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

Maike Holthuijzen [*] Brian Beckage [†] Dave Higdon[‡]
Patrick J. Clemins[§] Jonathan Winterr[¶]

Abstract

High-resolution, bias-corrected climate data is necessary for climate impact studies at local scales. Using gridded historical data for bias correction is convenient but may contain biases resulting from interpolation. Long-term, quality-controlled station data better represent true climatological measurements, but as the spatial distribution of climate stations over the landscape is irregular, station data are challenging to incorporate into downscaling and bias-correction approaches. The use of station data in creating full-coverage, bias-corrected climate products is not well-represented in the literature. In this study, we developed and compared six novel methodologies using station data to produce daily, high-resolution, bias-corrected climate products with maximum temperature simulations from a regional climate model (RCM). The methods differed with respect to interpolation methods and bias-correction techniques. We quantified performance of six methods with the root mean square error (RMSE) and Perkins skill score (PSS) and used two ANOVA models to analyze how performance metrics varied among methods. We temporally validated the six methods using two calibration sets of observed station data (1980-1989 and 1980-2014) and two testing sets of RCM data (1990-2014 and 1980-2014). RMSE for all methods varied considerably throughout the year and was larger in cold seasons, while PSS was more consistent. Quantile-mapping bias-correction techniques performed best in improving PSS, while simple linear transfer functions performed best in improving RMSE. For the 1980-1989 station calibration dataset, simple quantile-mapping techniques outperformed empirical quantile mapping (EQM) in improving PSS; conversely, when the calibration and testing sets represented the same time period, EQM performed best in improving PSS. No one method simultaneously improved RMSE and PSS; however, the simple quantile-mapping based techniques perform as well or better than more sophisticated methods such as empirical quantile mapping.

Key Words: bias-correction, downscaling, high-resolution, maximum temperature, kriging, inverse distance weighting

1 Introduction

High-resolution ($\leq$1km) gridded climate products with both fine spatial and temporal resolutions crucial to assessing the effects of a changing climate on social and

[*]University of Vermont, Burlington, VT, Complex Systems and Data Science Center
[†]University of Vermont, Burlington, VT, Dept. of Plant Biology
[‡]Virginia Tech, Blacksburg, VA, Dept of Statistics
[§]University of Vermont, Burlington, VT, Dept. of Computer Science
[¶]Department of Geography, Hanover, New Hampshire, Dept. of Geography

ecological systems at local scales (Flint and Flint, 2012; Holden et al., 2011; Franklin et al., 2013). Furthermore, such products are also important for climate impact assessments, agricultural modeling (Hansen, 2005), and ecological studies (Holden et al., 2011; Fridley, 2009). 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 (Ekström et al., 2015; Lafon et al., 2013) and require substantial computation power. In addition, GCMs simplify climate processes through parameterization schemes, resulting in the unrealistic representation of some climate processes (Maraun et al., 2017). Consequently, output from GCMs is characterized by a non-trivial degree of bias (Lafon et al., 2013; Cannon et al., 2020; Maraun et al., 2017). 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 this study, we develop six novel methodologies for generating daily, high-resolution, bias-corrected climate products. We apply the methods to maximum temperature simulations over a region northeastern United States.

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 a GCM or reanalysis data. An RCM simulates climate processes at a finer resolutions than forcing data by incorporating fine-scale landscape and atmospheric processes (Ekström et al., 2015; Caldwell et al., 2009; Leung et al., 2003; Wilby et al., 2004). RCMs are computationally intensive, although they typically require less processing power than GCMs (Feser et al., 2011; Giorgi et al., 2009). *Statistical downscaling*, in contrast, involves establishing statistical relationships between coarse-scale and fine-scale climate variables, often leveraging local, observed phenomena or attributes (Wilby et al., 2004). Statistical downscaling is computationally efficient and can be applied to both precipitation and temperature (Mearns et al., 2003; Fang et al., 2015). In contrast to dynamical downscaling, a substantial amount of observational data is necessary to derive statistical relationships necessary for statistical downscaling (Wilby et al., 2004). Approaches for statistical downscaling include regression-based methods (Ekström et al., 2015), principal components analysis (Huth, 1999; Kettle and Thompson, 2004), weather classification schemes, and weather generators (Wilby et al., 2004). Recently, machine learning methods such as artificial neural networks (Schoof and Pryor, 2001), deep learning (Vandal et al., 2017), and random forests (Hutengs and Vohland, 2016) have been used for downscaling both temperature and precipitation variables. Downscaling is especially important for accurate representation of temperature in regions characterized by topographically varied terrain (Hanssen-Bauer et al., 2005; Holden et al., 2011).

High-resolution climate data can also be generated by applying statistical downscaling to RCM output (Haas and Pinto, 2012). While this combination of dynamical and statistical downscaling is complex, it is an effective workflow for generating high-resolution climate data simulations (Engen-Skaugen, 2007; Winter et al., 2016; Han et al., 2019).

Bias-correction is another post-processing procedure that can correct the mean, variance, and higher moments of climatological variables (Lafon et al., 2013; Cannon et al., 2020). Generally, bias-correction methods can be classified into four categories: 1) linear scaling (Lenderink et al., 2007; Hay et al., 2000); 2) nonlinear scaling (Leander and Buishand, 2007); 3) distribution mapping (Piani et al., 2010); and 4) empirical (distribution-free) quantile mapping (Teutschbein and Seibert, 2012; Cannon et al., 2015; Wood et al., 2002). 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 (Maurer and Duffy, 2005; Hayhoe et al., 2008; Wood et al., 2004; Bennett et al., 2014; Fang et al., 2015). EQM, a sophisticated technique, can correct the mean, variance, and

higher moments of temperature and precipitation variables (Fang et al., 2015; Themeßl et al., 2011). 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 (Shrestha et al., 2017; Lenderink et al., 2007). Most bias-correction methods assume stationarity of model errors over time (Roberts et al., 2019), and sufficient observational data is necessary to derive robust transfer functions.

Gridded, observational climate products (e.g. Livneh, (Livneh et al., 2015); Daymet, (Thornton et al., 2012); and PRISM, (Daly et al., 2000)) are often used for bias-correction due to their extensive spatial and temporal coverage. However, the interpolation algorithms used to create gridded climate products can introduce bias (Behnke et al., 2016) and additional uncertainty when used for bias-correcting climate model output (Walton and Hall, 2018). In particular, (Behnke et al., 2016) 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, Bishop and Beier (2013) found that in the Northeastern US, PRISM data products (Daly et al., 2000) 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 (National Oceanic and Atmospheric Administration, 2018). Station data represent direct climatolgical measurements and have extensive global availability (Peterson and Vose, 1997; Durre et al., 2010). The use of station data, rather than gridded observational products, for bias-correction could potentially reduce uncertainty in bias-correction. Station data are often used to validate the accuracy bias-corrected climate model output but can also be effectual for bias-correcting output from climate models. For instance, Mejia et al. (2012) 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 hydrologic model. Poggio and Gimona (2015) showed that incorporating station data in a geostatstical 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.

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 are often irregular, especially in mountainous and high-elevation regions (Daly et al., 2000). Another challenge is that for constructing full-coverage, bias-corrected 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 used 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 datasets. The 1-km resolution was chosen based on spatial resolution requirements for local climate impact assessments (Wang et al., 2012; Winter et al., 2016). All of the six methods are specifically developed to address the challenge of creating full-coverage, high-resolution, bias-corrected climate products using only station data. We apply the methods to daily RCM simulations of 2-meter maximum air temperature (TMAX) over a region in the northeastern United States. Methods differ mainly with respect to bias-correction and interpolation techniques. We validate the methods using two calibration time periods, and we assess the ability of methods to bias-correct in a spatially coherent manner by applying a spatial cross-validation procedure.

This paper aims to address the following questions:

1. How do the different bias-correction techniques and interpolation methods affect performance metrics (root-mean square error and Perkins skill score)?
2. Does performance among methods vary by month, and do performance metrics improve when elevation lapse rates are used during downscaling?
3. Is any one method particularly well-suited for high-resolution downscaling and bias correction?

The article is organized as follows: in section 2, we describe the study area, station and WRF 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.

2 Methods

2.1 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 1). Four watersheds drain into Lake Champlain. The Green Mountains, Adirondack Mountains, and White Mountains span portions of Vermont, New York, and New Hampshire, respectively (Winter et al., 2016). Elevations in the study area range from 30 to 1500 m above mean sea level (MSL). The region is topographically varied; the northern portion of the study region is relatively flat, while mountain ranges cover the remaining portion.

Daily historical TMAX simulations over 1980-2014 were generated by the Advanced Weather and Research Forecasting model (WRF) version 3.9.1 (Skamarock et al., 2019). WRF is a widely used as both a regional climate model numerical weather prediction system (Skamarock et al., 2019). 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 (Dee et al., 2011) and was downscaled to 4 km using three one-way nests (36 km, 12 km, 4km) (Huang et al., 2020). Only output from the inner, 4km resolution domain was used in this study. Specific physics settings for WRF are shown in in Table 7. 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 (National Oceanic and Atmospheric Administration, 2018; Peterson and Vose, 1997). 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 by interpolating elevation from a 30m digital elevation model (DEM) (U.S. Geological Survey, 2019). Elevation values were interpolated to the 1km grid using inverse distance weighting (IDW).

2.2 Description of downscaling and bias-correction methods

Elevation has a major effect on climatological variables such as maximum temperature (Winter et al., 2016; Barry, 1992). Therefore, during downscaling, it is important to account for lapse rates, especially in topographically rich regions, such as the Lake Champlain Basin (Winter et al., 2016). 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 an 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 station locations or the fine-scale grid as downscaling.

2.2.1 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; for a detailed description, see Appendix, Figure 12). 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 (Wikle et al., 2019; Poggio and Gimona, 2015; Daly, 2006).

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 (Winter et al., 2016). IDW with topographic downscaling combines IDW with elevational lapse rates to adjust for elevation and has been applied to high-resolution downscaling (Winter et al., 2016) (full details on topographic downscaling and IDW are given in Supplementary Material, section 4). 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 Winter et al. (2016), who used a very 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 the methods in (Winter et al., 2016).

For EQM_krig, WRF simulations were interpolated via kriging. To account for fine-scale elevation, elevation (either at station locations or at the fine-scale grid) was included as a covariate in a universal kriging model. In the case when fine-scale elevation was not accounted for, ordinary kriging was used. IDW and kriging were implemented with the `gstat` package (Gräler et al., 2016) in R (R Core Team, 2018). 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, section 3). 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 metrics. Kriging methods often work better for interpolating sparsely distributed data (Varouchakis and Hristopulos, 2013; Hofstra et al., 2008), such as the GHCND station data, but IDW is simple and computationally efficient. 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 (Daly, 2006).

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 (1). The EQM transfer function is expressed by the empirical cumulative distribution function (ecdf) and its inverse (ecdf^{-1}).

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

In 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 (Gudmundsson, 2016) 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; for a detailed description see Appendix, Figure 13) has potential 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).

First, 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 suited for interpolating sparsely distributed data (Varouchakis and Hristopulos, 2013). Based on inspection of empirical variograms, the effective range, partial sill and nugget for the kriging model were set to 150, 15, and 0.2, respectively for all kriging models. 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 elevational 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 (1).

2.2.2 Linear transfer function-based methods: quantile mapping and simple 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 linear transfer functions derived from regression relationships between GHCND station data and WRF simulations (Figure 2). 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 a flexible 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 on the fine-scale grid.

The main difference between methods LTQM_grid_C/LTQM_grid_V and LT_grid is the ordering of the data used to construct the simple regressions, 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) (Seber and Lee, 2012), 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.

Using rank-ordered data results in a simple form of quantile mapping, but, in contrast to EQM, the quantile map between WRF and GHCND station data is modeled with a *linear* regression line. EQM is more flexible, as first quantiles of observed and station data are typically approximated using linear interpolation or local weighted least squares regression, and then the resulting quantile map is approximated via linear or spline interpolation (Gudmundsson, 2016). 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).

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, rather than station level.

LT_grid 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). Data were sorted in temporal order. Twelve LT functions (one for each month) were constructed for at each GHCND station location (2).month

$$TMAX_{station,i,m} = \beta_{0,i,m} + \beta_{1,i,m} \times WRF_{IDW,i,m} \tag{2}$$

In (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.

The exponential covariance function was used for all Bayesian kriging models. Prior distributions for covariance function parameters were selected based on recommendations in (Banerjee et al., 2004) and inspection of empirical variograms. Empirical variograms of estimated monthly slopes and intercept showed some degree of spatial autocorrelation, although the association was stronger in cold-season compared to warm-season months. We used non-informative priors for the intercept (β_0), the effective range (ϕ), partial sill (σ^2), and nugget (τ^2):

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

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 a some degree of uncertainty surrounding estimates of covariance function parameters (Pilz and Spöck, 2008). Finally, interpolated WRF simulations on the fine-scale grid were bias-corrected on grid-cell by grid-cell, using the corresponding kriged slope and intercept parameter estimates (3):

$$TMAX^*_{i,m} = \tilde{\beta}_{0,i,m} + \tilde{\beta}_{1,i,m} \times WRF_{1km-interp,i,m}. \quad (3)$$

In (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.

LTQM_grid_V and LTQM_grid_C 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, nearest WRF grid-cell values to GHCND station locations were the predictor variables and GHCND station data were the predictands, similar to the approach of (Berg et al., 2012), who applied rank-order regression to bias-correct temperature and precipitation simulations. Berg et al. (2012) 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 (4).

$$TMAX_{i,m} = \beta_{0,i,m} + \beta_{1,i,m} \times WRF_{NN_{i,m}}. \quad (4)$$

In (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 subtle, but important difference between 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 (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 (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. 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.

Examples of downscaled, bias-corrected data products over the study area for selected methods are shown in Supplementary Material.

2.3 Measures of performance and validation

2.3.1 Validation

To gain insight into the downscaling ability of each of the six methods, we used two calibration periods. Bias-correction was applied to 1980-2014 WRF simulations using 1980-2014 GHCND station data. In addition, 1990-2014 WRF simulations were bias-corrected with the 1980-1989 subset of 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").

Bias-corrected WRF data should exhibit day-to-day, as well as distributional, correspondence to GHCND station data. Thus, we chose performance metrics that 1) quantify daily discrepancies and 2) distributional similarity between WRF and GHCND station data. The root-mean-square prediction error (RMSE) and Perkins skill score (PSS) (Perkins et al., 2007) quantify daily errors and distributional similarity, 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 (Perkins et al., 2007). PSS is calculated by summing minimum densities of overlapping bins of discrete histograms of simulated and observed data. PSS is not influenced much by outliers, but it is sensitive to bin size (Perkins et al., 2007). However, large daily discrepancies between simulated and observed data influence RMSE.

Because our goal was to create a continuous, 1km gridded dataset over the study area, the ability of methods to bias-correct WRF simulations at locations where stations are not present is important to assess. Therefore, we also implemented a five-fold spatial cross-validation, where, 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.

2.3.2 Spatial cross-validation

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 $k \neq i$ folds 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 and 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.

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 (Perkins et al., 2007). 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 $K = 5$.

2.4 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). Based on our own observations of WRF simulations and previous work (Huang et al., 2020), WRF simulations of TMAX exhibit larger cold biases in winter and early spring than in summer and early fall, so we controlled for monthly variation in performance metrics. We also controlled for whether or not elevation was accounted for in downscaling to help disentangle the effects of downscaling and bias-correction on performance metrics. We also controlled for whether or not elevation was accounted for in downscaling via lapse rates to help disentangle the effects of downscaling and bias-correction on performance metrics. Finally,

we controlled for type of GHCND station data (1980-1989 or 1980-2014) that was used to bias-correct WRF simulations. We used 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 models can also help expand knowledge of more complex relationships among performance metrics, the six methods, and controlling variables (described below).

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 data calibration set ("1980-1989") or whether 1980-2014 WRF simulations were bias-corrected with the 1980-2014 GHCND time series ("1980-2014").

In addition, the initial full model fits included sensible two-way interactions: *Month* $\times$ *Method*, *Elevation*$\times$*Method*$\times$*Bias_correction_years*, *Elevation*$\times$*Method*, *Elevation*$\times$*Bias_correction_years*, and *Bias_correction_years* $\times$ *Method*. 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` (Lenth, 2020). Pairwise comparisons were performed using the with the Bonferroni correction for multiple comparisons. We also calculated η^2 for all effects in the final models for PSS and RMSE. η^2quantifies 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 (Levine and Hullett, 2002; Muller and Peterson, 1984). 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:

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

3 Results

3.1 Overall performance

Raw WRF interpolations at GHCND station locations exhibited a cold bias, and the bias was most pronounced in months 12, 1, 2, 3, and 4 (Figure 3). Generally, mean RMSE varied little among methods, ranging between 3.1 - 3.5 while mean PSS ranged between 0.94 - 0.96). 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.

Mean RMSE and PSS improved when bias-correction was based on the 1980-2014 GHCND dataset (and the correction was applied to 1980-2014 WRF data) compared to when bias-correction was based on the 1980-1989 GHCND subset (and the correction was applied to 1990-2014 WRF simulations) (Figures 4 (a) and (b), respectively). Generally, when the effect of elevation was accounted for during the downscaling step, mean RMSE decreased (Figure 4 (a)), but *Elevation* did not have an appreciable impact on mean PSS (Figure 4 (b)). In addition, performance metrics for all methods exhibited

considerable monthly variation: both mean monthly RMSE and PSS were worse in months 11, 12, and 1-4 compared to months 5-10 (Figures 5 (a) and (b)), although monthly variation was more pronounced for RMSE than PSS. There was no consistent relationship between low RMSE and high PSS.

Overall, methods LT_grid and LTQM_grid_V performed best and worst, respectively, in terms of mean RMSE (Figure 4 (a)), while methods EQM_grid and LTQM_grid_V performed best and worst, respectively in terms of mean PSS (Figure 4 (b)).

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* (Table 3; See Appendix, Table 5 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* (Table 4, see Appendix, Table 6 for the full model ANOVA). In contrast to the model for RMSE, the effect of *Elevation* was not significant in the full model for PSS.

3.2 Statistical analysis of error metrics

Due to the significance of interaction effects as well as main effects, main effects will be discussed in the context of interactions. Results for pairwise contrasts for each interaction term present in RMSE and PSS ANOVA models are shown in Supplementary Material.

3.2.1 *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 1). The large η^2 for *Month* indicates that *Month* was overwhelmingly the most important variable in the model (despite the statistical significance of the interaction *Month* × *Method*) and means that RMSE varied substantially by Month. Indeed, the monthly pattern of RMSE was consistent for all methods (Figure 6). The interaction plot shows that marginal mean RMSE of all methods was greater (3.2-4.2°C) in months 1,2,3,4,11, and 12 compared to months 5-10 (2.5-3°C) (Figure 6). Overall, marginal mean RMSE of LT_grid was lower than those of all other methods, and for months 2, 3, 4, 5, 6, 11, and 12, results for pairwise contrasts indicated it was significantly lower than RMSE of all other methods.

PSS In contrast to RMSE results, 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) in the model for PSS. This means that PSS varied more among the six methods, rather than among months (Figure 7). Specifically, marginal mean PSS for EQM_IDW, EQM_krig, and EQM_grid varied slightly between 0.92 and 0.95, regardless of month (Figure 7); however, marginal mean 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 (7). Marginal mean PSS for LT_grid followed a similar pattern as LTQM_grid_V and LTQM_grid_C in months 1-4, but in months 5-10, its marginal mean PSS was lower than that of all other methods, ranging between 0.90 and 0.91. Month for month, results for pairwise comparisons showed marginal mean PSS of LT_grid was signficantly lower than that of all other methods (Figure 7).

3.2.2 *Bias_correction_years* and *Bias_correction_years* × *Method*

RMSE 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*. RMSE

was overall lower when bias-correction was based on the 1980-2014 GHCND dataset compared to the 1980-1989 GHCND subset, although there were slight differences among methods. Marginal mean RMSE ranged between 3.18 and 3.57 when bias-correction was based on the 1989-1989 GHCND subset but ranged between 3.15-3.25 when bias-correction was based on the 1980-2014 (Figure 8). Marginal mean RMSE of LT_grid and LTQM_grid_V were overall lowest and highest, respectively, regardless of whether the 1980-2014 or 1989-1989 GHCND dataset was used for bias-correction. However, marginal mean RMSE of LT_grid was significantly lower (3.18) that of all other methods (3.3-3.56) when bias-correction was based on the 1980-1989 GHCND dataset (Figure 8). When bias-correction was applied using the 1980-2014 GHCND dataset marginal mean RMSE of LTQM_grid_V was significantly greater (3.33) than marginal mean RMSE of all other methods (3.15-3.25) (Figure 8). Finally, it is important to note that η^2 values for *Bias_correction_years* and *Bias_correction_years* $\times$ *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* $\times$ *Method*.

PSS Mean PSS generally increased when the 1980-2014, as compared to the 1980-1989 GHCND dataset, was used for bias-correction. However, the amount of increase varied among methods. In particular, the interaction *Bias_correction_years*$\times$*Method* was evident for methods LTQM_grid_C and LTQM_grid_V; marginal mean PSS for LTQM_grid_C and LTQM_grid_V were nearly identical and were significantly greater than that of all other methods only when the 1980-1989 GHCND dataset was used for bias correction (Figure 9). However, when 1980-2014 GHCND dataset was used for bias-correction, marginal mean PSS of EQM_IDW, EQM_krig, and EQM_grid was greater than that of LTQM_grid_C , LTQM_grid_V, and LT_grid (Figure 9). In contrast to results for RMSE, LT_grid performed worst overall; marginal mean PSS LT_grid was significantly lower than that of all other methods, regardless of which GHCND dataset was used for bias-correction (Figure 9). Similar to the results for RMSE, the main effect *Bias_correction_years* and interaction *Bias_correction_years*$\times$*Method* were comparatively less influential in the model. η^2 values for *Bias_correction_years* and interaction *Bias_correction_years*$\times$*Method* (0.09 and 0.14, respectively) were lower than both η^2 values of *Method* and the interaction *Month* $\times$ *Method* (η^2 = 0.43 and 0.28, respectively) (Table 2).

3.2.3 *Elevation*

RMSE Generally, RMSE decreased when elevation was accounted for compared to when it was not (Figure 10). η^2 for *Elevation* was nearly 19 times larger than that of *Elevation* $\times$ *Method* (η^2 = 0.013 and 0.00069, respectively; Table 1), indicating that the main effect of *Elevation* was more important in the RMSE ANOVA model than the interaction term. Additionally, results for pairwise contrasts showed that marginal mean RMSE of method LT_grid was significantly less than, and marginal mean RMSE of method 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 6), and *Elevation* did not have any appreciable effect on PSS (Figure 11).

4 Discussion

In this study, we developed six novel strategies for high-resolution downscaling and bias-correction of daily historical TMAX simulations from a regional climate model,

where bias-correction was based solely on station data. Although performances of all methods appeared similar, there were statistically significant differences in performance even after accounting for monthly variation, whether or not elevation was incorporated during the downscaling step, and which GHCND dataset (1980-2014 or 190-1989) was used for bias-correction. We found that most of the variation in performance among methods was due to the bias-correction technique, rather than interpolation technique, implemented in each of the six methods.

Generally, RMSE, and to a lesser degree, PSS, were better in warm-season months and worse in cold-season months, which is likely due to the pronounced cold bias in raw WRF data during winter and early spring months. EQM (EQM_grid, EQM_krig, and EQM_IDW) outperformed rank-ordered regression (LTQM_grid_C and LTQM_grid_-V) in improving PSS when the 1980-2014 GHCND dataset was used for bias-correction (and the correction was applied to 1980-2014 WRF simulations). However, the converse was true when the 1980-1989 GHCND dataset was used for bias-correction and the correction was applied to 1990-2014 WRF simulations. The bias-variance tradeoff, a well-known concept in statistical learning (Friedman et al., 2001), can help to explain this result. Simple statistical methods, such as linear models, have high bias but low variance, while highly flexible models have low bias but high variance (Friedman et al., 2001). Highly flexible models result in low training errors but are less able to generalize to new, unseen data, which is due to overfitting (Friedman et al., 2001). The EQM transfer function will be nearly perfect if observational and simulated data of the same time period are used. When that transfer function is subsequently used to bias-correct simulated data from the same time period, the applied correction will, by definition, adjust simulated quantiles to closely match those of observed quantiles. Since EQM is a flexible bias-correction method, it is not altogether surprising that it performed very well when bias-correction was based on the 1980-2014 GHCND station dataset, and the correction was applied to the 1980-2014 WRF dataset. For the bias-correction of historical simulations, bias-correction techniques such as EQM may improve PSS to a greater degree than ranked-ordered regression. However, a simple technique, ranked-order, regression may be better suited for correcting future projections, because the transfer function is more generalizable than that of EQM.

Incorporating elevation during interpolation steps in all of the methods was associated with improved RMSE but had no significant effect on PSS. In our study, adjusting temperature with lapse rates decreased the day-to-day discrepancies between simulated and observed data (improving RMSE). However, it is likely that quantile-mapping bias-correction techniques had a much greater influence on PSS than the adjustment provided by lapse rates, explaining why elevation did not appreciably improve PSS.

In addition, we found that no one method could concomitantly minimize RMSE and maximize PSS, which suggests that correcting overall distributional discrepancies as well as daily discrepancies between simulated and observed data is a challenging task. Maximizing PSS is achieved with quantile-mapping techniques, which works by matching the quantiles of 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 techniques such as EQM (EQM_grid, EQM_krig, and EQM_IDW) and rank-ordered regression (LTQM_grid_V and LTQM_grid_C) improve PSS but not necessarily RMSE, whereas temporally-ordered linear regression (LT_-grid) improve RMSE but not necessarily PSS. While the objectives of minimizing RMSE and maximizing PSS are not mutually exclusive, they may be difficult to attain concomitantly in practice.

5 Conclusion

The six high-resolution downscaling and bias-correction methods we presented in this study are efficient, easy to implement, and depending on the method, result in substantially improved RMSE and PSS compared to uncorrected WRF simulations. In addition, we presented methods for constructing full-coverage climate products in which station data are leveraged for bias-correction. Although we applied these methods to historical (1980-2014) daily maximum temperature simulations, most methods are suitable for future climate projections and any modeled temperature variable (minimum, maximum, or average).

References

Banerjee, S., B. P. Carlin, and A. E. Gelfand, 2004: *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC.

Barry, R. G., 1992: *Mountain weather and climate*. Psychology Press.

Behnke, R., S. Vavrus, A. Allstadt, T. Albright, W. E. Thogmartin, and V. C. Radeloff, 2016: Evaluation of downscaled, gridded climate data for the conterminous united states. *ECOAP*, **26 (5)**, 1338–1351.

Bennett, J. C., M. R. Grose, S. P. Corney, C. J. White, G. K. Holz, J. J. Katzfey, D. A. Post, and N. L. Bindoff, 2014: Performance of an empirical bias-correction of a high-resolution climate dataset. *International Journal of Climatology*, **34 (7)**, 2189–2204.

Berg, P., H. Feldmann, and H.-J. Panitz, 2012: Bias correction of high resolution regional climate model data. *J. Hydrol.*, **448**, 80–92.

Bishop, D. A., and C. M. Beier, 2013: Assessing uncertainty in high-resolution spatial climate data across the us northeast. *PloS one*, **8 (8)**, e70 260.

Caldwell, P., H.-N. S. Chin, D. C. Bader, and G. Bala, 2009: Evaluation of a WRF dynamical downscaling simulation over California. *CCH*, **95 (3-4)**, 499–521.

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

Cannon, A. J., S. R. Sobie, and T. Q. Murdock, 2015: Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *J. Climate*, **28 (17)**, 6938–6959.

Daly, C., 2006: Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **26 (6)**, 707–721.

Daly, C., G. Taylor, W. Gibson, T. Parzybok, G. Johnson, and P. Pasteris, 2000: High-quality spatial climate data sets for the united states and beyond. *Transactions of the ASAE*, **43 (6)**, 1957.

Dee, D. P., and Coauthors, 2011: The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, **137 (656)**, 553–597.

Durre, I., M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, 2010: Comprehensive automated quality assurance of daily surface observations. *JAMC*, **49 (8)**, 1615–1633.

Ekström, M., M. R. Grose, and P. H. Whetton, 2015: An appraisal of downscaling methods used in climate change research. *Climate Change*, **6 (3)**, 301–319.

Engen-Skaugen, T., 2007: Refinement of dynamically downscaled precipitation and temperature scenarios. *Climatic Change*, **84 (3-4)**, 365–382.

Fang, G., J. Yang, Y. Chen, and C. Zammit, 2015: Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in china. *HESS*, **19 (6)**, 2547–2559.

Feser, F., B. Rockel, H. von Storch, J. Winterfeldt, and M. Zahn, 2011: Regional climate models add value to global model data: a review and selected examples. *BAMS*, **92 (9)**, 1181–1192.

Finley, A., 2017: spNNGP. URL https://cran.r-project.org/web/packages/spNNGP/spNNGP.pdf, Online.

Finley, A. O., A. Datta, B. D. Cook, D. C. Morton, H. E. Andersen, and S. Banerjee, 2019: Efficient algorithms for bayesian nearest neighbor gaussian processes. *J. Comput. Graph Stat.*, 1–14.

Flint, L. E., and A. L. Flint, 2012: Downscaling future climate scenarios to fine scales for hydrologic and ecological modeling and analysis. *Ecol. Process.*, **1 (1)**, 2.

Franklin, J., F. W. Davis, M. Ikegami, A. D. Syphard, L. E. Flint, A. L. Flint, and L. Hannah, 2013: Modeling plant species distributions under future climates: how fine scale do climate projections need to be? *GLOB*, **19 (2)**, 473–483.

Fridley, J. D., 2009: Downscaling climate over complex terrain: high finescale (< 1000 m) spatial variation of near-ground temperatures in a montane forested landscape (great smoky mountains). *Journal of Applied Meteorology and Climatology*, **48 (5)**, 1033–1049.

Friedman, J., T. Hastie, and R. Tibshirani, 2001: *The elements of statistical learning*, Vol. 1. Springer series in statistics New York.

Giorgi, F., C. Jones, G. R. Asrar, and Coauthors, 2009: Addressing climate information needs at the regional level: the cordex framework. *WMO Bull.*, **58 (3)**, 175.

Gräler, B., E. Pebesma, and G. Heuvelink, 2016: Spatio-temporal interpolation using gstat. *The R Journal*, **8**, 204–218, URL https://journal.r-project.org/archive/2016/RJ-2016-014/index.html.

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

Haas, R., and J. G. Pinto, 2012: A combined statistical and dynamical approach for downscaling large-scale footprints of european windstorms. *Geophysical research letters*, **39 (23)**.

Han, Z., Y. Shi, J. Wu, Y. Xu, and B. Zhou, 2019: 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*, **58 (11)**, 2387–2403.

Hansen, J. W., 2005: Integrating seasonal climate prediction and agricultural models for insights into agricultural practice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360 (1463)**, 2037–2047.

Hanssen-Bauer, I., C. Achberger, R. Benestad, D. Chen, and E. Førland, 2005: Statistical downscaling of climate scenarios over Scandinavia. *Climate Research*, **29 (3)**, 255–268.

Hay, L. E., R. L. Wilby, and G. H. Leavesley, 2000: A comparison of delta change and downscaled gcm scenarios for three mountainous basins in the united states. *J. Am. Water. Resour. As.*, **36 (2)**, 387–397.

Hayhoe, K., and Coauthors, 2008: Regional climate change projections for the northeast usa. *Mitigation and Adaptation Strategies for Global Change*, **13 (5-6)**, 425–436.

Hofstra, N., M. Haylock, M. New, P. Jones, and C. Frei, 2008: Comparison of six methods for the interpolation of daily, european climate data. *Journal of Geophysical Research: Atmospheres*, **113 (D21)**.

Holden, Z. A., J. T. Abatzoglou, C. H. Luce, and L. S. Baggett, 2011: Empirical downscaling of daily minimum air temperature at very fine resolutions in complex terrain. *Agr. Forest Meteorol.*, **151 (8)**, 1066–1073.

Huang, H., J. M. Winter, E. C. Osterberg, J. Hanrahan, C. L. Bruyère, P. Clemins, and B. Beckage, 2020: Simulating precipitation and temperature in the lake champlain basin using a regional climate model: limitations and uncertainties. *Climate Dynamics*, **54 (1-2)**, 69–84.

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

Huth, R., 1999: Statistical downscaling in central europe: evaluation of methods and potential predictors. *Climate Research*, **13 (2)**, 91–101.

Kettle, H., and R. Thompson, 2004: Statistical downscaling in european mountains: verification of reconstructed air temperature. *Climate Research*, **26 (2)**, 97–112.

Lafon, T., S. Dadson, G. Buys, and C. Prudhomme, 2013: Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *Int. J. Climatol.*, **33 (6)**, 1367–1381.

Leander, R., and T. A. Buishand, 2007: Resampling of regional climate model output for the simulation of extreme river flows. *J. Hydrol.*, **332 (3-4)**, 487–496.

Lenderink, G., A. Buishand, and W. v. Deursen, 2007: Estimates of future discharges of the river rhine using two scenario methodologies: direct versus delta approach. *Hydrol. Earth. Syst. Sc.*, **11 (3)**, 1145–1159.

Lenth, R., 2020: *emmeans: Estimated Marginal Means, aka Least-Squares Means*. URL https://CRAN.R-project.org/package=emmeans, r package version 1.4.4.

Leung, L. R., L. O. Mearns, F. Giorgi, and R. L. Wilby, 2003: Regional climate research: needs and opportunities. *Bull. Amer. Meteor. Soc.*, **84 (1)**, 89–95.

Levine, T. R., and C. R. Hullett, 2002: Eta squared, partial eta squared, and misreporting of effect size in communication research. *Hum. Commun. Res.*, **28 (4)**, 612–625.

Liston, G. E., and K. Elder, 2006: A meteorological distribution system for high-resolution terrestrial modeling (micromet). *J. Hydrometeor.*, **7 (2)**, 217–234.

Livneh, B., T. J. Bohn, D. W. Pierce, F. Munoz-Arriola, B. Nijssen, R. Vose, D. R. Cayan, and L. Brekke, 2015: A spatially comprehensive, hydrometeorological data set for mexico, the us, and southern canada 1950–2013. *Scientific data*, **2 (1)**, 1–12.

Maraun, D., and Coauthors, 2017: Towards process-informed bias correction of climate change simulations. *Nat. Clim. Chang.*, **7 (11)**, 764.

Maurer, E. P., and P. B. Duffy, 2005: Uncertainty in projections of streamflow changes due to climate change in california. *Geophysical Research Letters*, **32 (3)**.

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

Mejia, J. F., J. Huntington, B. Hatchett, D. Koracin, and R. G. Niswonger, 2012: Linking global climate models to an integrated hydrologic model: using an individual station downscaling approach. *Journal of Contemporary Water Research & Education*, **147 (1)**, 17–27.

Muller, K. E., and B. L. Peterson, 1984: Practical methods for computing power in testing the multivariate general linear hypothesis. *Computational Statistics & Data Analysis*, **2 (2)**, 143–158.

National Oceanic and Atmospheric Administration, 2018: Global historical climatology network daily. NOAA, URL https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND, accessed: 2017-09-30.

Perkins, S., A. Pitman, N. Holbrook, and J. McAneney, 2007: Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J. Climate*, **20 (17)**, 4356–4376.

Peterson, T. C., and R. S. Vose, 1997: An overview of the global historical climatology network temperature database. *Bull. Amer. Meteor. Soc.*, **78 (12)**, 2837–2850.

Piani, C., J. Haerter, and E. Coppola, 2010: Statistical bias correction for daily precipitation in regional climate models over europe. *Theor. Appl. Climatol.*, **99 (1-2)**, 187–192.

Pilz, J., and G. Spöck, 2008: Why do we need and how should we implement bayesian kriging methods. *Stochastic Environmental Research and Risk Assessment*, **22 (5)**, 621–632.

Poggio, L., and A. Gimona, 2015: Downscaling and correction of regional climate models outputs with a hybrid geostatistical approach. *Spatial Statistics*, **14**, 4–21.

R Core Team, 2018: *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing, URL https://www.R-project.org/.

Roberts, D. R., W. H. Wood, and S. J. Marshall, 2019: Assessments of downscaled climate data with a high-resolution weather station network reveal consistent but predictable bias. *Int. J. Climatol.*, **39 (6)**, 3091–3103.

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

Schoof, J. T., and S. C. Pryor, 2001: Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **21 (7)**, 773–790.

Seber, G. A., and A. J. Lee, 2012: *Linear regression analysis*, Vol. 329. John Wiley & Sons.

Shrestha, M., S. C. Acharya, and P. K. Shrestha, 2017: Bias correction of climate models for hydrological modelling–are simple methods still useful? *Meteorological Applications*, **24 (3)**, 531–539.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, Z. Liu, J. Berner, and X. Huang, 2019: A description of the advanced research WRF model. URL https://opensky.ucar.edu/islandora/object/opensky:2898, accessed: 2019-03-04.

Teutschbein, C., and J. Seibert, 2012: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *J. Hydrol.*, **456**, 12–29.

Themeßl, M. J., A. Gobiet, and A. Leuprecht, 2011: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *Int. J. Climatol.*, **31 (10)**, 1530–1544.

Thornton, P. E., M. M. Thornton, B. W. Mayer, N. Wilhelmi, Y. Wei, R. Devarakonda, and R. Cook, 2012: 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)*.

U.S. Geological Survey, 2019: 3DEP 1/3 arc-second DEM. U.S. Geological Survey, URL https://viewer.nationalmap.gov/basic/.

Vandal, T., E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, 2017: Deepsd: Generating high resolution climate change projections through single image super-resolution. *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*, 1663–1672.

Varouchakis, E., and D. Hristopulos, 2013: Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins. *Environ. Monit. Assess.*, **185 (1)**, 1–19.

Walton, D., and A. Hall, 2018: An assessment of high-resolution gridded temperature datasets over california. *J. Climate*, **31 (10)**, 3789–3810.

Wang, T., A. Hamann, D. L. Spittlehouse, and T. Q. Murdock, 2012: ClimateWNA—high-resolution spatial climate data for western North America. *J. Appl. Meteor. Climatol.*, **51 (1)**, 16–29.

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

Wilby, R. L., S. Charles, E. Zorita, B. Timbal, P. Whetton, and L. Mearns, 2004: Guidelines for use of climate scenarios developed from statistical downscaling methods. *Supporting material of the Intergovernmental Panel on Climate Change, available from the DDC of IPCC TGCIA*, **27**.

Winter, J. M., B. Beckage, G. Bucini, R. M. Horton, and P. J. Clemins, 2016: Development and evaluation of high-resolution climate simulations over the mountainous northeastern united states. *J. Hydrol.*, **17 (3)**, 881–896.

Wood, A. W., L. R. Leung, V. Sridhar, and D. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic change*, **62 (1-3)**, 189–216.

Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier, 2002: Long-range experimental hydrologic forecasting for the eastern united states. *J. Geophys. Res-atmos.*, **107 (D20)**, ACL–6.

Table 1: η^2 for RMSE final 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: η^2 for PSS final model.

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

Table 3: ANOVA table for RMSE (final model)

	Df	Sum Sq	Mean Sq	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 4: ANOVA table for PSS (final model)

	Df	Sum Sq	Mean Sq	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		

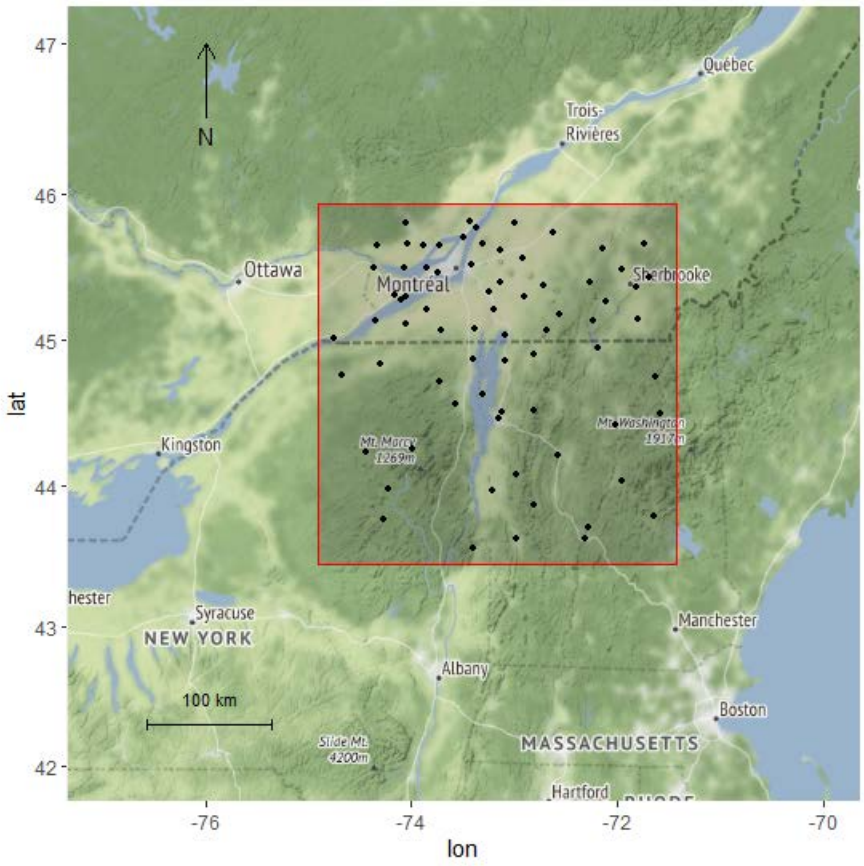

Figure 1: GHCND stations (black points) within the study area (red).

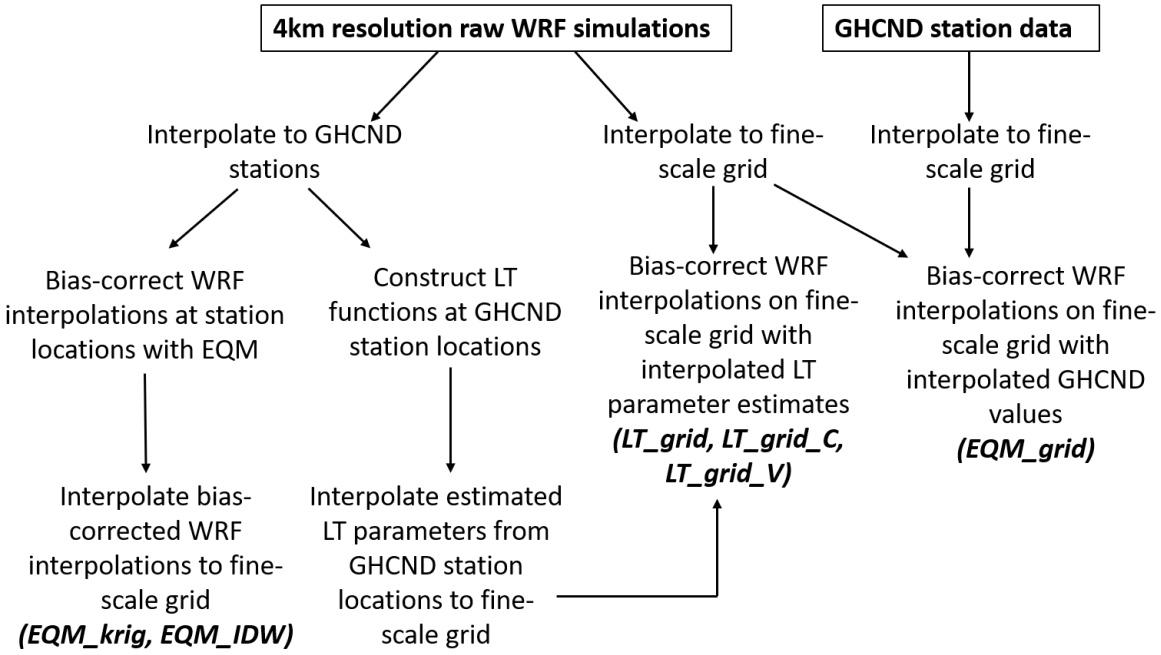

Figure 2: Workflows for the six bias-correction and downscaling methods described in this study. In EQM_IDW, EQM_krig, and EQM_grid, bias-correction was done with EQM. EQM_grid differs with respect to EQM_krig and EQM_IDW in that bias correction was done at the *grid* rather than station level. In LTQM_grid_V and LTQM_grid_C, LT functions were constructed using rank-ordered data, which results in a simple quantile-mapping transfer function. 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.

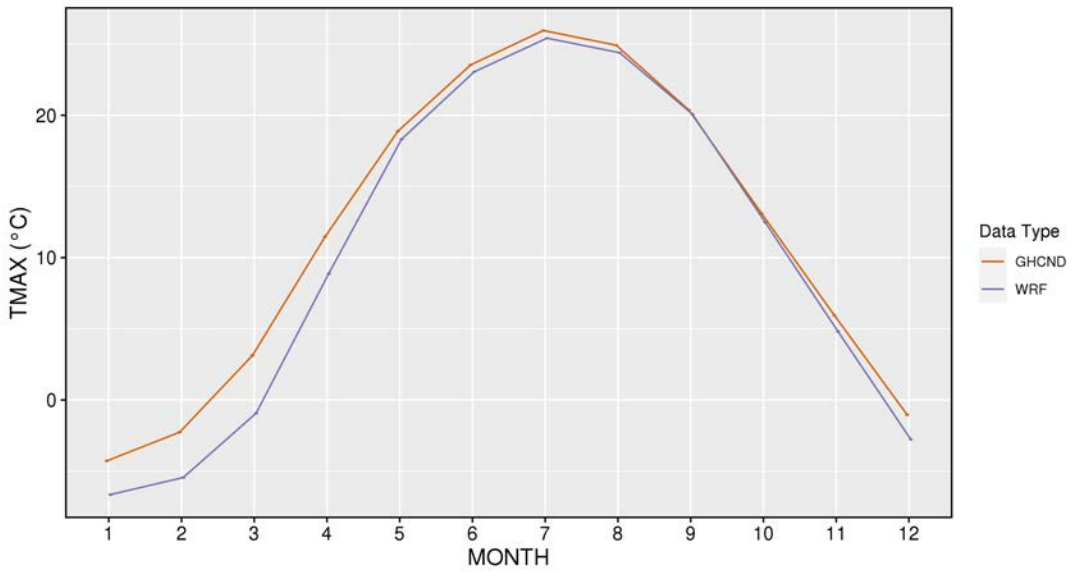

Figure 3: Monthly average TMAX (°C) of WRF interpolations at GHCND locations and GHCND station data from 1980-2014

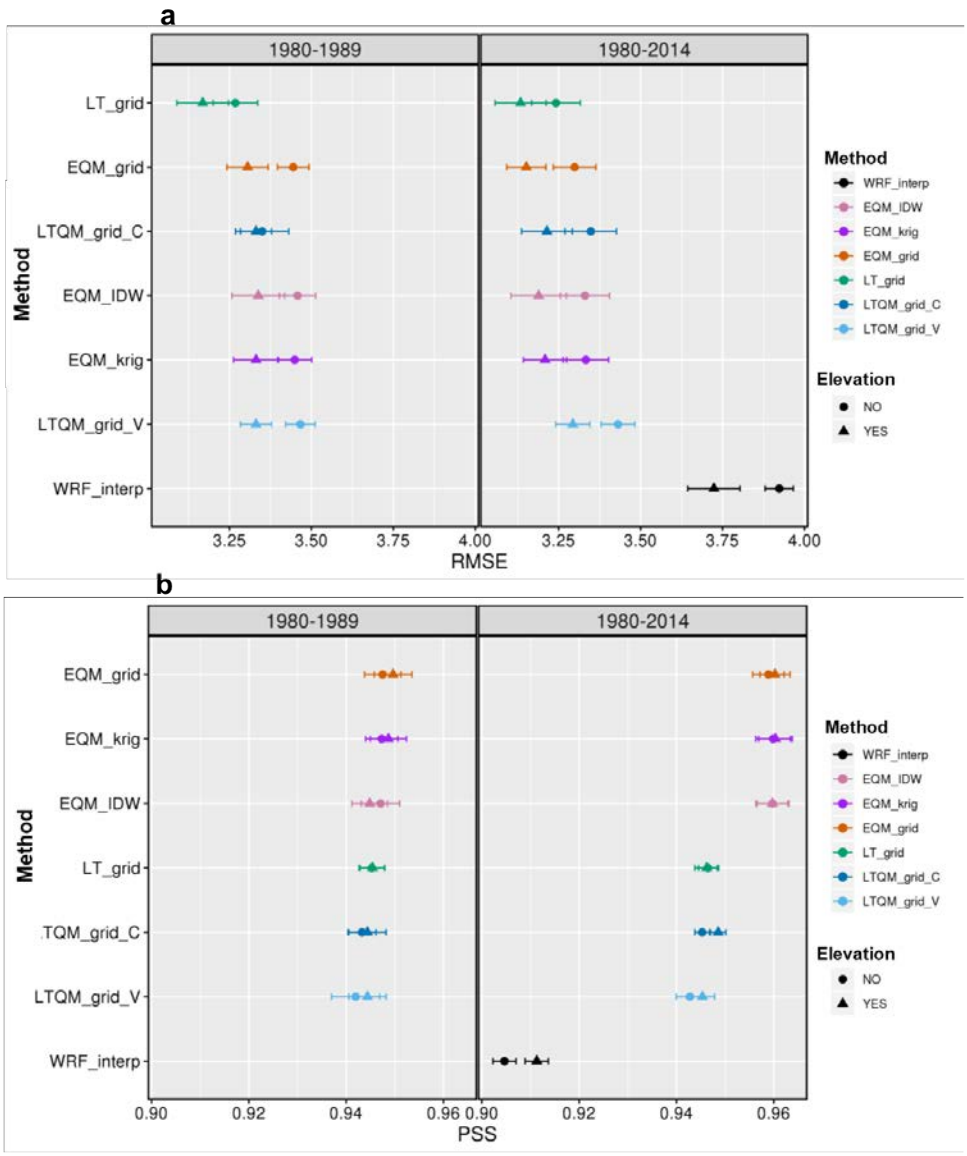

Figure 4: Mean RMSE (a) and PSS (b) by downscaling method and *Bias_correction_years*, where "1980-1989" and "1980-2014" are the GHCND station calibration datasets used to bias-correct 1990-2014 and 1980-2014 WRF simulations, respectively. RMSE and PSS values reflect mean performance metrics prior to linear model fits. 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 to indicate relative improvement of all methods over raw WRF interpolated values.

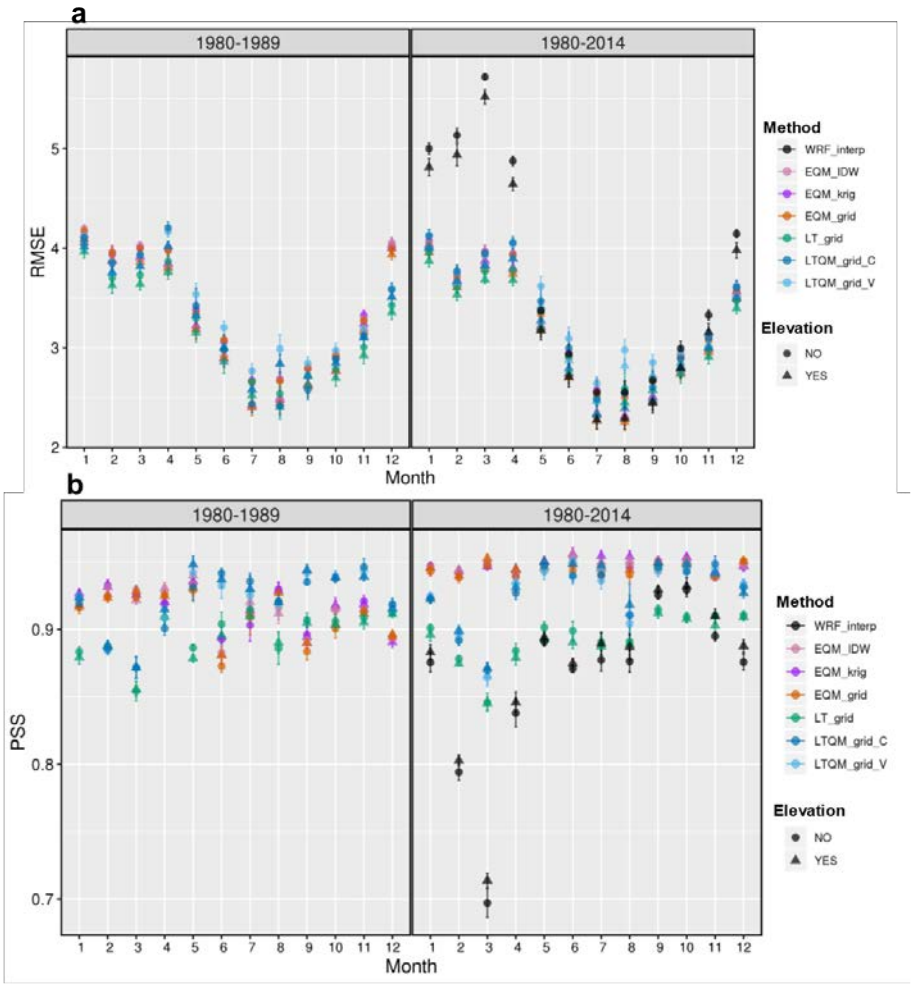

Figure 5: Mean RMSE (a) and PSS (b) by downscaling method, month and *Bias_correction_-years*, where "1980-1989" and "1980-2014" are the GHCND station calibration datasets used to bias-correct 1990-2014 and 1980-2014 WRF simulations, respectively. Error bars represent standard errors over five spatial cross-validation folds. “WRF_interp” denotes the raw WRF simulations interpolated to station locations and are shown to indicate relative improvement of all methods over raw WRF interpolated values.

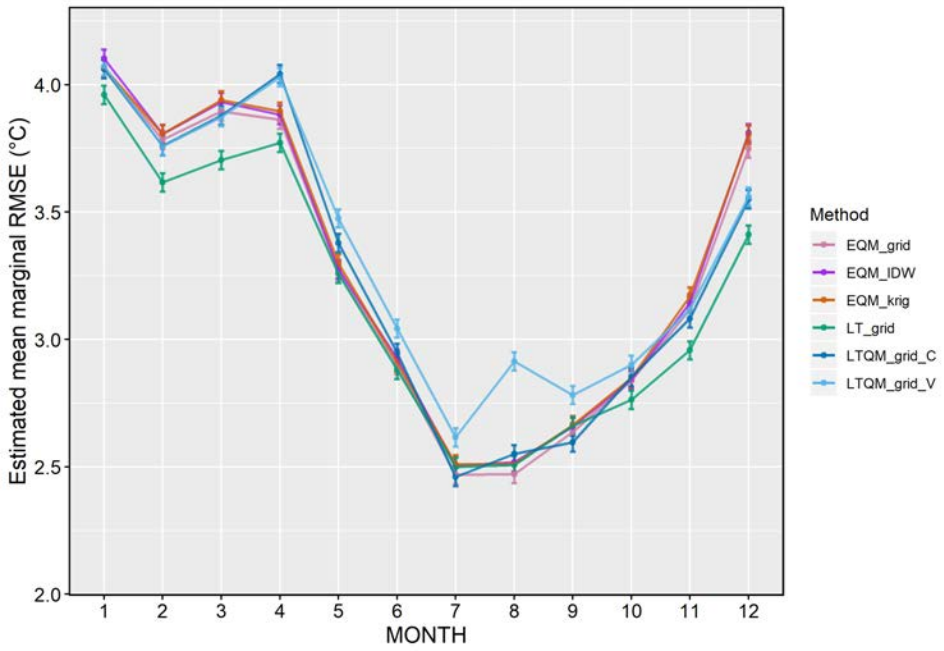

Figure 6: Interaction plot showing marginal mean RMSE by *Method* and *Month*. Error bars represent 95% confidence intervals.

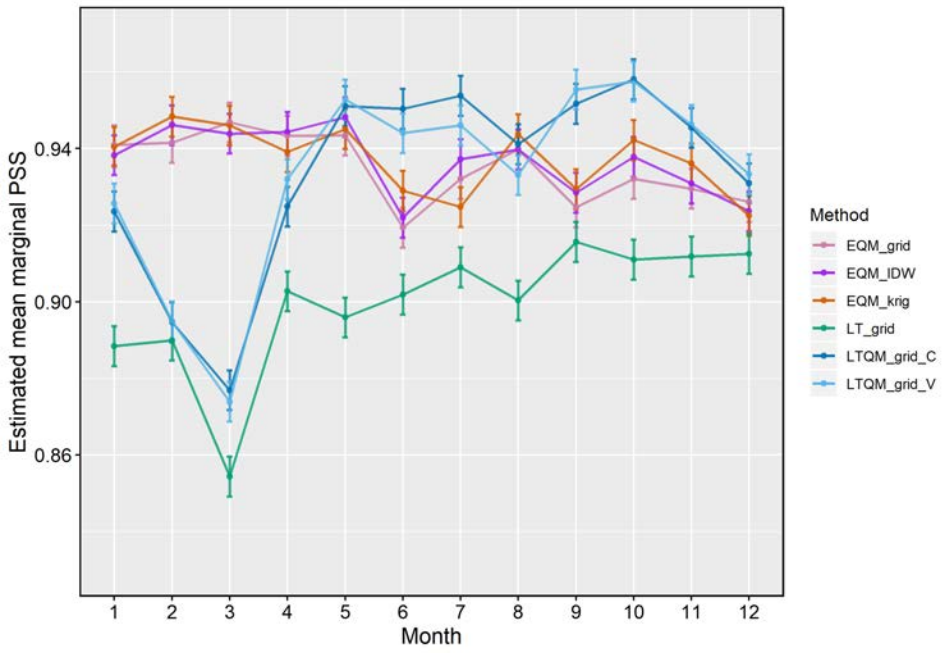

Figure 7: Interaction plot showing marginal means of PSS by *Method* and *Month*. Error bars represent 95% confidence intervals.

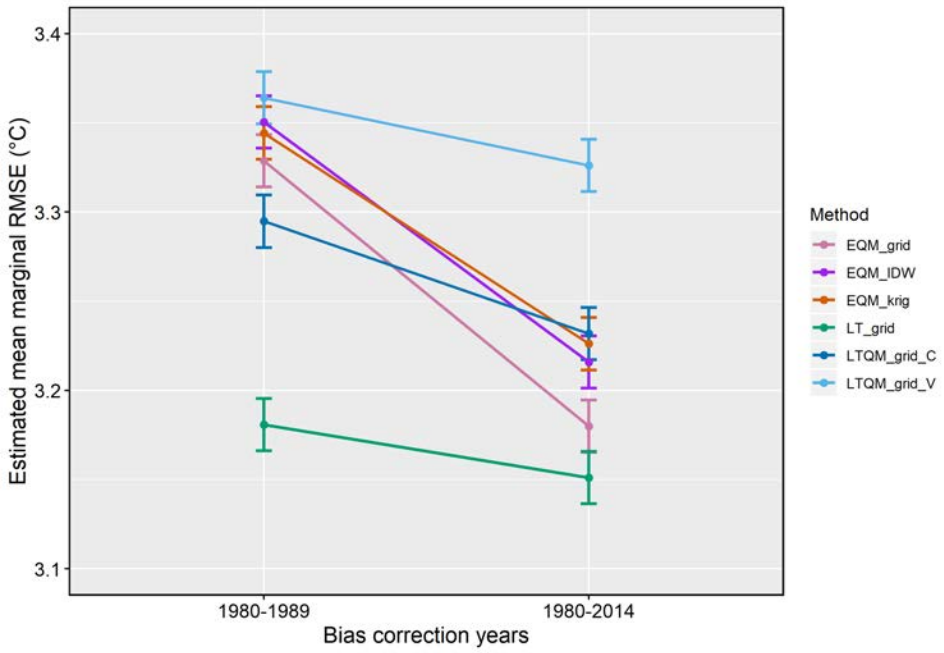

Figure 8: Interaction plot showing predicted marginal mean RMSE by *Method* and *Bias_correction_years* ("1980-1989" and "1980-2014" are the GHCND station datasets used to bias-correct 1990-2014 and 1980-2014 WRF simulations, respectively). Error bars represent 95% confidence intervals.

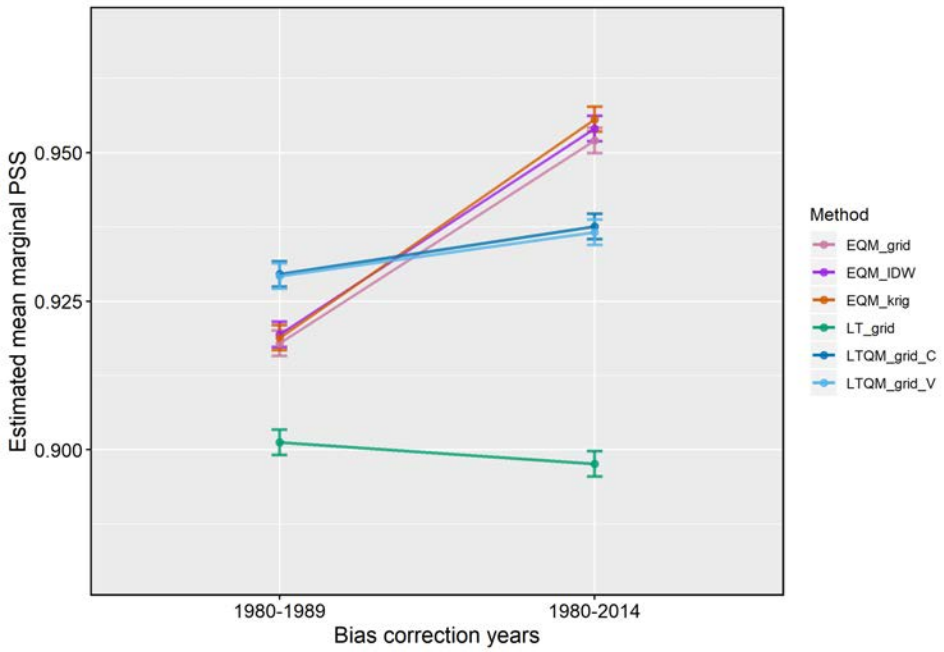

Figure 9: Interaction plot showing predicted marginal mean PSS by *Method* and *Bias_correction_years* ("1980-1989" and "1980-2014" are the GHCND station datasets used to bias-correct 1990-2014 and 1980-2014 WRF simulations, respectively). Error bars represent 95% confidence intervals.

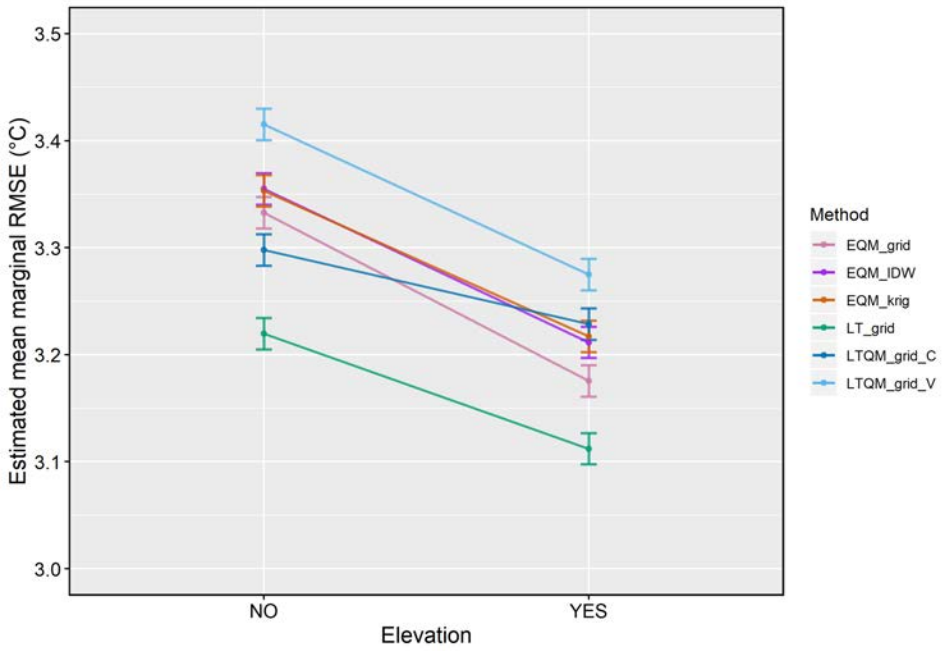

Figure 10: Interaction plot showing predicted marginal mean RMSE by *Method* and *Elevation*. Error bars represent 95% confidence intervals.

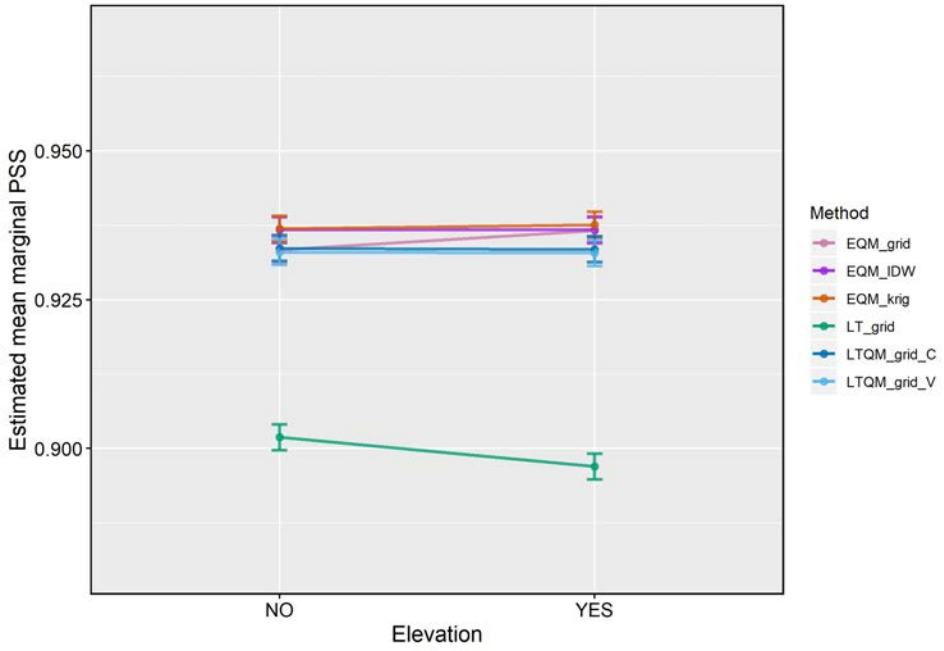

Figure 11: Interaction plot showing marginal mean PSS by *Method* and *Elevation* (results obtained from full model fit). Error bars represent 95% confidence intervals.

6 Appendix

Table 5: ANOVA table for full RMSE model

	Df	Sum Sq	Mean Sq	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 6: ANOVA table for full PSS model

	Df	Sum Sq	Mean Sq	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		

Table 7: A3. 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)

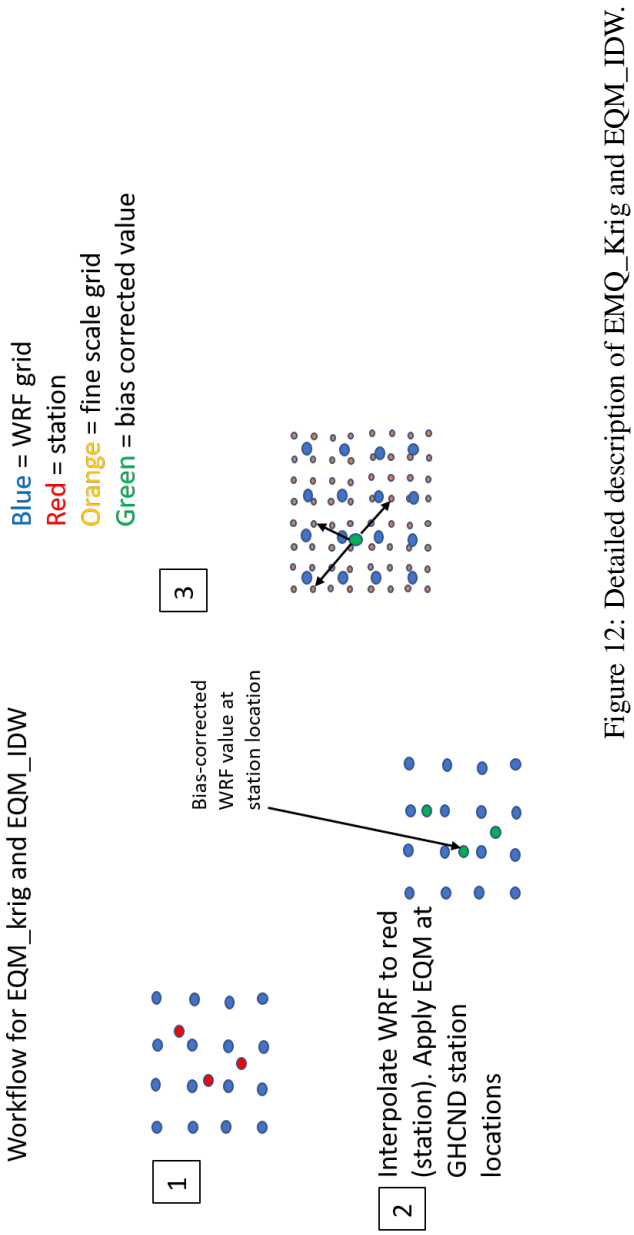

Figure 12: Detailed description of EMQ_Krig and EQM_IDW.

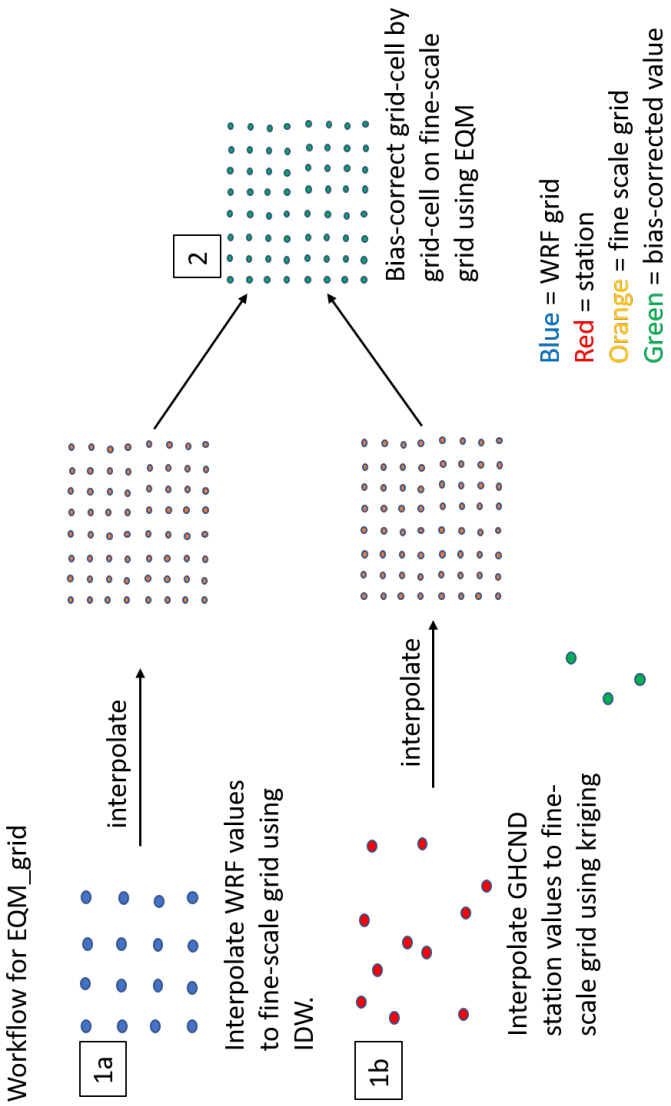

Figure 13: Detailed description of EMQ_grid.

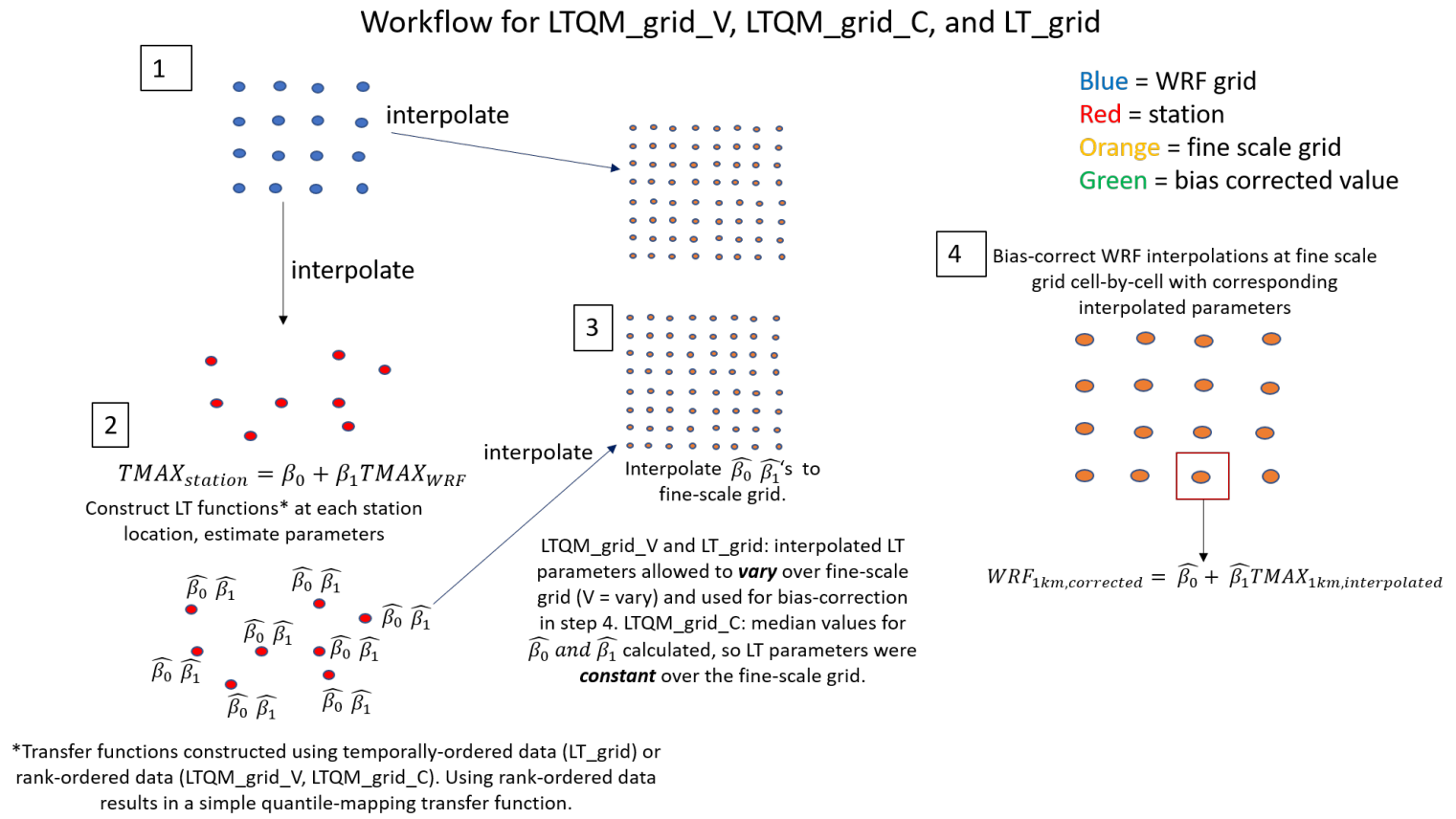

Figure 14: Detailed decriptions of LTQM_grid_V, LTQM_grid_C, and LT_grid.

7 Supplementary Material

1. Table 8. Pairwise contrasts for RMSE ANOVA model: *Month* × *Method*
2. Table 9. Pairwise contrasts for PSS ANOVA model: *Month* × *Method*
3. Table 10. Pairwise contrasts for RMSE ANOVA model: *bias_correction_years* × *Method*
4. Table 11. Pairwise contrasts for PSS ANOVA model: *bias_correction_years* × *Method*
5. Table 12. Pairwise contrasts for RMSE ANOVA model: *Elevation* × *Method*

Table 8: Contrasts for RMSE ANOVA model for interaction *Month* × *Method*

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

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
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

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

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
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

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 9: Contrasts for PSS ANOVA model for interaction: *Month* × Method

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

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
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

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

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
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 10: Contrasts for RMSE ANOVA model for interaction *Bias_correction_years* × *Method*

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 11: Contrasts for PSS model for interaction: *Bias_correction_years* × *Method*

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 12: Contrasts for RMSE model for interaction: *Elevation* × *Method*

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

7.1 Modeling details: 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 (Banerjee et al., 2004):

$$Y|\mu, \theta \sim N_n(\mu 1, \sigma(\theta)),$$

where N_n connotes the N dimensional normal distribution, μ is the constant mean, $\sigma(\theta)_{ii'}$ gives the covariance between $Y(s_i)$ and $Y(s_{i'})$, 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 either monthly slope or intercept 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:

$$\boldsymbol{Y}(s) = \mu(s) + w(s) + \epsilon(s), \tag{5}$$

where $Y(s)$ is the response at location s having a mean structure $\mu(s) = x^T(s)\beta$. The implementation of a full Bayesian spatial model is computationally intensive, due to the inversion of large $(n \times n)$ covariance matrices (Banerjee et al., 2004). To decrease computation time, we instead used a nearest-neighbor Gaussian process model (NNGP), which is computationally more efficient that the full Gaussian process model in (5). In NNGP models, the spatial process is estimated by a realization of the spatial process with its n nearest neighbors (Finley, 2017). The `spNNGP` function from the `spNNGP` package in R constructs an NNGP model (Finley, 2017). 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 (Banerjee et al., 2004). Predictions were made by passing the resulting model fit from `spNNGP` to the `spPredict` function (Finley, 2017), which carries out Bayesian kriging.

Based on exploratory variogram analysis, we used the exponential covariance function for fitting all models.

$$C(t) = \begin{Bmatrix} \tau^2 + \sigma^2 & \text{if} \quad d = 0 \\ \sigma^2 exp(-\phi d) & \text{if} \quad d > 0 \end{Bmatrix}, \tag{6}$$

In (6), $||h|| = d$, 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 $d_0 = \frac{3}{\phi}$.

Prior distributions were selected by following recommendations in (Banerjee et al., 2004). We inspected residual variograms to determine appropriate prior values for the effective range, ϕ. For Bayesian kriging in EQM_grid, priors for the intercept (β_0), the effective range (ϕ), partial sill (σ^2), and nugget (τ^2) were as follows:

$$\begin{aligned} \beta_0 &\sim N(0, 100) \\ \phi &\sim Unif(\frac{3}{D_{max}}, \frac{3}{10}) \\ \sigma^2 &\sim IG(2, 2) \\ \tau^2 &\sim IG(2, 0.1), \end{aligned}$$

where D_{max} was the maximum distance between any two GHCND station locations. The priors for all Bayesian kriging implemented in the LT_grid, LTQM_grid_V, and LTQM_grid_V methods were the same as those for EQM_krig, except we used an $IG(2, 0.02)$ prior for τ^2. All daily models used 5000 MCMC samples with a burn-in of 1250 iterations. Bayesian kriging was implemented as a nearest neighbor Gaussian process (NNGP), which is much more efficient than kriging and more accurate than predictive process models (Finley et al., 2019). All daily NNGP models were fit with the `spNNGP` package in R (Finley, 2017).

7.2 Modeling details: kriging

In methods EQM_grid and EQM_krig daily GHCND station values and bias-corrected WRF values at station locations were kriged to the 1km grid with non-Bayesian kriging. Non-Bayesian kriging can also be understood in the context of Gaussian processes. Suppose that, as in (5), spatial variates $Y = s_1...s_n$ are assumed to follow a multivariate normal distribution (Schabenberger and Gotway, 2017). A general expression for the spatial model is

$$\mathbf{Y} = X\beta + \epsilon, \text{ where } \epsilon \sim N(\mathbf{0}, \Sigma), \tag{7}$$

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 (8) is known as the kriging predictor.

$$E\left[(Y(s_0) - h(y))^2 | y\right] \tag{8}$$

It can be shown (e.g. Schabenberger and Gotway, 2017) 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 is prediction is to be made, and λ_i are weights chosen such that they satisfy the conditions of unbiasedness and minimize the kriging variance (Schabenberger and Gotway, 2017). Unlike Bayesian kriging, the variogram parameters must be estimated from the data.

Based on exploratory variogram analysis, we used the exponential covariance function (6) for all model fits (effective range = 150, partial sill = 15, and nugget= 0.2).

7.3 Inverse distance weighting

Inverse distance weighting is a deterministic interpolation technique in which 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:

$$\begin{aligned} Y(s_0) &= \sum_{i=1}^{n} w_i(s_0) Y(s_i), \\ w_i(s_0) &= \frac{\tilde{w}_i(s_0)}{\sum_{i=1}^{n}}, \\ Tilde w_i(s_0) &= \frac{1}{d(s_i, s_0)}. \end{aligned}$$

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.

7.4 Topographic downscaling

Topographic downscaling is a variation on IDW that is often used for high resolution downscaling (Winter et al., 2016). In contrast to kriging, IDW is deterministic; that is, the size of prediction errors cannot be quantified (Wikle et al., 2019). Topographic downscaling consists of three main steps:

1. Construct historical, empirical relationship between TMAX and elevation using regression;
2. Adjust WRF data to reference elevation (200m) using the estimated elevational lapse rate parameters and use IDW to interpolate to desired locations;
3. Back-transform interpolations using estimated elevational lapse rate parameters.

We used a weight of 2 and 9 nearest neighbors for all IDW interpolation, as suggested in (Winter et al., 2016).

Following methods by (Winter et al., 2016) and (Liston and Elder, 2006), 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 (9):

$$T_{sta} = T_0 - \beta\phi_{sta} - \gamma z_{sta}, \tag{9}$$

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}).

Our estimates for β and γ were -1.43 and -0.0059, respectively (Figure 15). 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 (Winter et al., 2016), as well as the standard elevational lapse rate (6.0°Ckm^{-1}) (Barry, 1992).

WRF projections for TMAX and were translated to reference elevation with (10)

$$T_{WRF,\,ref} = T_{WRF} - \gamma(z_{ref} - z_{WRF}), \tag{10}$$

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 9, 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. Interpolated WRF data were back-transformed to reflect the effect of elevation (11)

$$T_{sta,\,interp} = T_{ref,\,interp} - \gamma(z_{sta} - z_{ref}). \tag{11}$$

In (11) $T_{sta,interp}$ is the elevation-adjusted value for TMAX, $T_{ref,interp}$ is the interpolated value at a station location at reference elevation and z_{sta} and z_{ref} are the GHCND station and reference elevations, respectively. After back- transforming interpolated values at GHCND station locations, we applied empirical quantile mapping (EQM) at each station location.

$$X_{corr,t} = \text{ecdf}^{-1}_{obs,m}(\text{ecdf}_{raw,m}(X_{raw,t})), \tag{12}$$

In (12), $X_{corr,t}$ 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 (13)

$$T_{EQM,\,ref} = T_{EQM} - \gamma(z_{ref} - z_{sta}), \tag{13}$$

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 (11). Finally, the reference-adjusted, bias-corrected WRF interpolations at GHCND station locations were again interpolated to a 1km grid using IDW,

$$
\begin{aligned}
Y(s_0) &= \sum_{i=1}^{n} w_i(s_0) Y(s_i), \\
w_i(s_0) &= \frac{Tildew_i(s_0)}{\sum_{i=1}^{n}}, \\
Tildew_i(s_0) &= \frac{1}{d(s_i, s_0)},
\end{aligned}
$$

where 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 (14)

$$T_{fine,\,interp} = T_{ref,\,interp} - \gamma(z_{fine} - z_{ref}). \tag{14}$$

In (14), $T_{fine,\,interp}$ 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 (11).

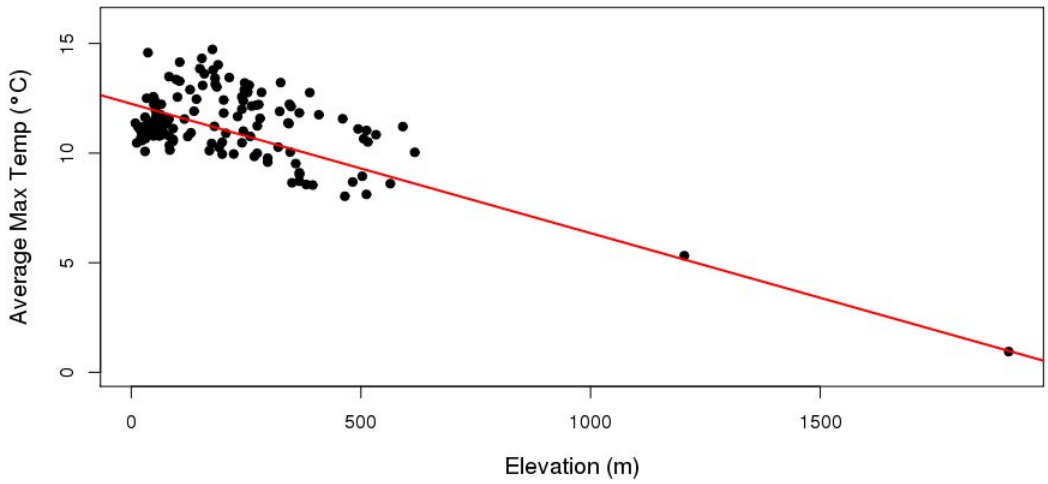

Figure 15: Elevational lapse rate adjustment for TMAX. Note: the elevational lapse rate did not change appreciably with omission of the two high elevation stations.

7.5 Results for sorted RMSE

An alternative metric for PSS is the sorted RMSE. Since PSS is more widely used 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 prior to the calculation. The full linear model included the same variables as the models for RMSE and PSS (Table 14). The final model included the main effects *Month*, *Bias_correction_years*, and *Method* as well as the interaction terms *Month* $\times$ *Method* and *Bias_correction_years* $\times$ *Method* (Table 13).

The results for sorted RMSE were very similar to those of PSS. Generally, SRMSE values were lower when bias correction was based on the 1980-2014 GHCND dataset. LT_grid performed worst overall regardless of whether the 1980-1989 or 1980-2014 GHCND dataset was used for bias-correction (Figures 16 and 17). In contrast to RMSE but similar to PSS, SRMSE exhibited less monthly variation. There was a positive association (0.60) between SRMSE and PSS when bias correction was conducted with the 1980-1989 GHCND dataset, but the correlation was strongly negative (-0.97) when bias correction was done with the 1980-2014 GHCND dataset (Figure 18).

The interaction of *Month* $\times$ *Method* was significant, and the interaction was most apparent for LT_grid. In contrast to other methods, mean SRMSE of LT_grid was significantly greater than that of all other methods in months 1-5 (Figure 19). The interaction plot for *Method* $\times$ *Bias_correction_years* shows that while EQM_IDW, EQM_krig, and EQM_grid performed better when bias corrected 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 (Figure 20). LT_grid performed worst overall, regardless of whether the complete 1980-2014 or 1980-1989 GHCND dataset was used for bias-correction (Figure 20).

Table 13: 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		

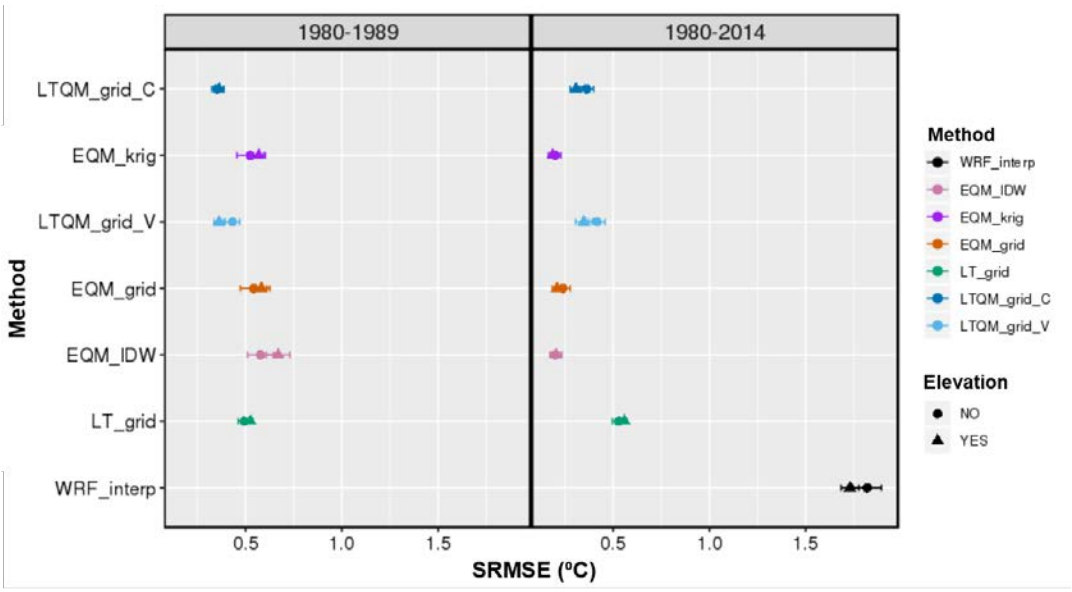

Figure 16: Mean SRMSE by *Method* and *Bias_correction_years*, where "1980-1989" and "1980-2014" denote the GHCND station datasets used to bias-correct 1990-2014 and 1980-2014 WRF simulations, respectively. Error bars represent 95% confidence intervals. "WRF_interp" denotes raw WRF simulations (not bias-corrected) interpolated to station locations and are shown to indicate relative improvement of all methods over raw WRF interpolated values.

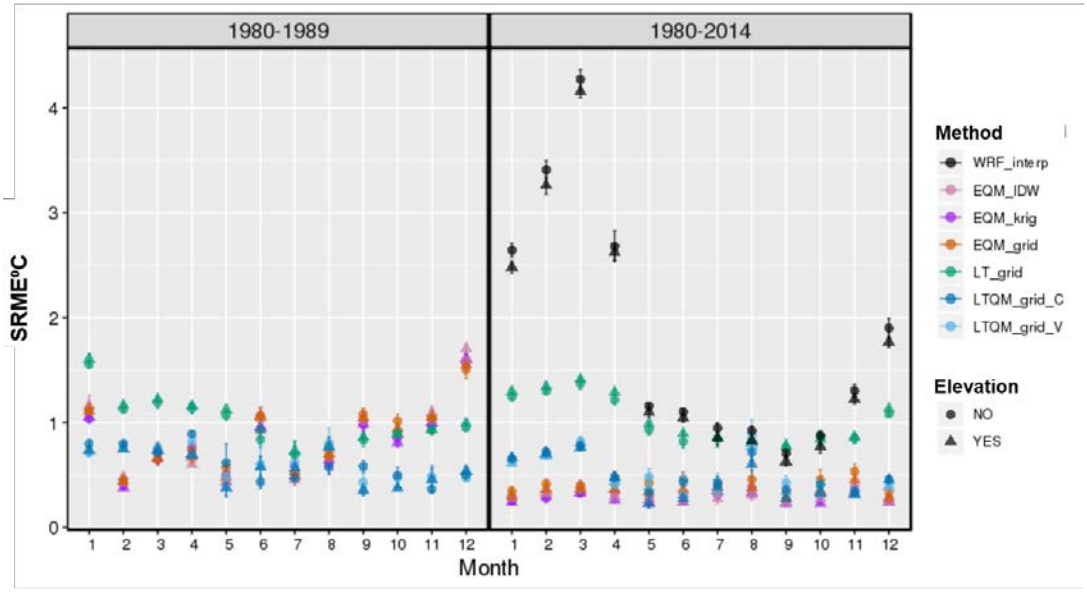

Figure 17: Mean SRMSE by *Method*, *Month*, and *Bias_correction_years*, where "1980-1989" and "1980-2014" denote the GHCND station datasets used to bias-correct 1990-2014 and 1980-2014 WRF simulations, respectively. Error bars represent 95% confidence intervals. "WRF_interp" denotes raw WRF simulations (not bias-corrected) interpolated to station locations and are shown to indicate relative improvement of all methods over raw WRF interpolated values.

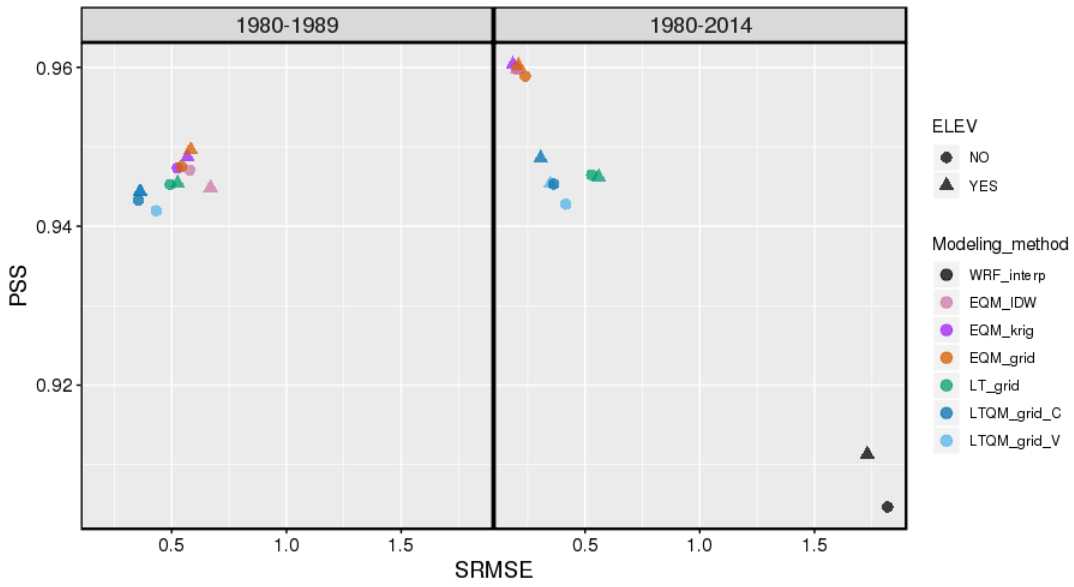

Figure 18: Scatterplot of mean SRMSE and PSS by *Bias_correction_years*, where "1980-1989" and "1980-2014" denote the GHCND station datasets used to bias-correct 1990-2014 and 1980-2014 WRF simulations, respectively. Error bars represent 95% confidence intervals. "WRF_interp" denotes raw WRF simulations (not bias-corrected) interpolated to station locations and are shown to indicate relative improvement of all methods over raw WRF interpolated values.

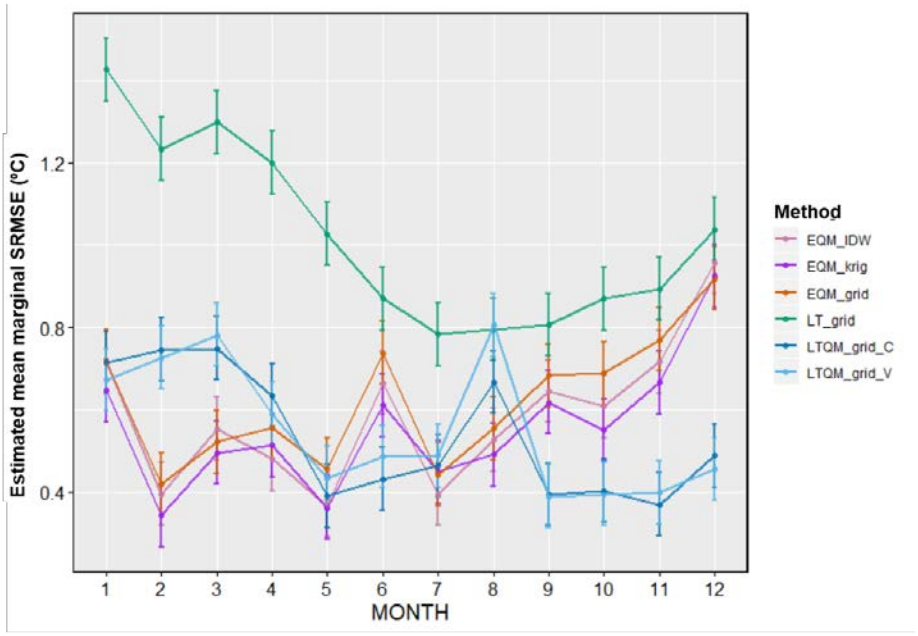

Figure 19: Interaction plot showing estimated mean marginal SRMSE by *Method* and *Month*. Error bars represent 95% confidence intervals.

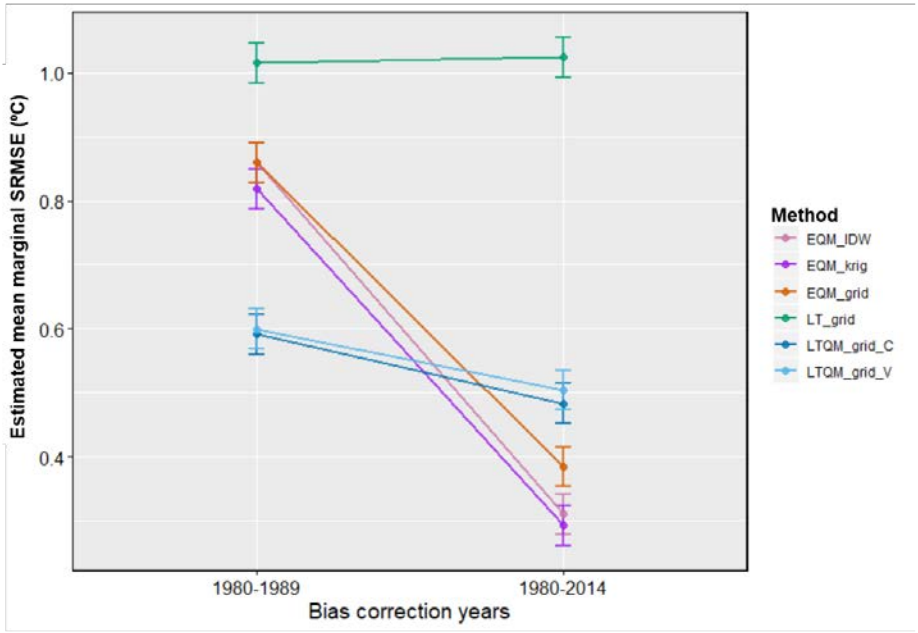

Figure 20: Interaction plot showing estimated mean marginal SRMSE by *Method* and *Bias_correction_years*, where "1980-1989" and "1980-2014" denote the GHCND station datasets used to bias-correct 1990-2014 and 1980-2014 WRF simulations, respectively. Error bars represent 95% confidence intervals. "WRF_interp" denotes raw WRF simulations (not bias-corrected) interpolated to station locations and are shown to indicate relative improvement of all methods over raw WRF interpolated values.

Table 14: 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		

7.6 Downscaling example plots

Figures 21 and 22 show original WRF TMAX simulations and downscaled WRF TMAX (°C) data for August 5, 1982 for methods EQM_IDW and LT_grid, respectively.

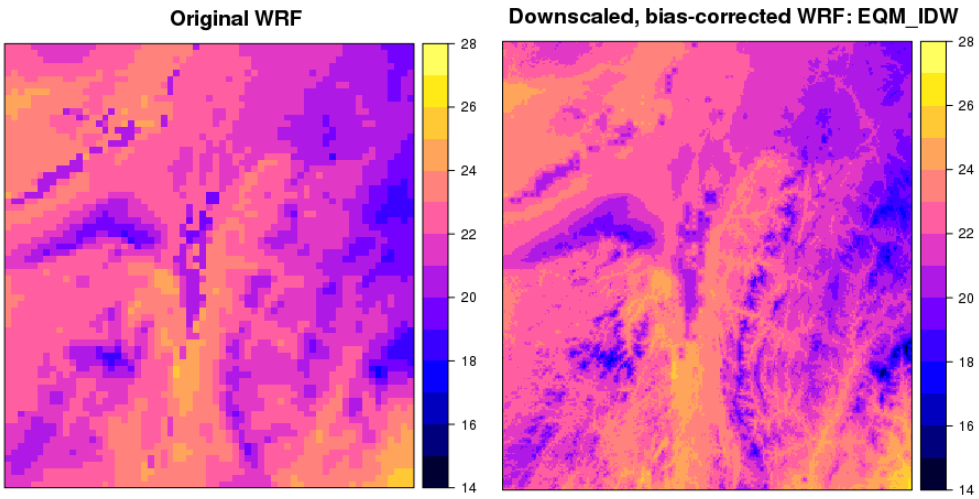

Figure 21: Original WRF simulations for TMAX (°C) and downscaled WRF TMAX (°C) using method EQM_IDW for August 5, 1982.

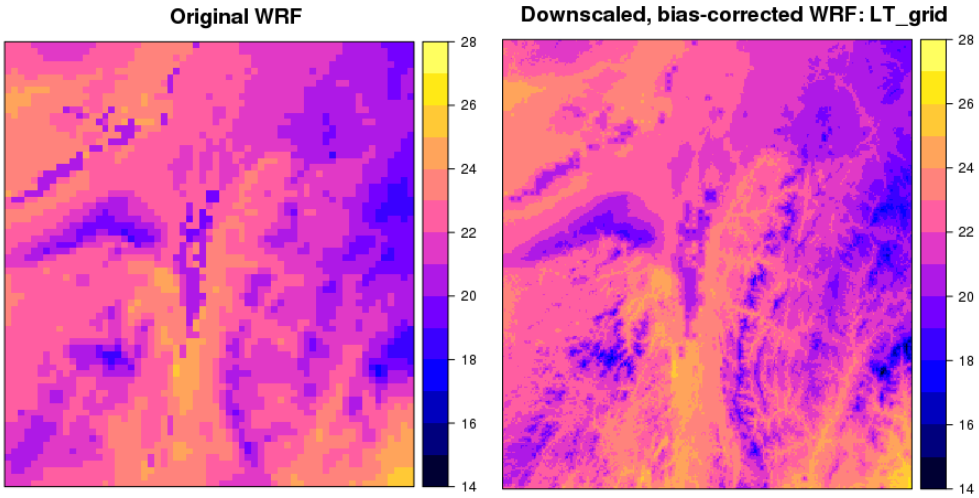

Figure 22: Original WRF simulations for TMAX (°C) and downscaled WRF TMAX (°C) using method LT_grid for August 5, 1982.