sigma_t \otimes \Sigma_{grid} \end{bmatrix} \right)$$

In this case the sizes of the components of the covariance matrix are

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} 30n \times 30n & 30n \times 30m \\ 30m \times 30n & 30m \times 30m. \end{bmatrix} \tag{5.24}$$

The conditional distribution $Z_{grid}|Z_n = z_n$ is a MVN normal distribution with

mean vector $\Sigma_{21}\Sigma_{11}^{-1}z_n$ and covariance matrix $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$:

$$
\begin{aligned}
Z_{grid}|Z_n = z_n &\sim N(\mu, V) \\
\mu &= \Sigma_t \otimes \Sigma(X_{grid}, X_n) \times (\Sigma_t \otimes \Sigma_{n\times n})^{-1} z_n \\
V &= \Sigma_t \otimes \Sigma_{grid} - \Sigma_t \otimes \Sigma(X_{grid}, X_n) \times (\Sigma_t \otimes \Sigma_{n\times n})^{-1} \times \Sigma_t \otimes \Sigma(X_n, X_{grid}),
\end{aligned}
$$

where z_n is a $30n$ vector consisting of n station measurements for a period of 30 days. Since $\Sigma_t \otimes \Sigma_{grid}$ will be a very large matrix, we need to use efficient computational methods to invert it. For instance for μ we have $\Sigma_t \otimes \Sigma(X_{grid}, X_n) \times u$, where $u =$ solve$(\Sigma_t \otimes \Sigma_{n\times n})$; here, u is a $30n$ vector.

In a grid (g_1, g_2), the distance d_{ij} is Haverstine$(g_{1,i}, g_{1,j})$ + Haverstine$(g_{2,i}, g_{2,j}) = d_{1,ij} + d_{2,ij}$. Then the covariance matrix $\Sigma(X_{grid}, X_n) = \Sigma(d1)\times\Sigma(d2) = \Sigma(d1) \bigotimes \Sigma(d2)$.

The mean vector μ of the conditional distribution $Z_{grid}|Z_n = z_n$ can be computed efficiently using selected properties of Kronecker products.
If we assume $A \in M^m$ and $B \in M^n$ are nonsingular, then the following property holds:

KRON 5.0.1 (Kronecker inverse) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

The product of two Kronecker products yields another Kronecker product:

KRON 5.0.2 (Product of two Kronecker products)

$$(A \otimes B)(C \otimes D) = AC \otimes BD, \forall A \in M^{p,q}, B \in M^{r,s}, C \in M^{q,k}, D \in M^{s,l}.$$

The following equality will also be used:

KRON 5.0.3

$$Y = CXB^T \Leftrightarrow \text{vec}(Y) = (B \otimes C)\text{vec}(X),$$

In **KRON** 5.0.3, $\text{vec}(X)$ performs the operation that stacks the columns of a matrix from left to right into a vector. The operation $\text{reshape}(\mathbf{A}, m, n)$ is the matrix of shape $m \times n$ defined by $\text{vec}(\text{reshape}(\mathbf{A}, m, n)) = \text{vec}(\mathbf{A})$. Generally, $\text{reshape}(y, n/n_r, n_r) = \mathbf{A} \times \text{reshape}(x, n/n_r, n_r) \times \mathbf{B^T}$. Suppose $\Sigma(X_{grid}, X_n) \in M^{m,n}$, $\Sigma_{n\times n} \in M^n$, $\Sigma_t \in M^p$, and Z_n is of length $n \times p$. The posterior mean vector μ can be computed efficiently as $\Sigma(X_{grid}, X_n) \times \Sigma_{n\times n}^{-1} \times Z$.

Proof

$$\begin{aligned}
\Sigma_t \otimes \Sigma(X_{grid}, X_n) \times (\Sigma_t \otimes \Sigma_{n\times n})^{-1} Z_n = \\
\Sigma_t \otimes \Sigma(X_{grid}, X_n) \times (\Sigma_t^{-1} \otimes \Sigma_{n\times n}^{-1}) Z_n = \\
\Sigma_t \Sigma_t^{-1} \otimes (\Sigma(X_{grid}, X_n) \times \Sigma_{n\times n}^{-1}) Z_n = \\
I_p \otimes (\Sigma(X_{grid}, X_n) \times \Sigma_{n\times n}^{-1}) Z_n = \\
(\Sigma(X_{grid}, X_n) \times \Sigma_{n\times n}^{-1}) Z (I_p)^T = \\
\Sigma(X_{grid}, X_n) \times \Sigma_{n\times n}^{-1} \times Z \square
\end{aligned}$$

Evaluation of the Kronecker product $\Sigma_t \otimes \Sigma(X_{grid}, X_n)$ has computational complexity $O(nmp^2)$, where $m > p > n$. Evaluation of the Kronecker product $\Sigma_t \otimes \Sigma_{n\times n})$ is $O(n^2p^2)$, where $p > n$. However, the computation time of the simplified calculation is the computation of the matrix-matrix product $\Sigma(X_{grid}, X_n) \times \Sigma_{n\times n}^{-1}$, which is $O(mnp) < O(nmp^2))$.

Similarly, the conditional covariance matrix V, can be computed as $\Sigma_t \otimes (\Sigma_{grid} - \Sigma(X_{grid}, X_n)\Sigma_{n\times n}^{-1}\Sigma(X_n, X_{grid}))$:

Proof

$$V = \Sigma_t \otimes \Sigma_{grid} - \Sigma_t \otimes \Sigma(X_{grid}, X_n) \times (\Sigma_t \otimes \Sigma_{n\times n})^{-1} \times \Sigma_t \otimes \Sigma(X_n, X_{grid}) =$$

$$\Sigma_t \otimes \Sigma_{grid} - \Sigma_t \otimes \Sigma(X_{grid}, X_n) \times \Sigma_t^{-1} \otimes \Sigma_{n\times n}^{-1} \times \Sigma_t \otimes \Sigma(X_n, X_{grid}) =$$

$$\Sigma_t \otimes \Sigma_{grid} - \Sigma_t\Sigma_t^{-1} \otimes \Sigma(X_{grid}, X_n)\Sigma_{n\times n}^{-1}\Sigma(X_n, X_{grid}) =$$

$$\Sigma_t \otimes \Sigma_{grid} - I_p\Sigma_t \otimes \Sigma(X_{grid}, X_n)\Sigma_{n\times n}^{-1}\Sigma(X_n, X_{grid}) =$$

$$\Sigma_t \otimes (\Sigma_{grid} - \Sigma(X_{grid}, X_n)\Sigma_{n\times n}^{-1}\Sigma(X_n, X_{grid}))\square$$

Unfortunately, there is no way to eliminate the Kronecker product in this computation. Instead, approximation or simulation of the posterior covariance matrix V using [215] would be a better option. This will be discussed later. The conditional posterior covariance matrix V is of less importance than the posterior mean μ for this application.

Efficient spatiotemporal models: adding the nugget and marginal variance

We now consider the (more realistic) case in which the marginal variance and nugget effect are incorporated into the model. We begin as before, considering

$$y_t = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{nt} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \sigma^2\Sigma_{n\times n} + \tau^2\mathbf{I}\right)$$

Again, $\Sigma_{n\times n}$ is assumed known, and $\Sigma_{n\times n_{i,j}} = C(s_i, s_j; \theta)$, where $C(\theta)$ is the Gaussian covariance function with parameter vector θ. σ^2 is the marginal variance (partial sill), and τ^2 is the nugget effect. The parameter vector θ will be estimated using maximum likelihood, as discussed later.

Suppose $Z_n = (y_1 \dots y_{30})$, where $y_i, i = 1 \dots 30$, is a vector of length n, where n is the number of stations.

$$Z_n = \begin{cases} \mathbf{y_1} \\ \vdots \\ \mathbf{y_{30}} \end{cases} \sim N\left(\mathbf{0_{n\times 30}}, \sigma^2(\Sigma_t \otimes \Sigma_{n\times n}) + \tau^2\mathbf{I}\right) \tag{5.25}$$

Σ_t is a correlation matrix of dimension 365.

The ultimate goal is to make predictions at locations over a fine-scale (1km) grid. Let Z_{grid} denote locations at which predictions are to be made. The vector Z_{total} can be partitioned as $(Z_n, Z_{grid})^T$, where Z_n is defined as above. The distribution of Z_{total} can be written as

$$\begin{pmatrix} Z_n \\ Z_{grid} \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_1 = \mathbf{0} \\ \mu_2 = \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma^2(\Sigma_t \otimes \Sigma_{n\times n}) + \tau^2\mathbf{I} & \sigma^2\mathbf{\Sigma}_t \otimes \mathbf{\Sigma}(X_n, X_{grid}) \\ \sigma^2\mathbf{\Sigma}_t \otimes \mathbf{\Sigma}(X_{grid}, X_n) & \sigma^2\mathbf{\Sigma}_t \otimes \mathbf{\Sigma}_{grid} \end{bmatrix}\right)$$

In this case the sizes of the components of the covariance matrix are

$$\begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} 30n \times 30n & 30n \times 30m \\ 30m \times 30n & 30m \times 30m. \end{bmatrix} \tag{5.26}$$

The conditional distribution $Z_{grid}|Z_n = z_n$ is a MVN normal distribution with

mean vector $\mu_1 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(z_n - \mu_2)$ and covariance matrix $\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$. Since we are assuming $\mu_1 = \mu_2 = \mathbf{0}$ we will omit μ_1 and μ_2 for simplicity:

$$Z_{grid}|Z_n = z_n \sim N(\mu, V)$$
$$\mu = \sigma^2\mathbf{\Sigma}_t \otimes \mathbf{\Sigma}(X_{grid}, X_n) \times (\sigma^2(\mathbf{\Sigma}_t \otimes \mathbf{\Sigma}_{n\times n}) + \tau^2\mathbf{I})^{-1} z_n$$

$$V = \sigma^2\mathbf{\Sigma}_t \otimes \mathbf{\Sigma}_{grid} - \sigma^2\mathbf{\Sigma}_t \otimes \mathbf{\Sigma}(X_{grid}, X_n) \times (\sigma^2(\mathbf{\Sigma}_t \otimes \mathbf{\Sigma}_{n\times n}) + \tau^2\mathbf{I})^{-1} \times \sigma^2\mathbf{\Sigma}_t \otimes \mathbf{\Sigma}(X_n, X_{grid}),$$

where z_n is a $30n$ vector consisting of n station measurements for a period of 30 days. However, calculating the conditional mean and covariance matrix is computationally infeasible for large n.

We will simplify the calculations by using the eigendecomposition of $\mathbf{\Sigma}_t$ and $\mathbf{\Sigma}_{n\times n}$. To reduce confusion, we will define $\mathbf{C} = \mathbf{\Sigma}_t, \mathbf{R} = \mathbf{\Sigma}_{n\times n}$, and $\mathbf{G} = \mathbf{\Sigma}(X_{grid}, X_n)$. Now let the eigendecomposition of $\mathbf{C}$ and $\mathbf{R}$ be $\mathbf{U_c S_c U_c^T}$ and $\mathbf{U_r S_r U_r^T}$, respectively. First, we will work through the simplification of $(\sigma^2(\mathbf{\Sigma}_t \otimes \mathbf{\Sigma}_{n\times n}) + \tau^2\mathbf{I})^{-1} z_n = (\sigma^2(\mathbf{C} \otimes \mathbf{R}) + \tau^2\mathbf{I})^{-1} z_n$. After substituting, the expression for the posterior conditional mean vector μ is:

$$\sigma^2(\mathbf{C} \otimes \mathbf{G}) \times (\sigma^2\mathbf{C} \otimes \mathbf{R} + \tau^2\mathbf{I})^{-1} z_n. \tag{5.27}$$

Simplifying the right hand side of (5.27) $(\sigma^2\mathbf{C} \otimes \mathbf{R} + \tau^2\mathbf{I})^{-1} z_n$ gives us :

$$
\begin{aligned}
(\sigma^2\mathbf{C}\otimes\mathbf{R}+\tau^2\mathbf{I})^{-1}z_n = \\
(\sigma^2\mathbf{U_c}\mathbf{S_c}\mathbf{U_c^T}\otimes\mathbf{U_r}\mathbf{S_r}\mathbf{U_r^T}+\tau^2\mathbf{I})^{-1}z_n = \\
(\sigma^2\mathbf{U_c}\mathbf{S_c}\otimes\mathbf{U_r}\mathbf{S_r}\times\mathbf{U_c^T}\otimes\mathbf{U_r^T}+\tau^2\mathbf{I})^{-1}z_n = \\
((\mathbf{U_c}\otimes\mathbf{U_r})\times(\sigma^2\mathbf{S_c}\otimes\mathbf{S_r}+\tau^2\mathbf{I})(\mathbf{U_c^T}\otimes\mathbf{U_r^T}))^{-1}z_n = \\
(\mathbf{U_c^T}\otimes\mathbf{U_r^T})^{-1}\times(\sigma^2\mathbf{S_c}\otimes\mathbf{S_r}+\tau^2\mathbf{I})^{-1}\times(\mathbf{U_c}\otimes\mathbf{U_r})^{-1}z_n
\end{aligned}
$$

Let $\mathbf{X}=(\sigma^2\mathbf{S_c}\otimes\mathbf{S_r}+\tau^2\mathbf{I})^{-1}$. Since $\mathbf{S_c}$ and $\mathbf{S_r}$ are diagonal matrices of eigenvalues, the inverse is computationally cheap to calculate. It is easy to see that the eigenvalues of $(\sigma^2\mathbf{C}\otimes\mathbf{R})+\tau^2\mathbf{I}$ are the diagonal elements of $(\sigma^2\mathbf{S_c}\otimes\mathbf{S_r}+\tau^2\mathbf{I})$ (Theorems 5.2.1 and 5.2.2). Thus, we have

$$
(\mathbf{U_c^T}\otimes\mathbf{U_r^T})^{-1}\times\mathbf{X}\times\mathrm{vec}(\mathbf{U_r}^{-1}\mathbf{Z_n}\mathbf{U_c^{-T}}). \tag{5.28}
$$

In (5.28), $\mathbf{Z_n}$ is a matrix of appropriate dimension, constructed from stacking z_n into columns from left to right.

Now, substituting (5.28) for $(\mathbf{C}\otimes\mathbf{R}+\tau^2\mathbf{I})^{-1}z_n$ in (5.27) we obtain:

$$
\sigma^2(\mathbf{C}\mathbf{U_c^{-T}}\otimes\mathbf{G}\mathbf{U_r^{-T}})\times\mathbf{X}\times\mathrm{vec}(\mathbf{U_r}^{-1}\mathbf{Z_n}\mathbf{U_c^{-T}}).
$$

Now, let $\mathrm{vec}(\mathbf{D})=\mathbf{X}\times\mathrm{vec}(\mathbf{U_r}^{-1}\mathbf{Z_n}\mathbf{U_c^{-T}})$. The expression can then be further simplified:

$$\sigma^2(\mathbf{C}\mathbf{U_c}^{-\mathbf{T}} \otimes \mathbf{G}\mathbf{U_r}^{-\mathbf{T}}) \times \mathrm{vec}(\mathbf{D}) =$$
$$\sigma^2\mathbf{G}\mathbf{U_r}^{-\mathbf{T}}\mathbf{D}(\mathbf{C}\mathbf{U_c}^{-\mathbf{T}})^{\mathbf{T}} =$$
$$\sigma^2\mathbf{G}\mathbf{U_r}^{-\mathbf{T}}\mathbf{D}\mathbf{U_c}^{-\mathbf{1}}\mathbf{C}^{\mathbf{T}}.$$

It is reasonable to assume the addition of a temporal nugget, α^2, to the temporal covariance matrix to account for small scale variation over time. In this case, $\Sigma_t' = \Sigma_t + \alpha^2$. In the following calculation, Σ_t' will denote the temporal covariance matrix including the nugget, and Σ_t will denote the temporal covariance matrix without the nugget.

$$\begin{pmatrix} Z_n \\ Z_{grid} \end{pmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma^2(\Sigma_t' \otimes \Sigma_{n\times n}) + \tau^2\mathbf{I} & \sigma^2\mathbf{\Sigma}_t \otimes \mathbf{\Sigma}(X_n, X_{grid}) \\ \sigma^2\mathbf{\Sigma}_t \otimes \mathbf{\Sigma}(X_{grid}, X_n) & \sigma^2\mathbf{\Sigma}_t \otimes \mathbf{\Sigma}_{grid} \end{bmatrix}\right)$$

.

The calculation of the conditional mean vector is nearly identical to the steps in equations 5.6 and 5.7, except that Σ_t' will be denoted by $\mathbf{C}'$. The eigendecomposition of $\mathbf{C}'$ will by denoted by $\mathbf{U_c'}\mathbf{S_c'}\mathbf{U_c^{T'}}$. After substituting (5.28) for $(\mathbf{C}' \otimes \mathbf{R} + \tau^2\mathbf{I})^{-1}z_n$, we obtain:

$$\sigma^2(\mathbf{C}\mathbf{U_c}^{-\mathbf{T}'} \otimes \mathbf{G}\mathbf{U_r}^{-\mathbf{T}}) \times \mathbf{X} \times \mathrm{vec}(\mathbf{U_r}^{-1}\mathbf{Z_n}\mathbf{U_c}^{-\mathbf{T}'}).$$

Further simplification yields:

$$\sigma^2(\mathbf{C}\mathbf{U_c}^{-\mathbf{T}'} \otimes \mathbf{G}\mathbf{U_r}^{-\mathbf{T}}) \times \mathrm{vec}(\mathbf{D}) =$$
$$\sigma^2\mathbf{G}\mathbf{U_r}^{-\mathbf{T}}\mathbf{D}(\mathbf{C}'\mathbf{U_c}^{-\mathbf{T}'})^{\mathbf{T}} =$$
$$\sigma^2\mathbf{G}\mathbf{U_r}^{-\mathbf{T}}\mathbf{D}\mathbf{U_c}^{-\mathbf{1}'}\mathbf{C}^{\mathbf{T}}.$$

Because the Σ_{grid} is too large to be stored in memory, realizations of the conditional distribution could be approximated using a method described by [215]. However, in this context only the conditional mean is of interest.

Theorem 5.0.1 *Suppose λ is an eigenvalue of an $n \times n$ matrix $\mathbf{A}$. Then $\alpha\lambda$ is an eigenvalue of $\alpha\mathbf{A}$.*

Proof Let $x \neq 0$ be an eigenvector of $\mathbf{A}$ for λ. Then

$$\begin{aligned}(\alpha\mathbf{A})x &= \alpha(\mathbf{A}x) \\ &= \alpha(\lambda x) \\ &= (\alpha\lambda)x \quad \square\end{aligned}$$

Theorem 5.0.2 *Suppose $\lambda_i, i = 1...n$ are eigenvalues of an $n \times n$ matrix $\mathbf{A}$. Let c be a scalar, and let $\mathbf{I}$ be the $n \times n$ identity matrix. Then $\lambda_i + c$ are the eigenvalues of $\mathbf{A} + c\mathbf{I}$.*

Proof Let $x \neq 0$ be an eigenvector of $\mathbf{A}$ for λ. Then

$$\begin{aligned}\mathbf{A}x = \lambda x; c\mathbf{I}x = c\mathbf{I}x \\ \mathbf{A}x + c\mathbf{I}x = \lambda x + c\mathbf{I}x \\ (\mathbf{A} + c\mathbf{I})x = (\lambda + c\mathbf{I})x \quad \square\end{aligned}$$

Estimating model parameters

Because it is of interest to obtain smooth estimates of the posterior mean over both space and time, a smooth covariance function must be selected. I chose the infinitely-differentiable

Gaussian covariance function (5.29) to construct the spatial correlation matrix and a periodic covariance function for the temporal correlation matrix (5.30). The Gaussian covariance function is infinitely differentiable and leads to smooth realizations.

$$C(d, \phi) = \exp(-(d/\phi)^2) \tag{5.29}$$

$$C(d, \ell, \rho) = \exp(-2\ell^2 \sin^2(\pi d/\rho)) \tag{5.30}$$

$d = d(i, j)$ represents the distance between locations i and j in (5.29)and between days i and j in (5.30). Distances are calculated in km and in days for spatial and temporal correlation matrices, respectively. In 5.29, ϕ is the spatial scale parameter governing the strength of spatial correlation. In (5.30) ℓ controls the strength of temporal correlation, and ρ is the period, which is set to 365, as there are 365 days in a year.

Maximum likelihood (ML) is used to estimate ϕ, ℓ, the marginal variance σ^2, the overall nugget τ^2, and the temporal nugget, α^2. When no nugget effects are assumed, only a marginal variance, the distribution of $\mathbf{Y}$ is

$$\mathbf{Y} \sim MVN(\mathbf{0_{p \times n}}, \sigma^2 \boldsymbol{\Sigma_t} \otimes \boldsymbol{\Sigma_{n \times n}}) \tag{5.31}$$

where σ^2 is the marginal variance of Y. In this study, Y is typically a daily, mean-0 time series that has a marginal standard deviation of 1. Now, let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma_t} \otimes \boldsymbol{\Sigma_{n \times n}}$, $N = p \times n$, μ be a $\mathbf{0}$ vector of length N and Y be a vector of observed, mean-0 data of length N. Then the log likelihood function for (5.31) can be written as

$$\log(\ell(\mathbf{Y})) = -N/2(2\pi) - \log(|\boldsymbol{\sigma^2 \Sigma}|^{1/2}) - \frac{1}{2}\Sigma_{n=1}^{N}(\mathbf{Y} - \mu)^T \boldsymbol{(\sigma^2 \Sigma)^{-1}}(\mathbf{Y} - \mu) =$$
$$-\frac{N}{2}(2\pi) - \log(|\boldsymbol{\sigma^2 \Sigma}|^{1/2}) - \frac{1}{2}\Sigma_{n=1}^{N}\mathbf{Y}^T(\sigma^2 \boldsymbol{\Sigma})^{\mathbf{-1}}\mathbf{Y}. \tag{5.32}$$

ML estimation of the scale parameters for $\mathbf{\Sigma_t}$ and $\mathbf{\Sigma_{n\times n}}$, ϕ_t and $\phi_{n\times n}$, respectively, was done by minimizing $\log(\ell(Y))$ (5.32) using the Nelder Mead algorithm implemented in the `dfoptim` package [216] in R.

Observe that $|\mathbf{\Sigma}|$ can be expressed as $|\mathbf{\Sigma_t}|^n|\mathbf{\Sigma_{n\times n}}|^p$, where here $p = 365$. Let $\mathbf{L_t}$ and $\mathbf{L_s}$ be the Cholesky factors of Σ_t and $\Sigma_{n\times n}$, respectively. We can then write $\log(|\mathbf{\Sigma}|) = \log(|\mathbf{L_t}|^{\mathbf{n}}|\mathbf{L_s}|^{\mathbf{p}}) = n \times \sum_{i=1}^{p} \mathbf{L_{s_{ii}}} + p \times \sum_{j=1}^{n} \mathbf{L_{s_{jj}}}$. This is a more stable and efficient computation than calculating $\log|\mathbf{\Sigma}|$.

In the case when both the overall nugget, τ^2 and the temporal nugget α^2 are incorporated in the model, the distribution of $\mathbf{Y}$ can be expressed as a matrix-variate normal distribution:

$$Y \sim \mathcal{MN}_{n\times p}(\text{vec(Y)}|\mathbf{0}_{n\times p}, \sigma^2\mathbf{\Sigma'_t} \otimes \mathbf{\Sigma_{n\times n}} + \tau^2\mathbf{I}_{n\times p}). \tag{5.33}$$

In 5.33, $\mathbf{\Sigma}'_t$ denotes the temporal covariance matrix with the temporal nugget added ($\mathbf{\Sigma}'_t = \mathbf{\Sigma}_t + \alpha^2\mathbf{I}$). Let $\mathbf{C} = \mathbf{\Sigma}'_t, \mathbf{R} = \mathbf{\Sigma}_{n\times n}$, $\mathbf{G} = \mathbf{\Sigma}(X_{grid}, X_n)$ and the eigendecomposition of $\mathbf{C}$ and $\mathbf{R}$ be $\mathbf{U_c S_c U_c^T}$ and $\mathbf{U_r S_r U_r^T}$, respectively. The log likelihood for (5.33) is written as [217]:

$$\log(\ell(Y)) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln|(\sigma^2\mathbf{S_c}\otimes\mathbf{S_r} + \tau^2\mathbf{I})| - \frac{1}{2}\text{vec}(\mathbf{U_r^T Y U_c})^T(\sigma^2\mathbf{S_c}\otimes\mathbf{S_r} + \tau^2\mathbf{I})^{-1}\text{vec}(\mathbf{U_r^T Y U_c}) \tag{5.34}$$

In (5.34), $\ln|(\sigma^2\mathbf{S_c} \otimes \mathbf{S_r} + \tau^2\mathbf{I})|$ is computationally cheap to evaluate because $(\sigma^2\mathbf{S_c} \otimes \mathbf{S_r} + \tau^2\mathbf{I})$ is a diagonal matrix. The log determinant is the sum of the diagonal elements.

All parameters ($\sigma^2, \tau^2, \alpha^2, \phi$, and ρ) were estimated using the data, and estimates were used as initial values during optimization.

The variance of $\mathbf{Y}$ was obtained by computing the daily variance and then averaging over all days. This estimated variance is the total variance of all the variance components (σ^2, τ^2, and α^2); the sum of all variance components should not exceed the estimated total variance. To implement constraints on variance components, an additional penalty term

was added to the likelihood function, $-0.5\lambda(\nu^2 - \sigma^2 - \tau^2 - \alpha^2)$, where $\lambda > 0$ controls the strength of the penalty, and ν is the estimated sum of all variance components (for this data, the estimated sum of all variance components was 0.285^2). As λ increases, the penalty for exceeding the overall variance increases. In practice, consistent estimation of all parameters is difficult, as shown in [218]. Estimation can be improved by specifying either the spatial or temporal range parameter (ϕ or ρ), but with noisy data, estimates may warrant further scrutiny.

Conclusion

The Gaussian spatio-temporal model described here can be efficiently implemented by using various properties of the Kronecker product in combination with known matrix properties. While the model is used for an application of bias-correction, the model is highly generalizable and could be used in other applications that require modeling of spatiotemporal data. The main limitation of this model is in the estimation of parameters, as it is difficult to obtain consistent estimation of both the marginal variance and range parameters, especially with noisy data. More research is needed to refine the estimation of parameters in spatiotemporal models with noisy data.

Appendix

Simulating multivariate normal random variables

Suppose we wish to simulate random variables from a multivariate distribution $\mathbf{X} \sim MVN(\mu, \mathbf{\Sigma})$. Define a collection of n univariate normal random vectors, $Z_i \sim N(0,1), i = 1, \ldots, n$.

Any semi-positive definite matrix can be decomposed via the Cholesky decomposition as $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^{\mathbf{T}}$, where $\mathbf{L}$ is lower triangular. Then

$$\mathbf{X} = \mu + \mathbf{L}\mathbf{Z} \tag{5.35}$$

has mean vector μ and covariance matrix $\mathbf{L}\mathbf{L}^{\mathbf{T}} = \Sigma$. Cholesky decomposition can only be applied to Hermitian positive definite matrices, but if pivoting is utilized, Cholesky factorization can be applied to positive semidefinite matrices [219]. Because the covariance matrices used in this application are in general positive semi-definite, symmetric pivoting of the form $\mathbf{A} \rightarrow \mathbf{P}\mathbf{A}\mathbf{P}^{\mathbf{T}}$, where $\mathbf{P}$ is a permutation matrix, will be used. The permutation matrix $\mathbf{P}$ reorders the diagonal elements of $\mathbf{A}$. This results in near-zero entries are not encountered until the algorithm has completed the r^{th} iteration of the main loop in Cholesky decomposition (r is the perceived rank of the matrix). Choleskty decomposition with complete pivoting is regarded as a stable algorithm [219].

In general, any symmetric positive semi definite matrix $\mathbf{A}$ has a factorization

$$\mathbf{P}^{\mathbf{T}}\mathbf{A}\mathbf{P} = \mathbf{R}^{\mathbf{T}}\mathbf{R},$$

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix},$$

where $\mathbf{P}$ is a permutation matrix, R_{11} is an upper triangular matrix of dimension $r \times r$ with positive elements on the diagonal, and $\text{rank}(\mathbf{A}) = r$. This factorization is accomplished via complete pivoting, in which at each stage, the largest diagonal element in the active submatrix is set to the pivot position.

An alternative method for finding $\mathbf{L}\mathbf{L}^{\mathbf{T}}$ is using singular value decomposition (SVD). In contrast to the Cholesky decomposition, any square or rectangular, real-valued matrix $\mathbf{A}$

with dimension $m \times n$ can be factorized as $\mathbf{UDV^T}$, where $\mathbf{U}$ is an $m \times m$ orthogonal matrix of left singular values, $\mathbf{D}$ is an $m \times n$ matrix with non-negative real numbers (eigenvalues) on the diagonal, and $\mathbf{V}$ is an $n \times n$ orthogonal matrix of right singular values. For symmetric matrices of size $m \times m$, $\mathbf{U} = \mathbf{V^T}$, so the factorization may be written as $\mathbf{UDU^T}$. In this case $\mathbf{D}$ is a square matrix of size $m \times m$ with eigenvalues along the diagonal. The columns of $\mathbf{U}$ are the eigenvectors corresponding to the eigenvalues in $\mathbf{D}$. Since covariances matrices are always symmetric, this is the factorization we will use. SVD can be used to find $\mathbf{L^*L^{*T}}$:

$$\mathbf{UDU^T} = \mathbf{UD^{1/2}D^{1/2}U^T} = \mathbf{L^*L^{*T}}$$

although in general $\mathbf{L^*L^{*T}} \neq \mathbf{LL^T}$. MVN realizations can be constructed similarly as in (5.35), except that $\mathbf{L}$ is replaced by $\mathbf{L^*} = \mathbf{UD^{1/2}}$.

The covariance matrix for separable Gaussian covariance functions will be of the form $\Sigma = aA \otimes B + c\mathbf{I}$, for scalar values a and $c > 0$; for large n, simulating realizations will become computationally infeasible. However, computation time can be expedited by utilizing the matrix-normal, rather than multivariate normal, distribution. First, we will consider the case when $a = 1$ and $c = 0$ (i.e. the marginal variance is 1 and there is no nugget effect). The matrix-normal distribution is related to the multivariate normal distribution as follows:

$$\mathbf{X} \sim \mathcal{MN}_{n \times p}(\mathbf{M}, \mathbf{U}, \mathbf{V}),$$

if and only if

$$\text{vec}(\mathbf{X}) \sim \mathcal{MVN}_{np}(\text{vec}(\mathbf{M}), \mathbf{V} \otimes \mathbf{U}). \tag{5.36}$$

The probability distribution function (PDF) of the matrix-normal distribution is given

by:

$$P(\mathbf{X}|\mathbf{M},\mathbf{U},\mathbf{V}) = \frac{\exp(\mathrm{tr}([\mathbf{V^{-1}}(\mathbf{X}-\mathbf{M})^{\mathbf{T}}\mathbf{U^{-1}}(\mathbf{X}-\mathbf{M})])}{(2\pi)^{np/2}|\mathbf{V}|^{\mathbf{n/2}}|\mathbf{U}|^{\mathbf{p/2}}}$$

$$\mathbf{M} = n \times p$$

$$\mathbf{U} = n \times n$$

$$\mathbf{V} = p \times p.$$

Proof:

$$\begin{aligned} \frac{-1}{2}\mathrm{tr}[\mathbf{V^{-1}}(\mathbf{X}-\mathbf{M})^{\mathbf{T}}\mathbf{U^{-1}}(\mathbf{X}-\mathbf{M})] = \\ \frac{-1}{2}\mathbf{vec}(\mathbf{X}-\mathbf{M})^{\mathbf{T}}\mathbf{vec}(\mathbf{U^{-1}}(\mathbf{X}-\mathbf{M})\mathbf{V^{-1}}) = \\ \frac{-1}{2}\mathbf{vec}(\mathbf{X}-\mathbf{M})(\mathbf{V^{-1}}\otimes\mathbf{U^{-1}})\mathbf{vec}(\mathbf{X}-\mathbf{M}) \end{aligned} \tag{5.37}$$

,

where (5.37) is the argument of the exponent of the multivariate normal PDF. The proof is complete by noting that $|\mathbf{V}\otimes\mathbf{U}| = |\mathbf{V}|^{\mathbf{n}}|\mathbf{U}|^{\mathbf{p}}\square$.

Suppose $\mathbf{Y}$ has a matrix normal distribution with mean matrix $Z = \mathbf{0}_{n\times p}$ and covariance matrices $\mathbf{U}$ and $\mathbf{V}$. Then for a mean matrix $\mathbf{M}$ and linear transformations $\mathbf{L}$ and $\mathbf{R}$, $\mathbf{Y}$ can be expressed as:

$$\mathbf{Y} = \mathbf{M} + \mathbf{LZR}, \tag{5.38}$$

where $\mathbf{Y}$ has a matrix-normal distribution with parameters $\mathbf{M}, \mathbf{LL^T}, \mathbf{R^TR}$. In (5.38), Cholesky decomposition is used to construct the matrices $\mathbf{L}$ and $\mathbf{R}$, the Cholesky factors of $\mathbf{U}$ and $\mathbf{V}$, respectively. If SVD is used, then $\mathbf{LL^T} = \mathbf{U_u D_u^{1/2} D_u^{1/2} U_u^T}$ is the SVD of $\mathbf{U}$ and $\mathbf{R^TR} = \mathbf{U_v D_v^{1/2} D_v^{1/2} U_v^T}$ is the SVD of $\mathbf{V}$. This, in turn implies that (5.38) can be used to generate MVN variates with $\mathbf{L} = \mathbf{U_u D_u^{1/2}}$ and $\mathbf{R} = \mathbf{D_v^{1/2} U_v^T}$

Suppose we wish to sample from (5.22), for t timesteps and m spatial locations. We begin by expressing (5.22) as a matrix-normal distribution with a mean matrix $\mathbf{M}$ of size $m \times t$ consisting entirely of 0s. The matrix of N(0,1) random variates $\mathbf{Z}$ will also be of size $m \times t$. We set U =: $\Sigma_{n\times n}$, and V =: Σ_t and use either Cholesky decomposition or SVD to obtain $\mathbf{L}$ and $\mathbf{R}$. The decompositions of $\mathbf{U}$ and $\mathbf{V}$ represent the computationally most expensive steps in the linear transformation (5.38). Cholesky decomposition, which requires $n^3/4$ flops, is slightly faster than SVD. However, the time savings using either Cholesky or SVD over (5.35) are substantial, because the Kronecker product ($O(n^6)$) does not have to be computed.

If the marginal variance a and nugget c ($a > 0, c > 0$) are added, the linear transformation in (5.38) can still be used, but will be modified as

$$\mathbf{Y} = \mathbf{M} + a\mathbf{LZR} + c\mathbf{I}. \tag{5.39}$$

In (5.39), $c\mathbf{I}$ is easily simulated, as it represents a $N(0, c)$ process.

Eigendecomposition of symmetric matrices

Eigendecomposition is crucial for obtaining computationally efficient solutions in large-n problems, such as the one discussed in this chapter. Recall that a nonzero vector x is an eigenvector of an $n \times n$ matrix $\mathbf{A}$ with eigenvalue λ if $\mathbf{A}x = \lambda x$. The matrix $\mathbf{A} - \lambda\mathbf{I}$ is singular and x is a nonzero vector in the nullspace of $(\mathbf{A} - \lambda\mathbf{I})$. The eigenvalues of $\mathbf{A}$ are the roots of the characteristic polynomial $p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = 0$. Every square matrix has at least one unique eigenvalue, and when counted with multiplicity, an $n \times n$ matrix has n eigenvalues (which may or may not be unique).

For symmetric matrices eigenvectors corresponding to eigenvalues are orthogonal, and all the eigenvalues are real. Every real symmetric matrix can be decomposed as $\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^\mathbf{T}$,

where $\Lambda = diag(\lambda_1, ...\lambda_n)$ is diagonal with real diagonal elements. The diagonal elements of λ are the eigenvalues of $\mathbf{A}$ in decreasing order. Also, $\mathbf{Q^{-1}} = \mathbf{Q^T}$ because $\mathbf{Q}$ is an orthogonal matrix. If the matrix is symmetric and positive definite, then all the eigenvalues are positive.

Gaussian process regression

A process convolution is an efficient way to implement Gaussian (GP) regression. A GP is a collection of finite realizations that follow a multivariate normal (MVN) distribution. The realizations can be completely characterized by their mean vector μ and covariance matrix Σ. The covariance function $C(\cdot)$ must result in a valid covariance matrix; that is $\Sigma_n(x, x')$ must be positive semi definite:

$$a\Sigma_n a^T \geq 0, \forall a.$$

In GP regression, a prior is placed on the function $f(x)$ that generates $y_1 \dots y_n$ given input $x_1 \dots x_n$. Thus $f(x) \sim \mathcal{GP}$. The data, $Z_n(X_n, Y_n)$ have a MVN likelihood. In GP regression, we wish to determine which function could have resulted in the data Z_n; this is the conditional distribution $f(x)|Z_n$ and it is also the posterior distribution. Recall that

the conditional distribution of a MVN distribution is also MVN normal:

$$
\begin{aligned}
X &= \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \begin{pmatrix} q \times 1 \\ (N-q) \times 1 \end{pmatrix} \\
\mu &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} q \times 1 \\ (N-q) \times 1 \end{pmatrix} \\
\Sigma &= \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \begin{pmatrix} q \times q & q \times (N-q) \\ (N-q) \times q & (N-q) \times (N-q) \end{pmatrix} \\
\tilde{\mu} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\
\tilde{\Sigma} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}
\end{aligned}
$$

Now, suppose we want to make predictions at a vector of new locations. It is obvious that the observations Y_n and $f(\mathcal{X})$ are multivariate normal. Suppose that $f(x)|Z_n \sim MVN(\mu(\mathcal{X}), \Sigma(\mathcal{X}))$ and that $\mathcal{X}$ has m rows. We can apply the previous result for conditioning on a MVN distribution to obtain the posterior predictive distribution. Here, X_n are training data, $\mathcal{X}$ are new test data locations, and Y_n are observations:

$$
\begin{aligned}
\mu(\mathcal{X}) &= \Sigma(\mathcal{X}, X_n)\Sigma_n^{-1}Y_n \\
\Sigma(\mathcal{X}) &= \Sigma(\mathcal{X}, \mathcal{X}) - \Sigma(\mathcal{X}, X_n)\Sigma_n^{-1}\Sigma(\mathcal{X}, X_n)^T
\end{aligned}
$$

Clearly, this approach is not feasbile for large n, as it requires the inversion of an $n \times n$ matrix. If n is large, the covariance matrix can be approximated using singular value decomposition (SVD). SVD can be used to extract the most important information from a matrix and thus reduce its rank from m to k, where $k < m$. When SVD is applied to a covariance matrix, Σ, we obtain the following:

$$\Sigma = UDV^T = UDU^T = UD^{1/2}D^{1/2}U^T = KK^T.$$

where D is a diagonal matrix of singular values. Since Σ is symmetric, $U = V$. Also, U is an $m \times m$ orthogonal matrix and Σ is an $m \times n$ matrix, but here $n \times n$ diagonal matrix ($\Sigma_{ij} = 0$ for $i \neq j$) with entries $\sigma_i \geq 0, \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k \geq 0, andk = min(m, n)$. V is an $n \times n$ orthogonal matrix.

Now, say $Z = Kx$, where $x \sim N(0, I)$. Realizations from Z can now be generated and the K matrix can be expressed as:

$$K = \left[u_1 d_1^{1/2} | \quad u_2 d_2^{1/2} | \quad \ldots | \quad u_\ell d_\ell^{1/2}\right],$$

where $\ell < m$.

Approximation of covariance matrices

Often, computing realizations from a MVN distribution is limited by the size of the covariance matrix. For large covariance matrices, direct simulation of MVN realizations may not be feasible. However, with approximation techniques such as that proposed by [215], the simulation of MVN realizations is possible. Here, I prove the proposed method by [215].

I will begin with a review of the conditional Gaussian distribution. Suppose $Z_2(s) \sim MVN(0, \Sigma)$, and $Z_1(s) = Z_2 + \epsilon$, where $\epsilon \sim MVN(0, \sigma^2)$. Thus, Z_1 is a "noisy" version of Z_2. Now, suppose we observe noisy observations at locations s_2 and s_5 ($Z_1 = (Z_1(s_2), Z_1(s_5))^T$), and we wish conduct conditional inference around locations $s_1 \ldots s_{10}$ ($Z_2 = (Z_2(s_1) \ldots Z_2(s_{10}))^T$). Given that Z_1 and Z_2 are Gaussian-distributed, the vector $Z = [Z_1\ Z_2]^T$ is also Gaussian-distributed with a mean of μ and a covariance matrix Ω:

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}, \begin{pmatrix} q \times 1 \\ (N-q) \times 1 \end{pmatrix}$$

$$\mu = \begin{pmatrix} \mu_1 = 0 \\ \mu_2 = 0 \end{pmatrix}, \begin{pmatrix} q \times 1 \\ (N-q) \times 1 \end{pmatrix}$$

$$\Omega = \begin{pmatrix} \Sigma_{11} + \sigma^2 I & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \begin{pmatrix} q \times q & q \times (N-q) \\ (N-q) \times q & (N-q) \times (N-q) \end{pmatrix}$$

To conduct inference about $s_1 \ldots s_{10}$ given noisy observations s_2, s_5 we obtain the conditional distribution $Z_2|Z_1 = z_1$, which is also Gaussian-distributed with a mean of $\hat{\mu}$ and a covariance matrix $\hat{\Omega}$:

$$\hat{\mu} = \mu_1 + \Sigma_{21}(\Sigma_{11} + \sigma^2)^{-1} z_1$$

$$\hat{\Omega} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

Figure 5.22 shows noisy observations at locations s_2 and s_5, along with realizations taken from the conditional distribution $Z_2|Z_1 = z_1$. Note that the values in the conditional mean vector at locations s_2 and s_5 are exactly the same as the observed values at those locations. However, there is some uncertainty around locations s_2 and s_5.

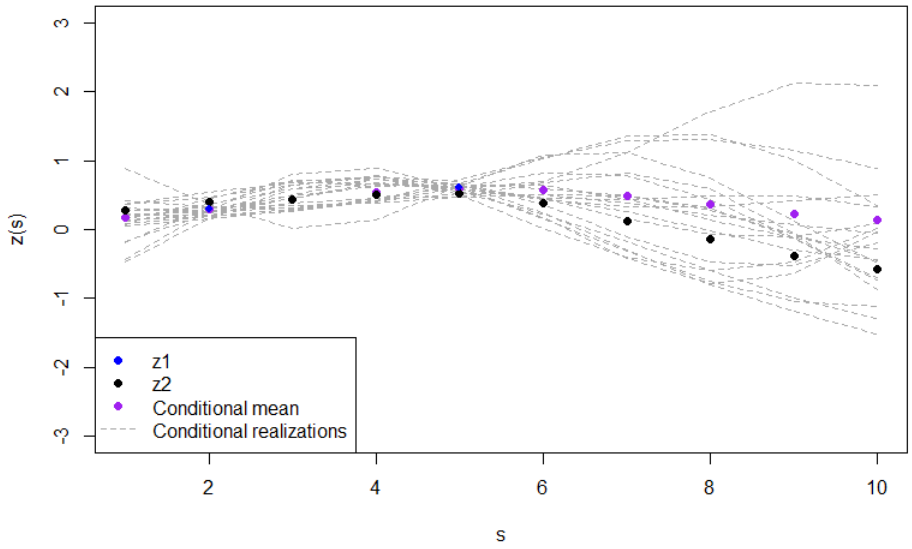

Figure 5.22: Noisy observations s_2 amd s_5 (blue), conditional mean (purple), conditional realizations (gray).

Implementation [215] outlines an approach for the efficient approximation of covariance matrices. I will prove the correctness of the approach below and implement the associated algorithm. The goal is to develop a Monte Carlo approach to approximate the posterior conditional covariance matrix. Specifically, using the nomenclature in [215], the goal is to approximate the conditional covariance matrix of $z|y$. We are given that $z|y \sim MVN(\hat{z}, V)$, where $\hat{z} = \Sigma K^T(K\Sigma K + \sigma^2\mathbf{I})^{-1}y$ and $V = \Sigma - \Sigma K^T(K\Sigma K^T + \sigma^2\mathbf{I})^{-1})K\Sigma$. [215]'s approach involves taking a sample $u \sim MVN(0, \Sigma)$. Next, a psuedo y is constructed as $y^* = Ku + \epsilon$, where $\epsilon \sim MVN(0, \sigma^2\mathbf{I})$. Then u^* is constructed: $u^* = u - \Sigma K^T(K\Sigma K + \sigma^2\mathbf{I})^{-1}y^*$. A sample from the appropriate conditional distribution can then be obtained as $\tilde{z} = \hat{z} + u^*$. It is easy to see that $y^* \sim MVN(0, K\Sigma K^T + \sigma^2\mathbf{I})$. To find the distribution of u^*, we must first find the distribution of $\Sigma K^T(K\Sigma K + \sigma^2\mathbf{I})^{-1}y^*$.

Because y^* is Gaussian-distributed, we can use the linear transformation properties of

random variables to determine the distribution of $\Sigma K^T(K\Sigma K + \sigma^2\mathbf{I})^{-1}\mathbf{y}^*$.

$$\Sigma K^T(K\Sigma K + \sigma^2\mathbf{I})^{-1}y^* \sim MVN(0, \Omega)$$
$$\Omega = \Sigma K^T(K\Sigma K + \sigma^2\mathbf{I})^{-1}(K\Sigma K^T + \sigma^2\mathbf{I})\left[\Sigma K^T(K\Sigma K^T + \sigma^2\mathbf{I})^{-1}\right]^T =$$
$$\Sigma K^T\mathbf{I}\left[\Sigma K^T(K\Sigma K^T + \sigma^2\mathbf{I})^{-1}\right]^T =$$
$$\Sigma K^T(K\Sigma K^T + \sigma^2\mathbf{I})^{-1}K\Sigma. \tag{5.40}$$

Thus, $\Sigma K^T(K\Sigma K^T + \sigma^2\mathbf{I})^{-1}\mathbf{y}^* \sim MVN(0, \Sigma K^T(K\Sigma K^T + \sigma^2\mathbf{I})^{-1}K\Sigma)$.
Also, observe that the vector $[u, y^*]^T$ has the following distribution:

$$\begin{bmatrix} u \\ y^* \end{bmatrix} \sim MVN \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma & K\Sigma \\ \Sigma K^T & K\Sigma K^T + \sigma^2 I \end{bmatrix}$$

The vector $[u, Dy^*]$, where $D = \Sigma K^T(K\Sigma K^T + \sigma^2 I)^{-1}$, has the following distribution:

$$\begin{bmatrix} u \\ Dy^* \end{bmatrix} \sim MVN \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma & \Sigma K^T(K\Sigma K^T + \sigma^2 I)^{-1}K\Sigma \\ \Sigma K^T(K\Sigma K^T + \sigma^2 I)^{-1}K\Sigma & \Sigma K^T(K\Sigma K^T + \sigma^2 I)^{-1}K\Sigma \end{bmatrix},$$

where $Cov(u, Dy^*)$ and $Cov(Dy^*, u)$ can be derived using the fact that for constant matrices A and B and random vectors X and Y, $Cov(AX, BY) = ACov(X, Y)B^T$. Note that here, D is a constant matrix. We have

$$Cov(u, Dy^*) =$$

$$ICov(u, y^*)D^T =$$

$$\Sigma K^T(K\Sigma K^T + \sigma^2 I)^{-1}K\Sigma$$

and

$$Cov(Dy^*, u) =$$

$$DCov(y^*, u)I^T =$$

$$\Sigma K^T(K\Sigma K^T + \sigma^2 I)^{-1}K\Sigma$$

.

Thus, $Cov(u, Dy^*) = Cov(Dy^*, u) = Var(Dy^*, Dy^*)$.

Now, we have that

$$u - Dy^* = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} \begin{bmatrix} u & Dy^* \end{bmatrix}.$$

Thus, by linear transformation of a MVN random variable, we have that

$$u - Dy^* \sim MVN \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} \begin{bmatrix} \Sigma & \Sigma K^T(K\Sigma K^T + \sigma^2 I)^{-1}K\Sigma \\ \Sigma K^T(K\Sigma K^T + \sigma^2 I)^{-1}K\Sigma & \Sigma K^T(K\Sigma K^T + \sigma^2 I)^{-1}K\Sigma \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}^T.$$

After expanding the matrix multiplication, we have

$$u - Dy^* \sim MVN \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma & -\Sigma K^T(K\Sigma K^T + \sigma^2 I)^{-1}K\Sigma \\ -\Sigma K^T(K\Sigma K^T + \sigma^2 I)^{-1}K\Sigma K^T & \Sigma K^T(K\Sigma K^T + \sigma^2 I)^{-1}K\Sigma \end{bmatrix}$$

Now, we must determine the structure of the covariance matrix of $[u - Dy^*]$. The

variance of the sum of two random variables X and Y is

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

$$Var(X+Y) = E((X+Y) - E(X+Y))^2 =$$

$$E((X+Y) - (E(X) - E(Y)))^2 =$$

$$E((X - E(X)) + (Y - E(Y)))^2 =$$

$$E\left[(X - E(X))^2 + (Y - E(Y))^2 + 2((X - E(X))(Y - E(Y)))\right] =$$

$$Var(X) + Var(Y) + 2Cov(X,Y).$$

Thus, the variance-covariance matrix of $u^* = [u - Dy^*] =$

$$Var(u) + Var(Dy^*) + 2Cov(u, Dy^*) =$$

$$\Sigma + \Sigma K^T (K\Sigma K^T + \sigma^2 I)^{-1} K\Sigma - 2(\Sigma K^T (K\Sigma K^T + \sigma^2 I)^{-1} K\Sigma =$$

$$\Sigma - \Sigma K^T (K\Sigma K^T + \sigma^2 I)^{-1} K\Sigma$$

Thus, $u^* \sim MVN(0, \Sigma - \Sigma K^T (K\Sigma K^T + \sigma^2 \mathbf{I})^{-1} K\Sigma)$. Next, we must show that $\tilde{z} = \hat{z} + u^*$ is a sample from $z|y$. Since $\hat{z}$ is a constant mean vector, $\tilde{z}$ is a simple linear transformation of u^*, and

$$\tilde{z} \sim MVN(\Sigma K^T (K\Sigma K + \sigma^2 I)^{-1} y, \Sigma - \Sigma K^T (K\Sigma K^T + \sigma^2 I)^{-1} K\Sigma),$$

which is the desired result. Figure 5.23 plot shows the same data as in Figure 5.22 except that realizations are generated using the approximation method proposed by [215]. The approximated realizations in Figure 5.23 are nearly indistinguishable from actual realizations shown in Figure 5.22.

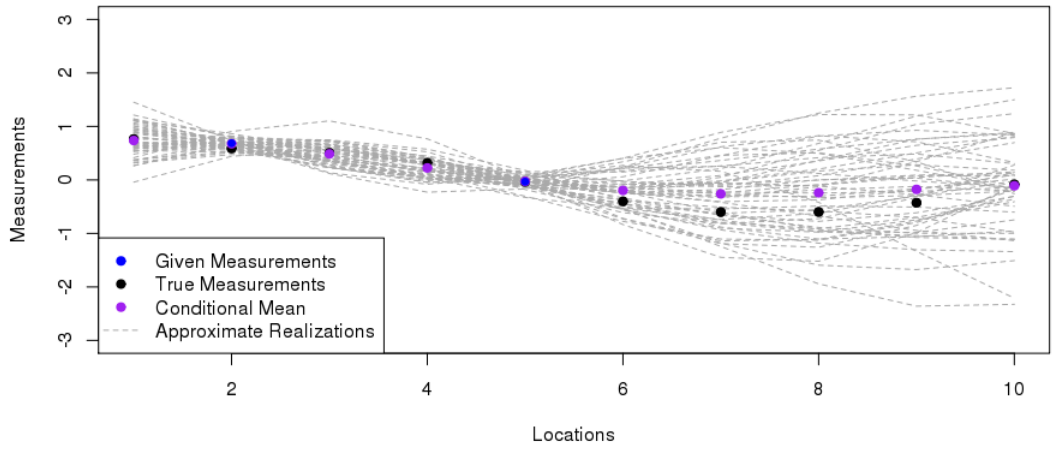

Figure 5.23: Simulation of realizations using the covariance matrix approximation method proposed by [215]

References

[215] D. Nychka, C. Wikle, and J. A. Royle, "Multiresolution models for nonstationary spatial covariance functions," *Statistical Modelling*, vol. 2, no. 4, pp. 315–331, 2002.

[216] R. Varadhan, H. W. Borchers, and M. R. Varadhan, *Package 'dfoptim'*, 2016.

[217] O. Stegle, C. Lippert, J. M. Mooij, N. Larence, and K. Borgwardt, "Efficient inference in matrix-variate gaussian models with iid observation noise," 2011.

[218] H. Zhang, "Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 250–261, 2004.

[219] N. J. Higham, *Analysis of the Cholesky decomposition of a semi-definite matrix*. Oxford University Press, 1990.