THE EFFECTS OF WEATHER CLASSIFICATION ON REGRESSION-BASED DOWNSCALING OF DAILY TEMPERATURE EXTREMA IN THE UNITED STATES

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

QUINN PALLARDY

Dr. Neil Fox, Dissertation Supervisor

DECEMBER 2022

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

THE EFFECTS OF WEATHER CLASSIFICATION
ON REGRESSION-BASED DOWNSCALING OF DAILY
TEMPERATURE EXTREMA IN THE UNITED STATES

presented by Quinn Pallardy,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Neil Fox

Professor Christopher Wikle

Professor Anthony Lupo

Professor Patrick Market

Professor Patrick Guinan

ACKNOWLEDGEMENTS

I would like to thank Dr. Neil Fox (University of Missouri) and Dr. Stephen Pallardy (University of Missouri) for their assistance with proofreading and editing, Dr. Christopher Wikle (University of Missouri) for his assistance with review of the methodology for Chapter 1, and Dr. Bohumil Svoma (formerly University of Missouri) for his contributions to the software used for this research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THE EFFECTS OF WEATHER CLASSIFICATION
ON REGRESSION-BASED DOWNSCALING OF DAILY
TEMPERATURE EXTREMA IN THE UNITED STATES

Quinn Pallardy

Dr. Neil Fox, Dissertation Supervisor

ABSTRACT

The focus of this dissertation was on the role played by weather classification in regression-based downscaling of daily temperature extrema. Three closely related studies were conducted, each using a different criterion for weather classification. The primary objective of all these studies was to evaluate changes in downscaling model performance as meteorological properties of the training periods were varied. This objective was of interest due to potential improvements in downscaling performance when accounting for non-static relationships between predictors and predictands. The first study used the time of day of the temperature extremum as the weather classification, while the third study used the direction of the wind as the weather classification. The second study used temperature as the weather classification, with a focus on possible consequences for downscaling in warmer conditions that were not present in the training conditions. Results from all three studies indicated that downscaling performance had the potential to be affected by the weather conditions seen in the training periods.

Chapter 1. The Effects of Atypical Diurnal Temperature Cycles on Regression-Based Downscaling of Daily Temperature Extrema in the Central United States

Abstract

The effects of variations in time of day of daily temperature extrema on regression-based statistical downscaling of daily temperature extrema were examined. These effects were analyzed by evaluating the performance of a regression-based downscaling model with multiple approaches to the incorporation of the relevant temperature data. The differing approaches included which predictor variables were selected for inclusion in the model, as well as variations in model methodology. Three different versions of the downscaling model were evaluated: (i) standard multiple linear regression, (ii) a weather classification scheme combined with multiple linear regression, and (iii) a weather classification scheme combined with multiple linear regression using dynamic time-step predictors. Bias and accuracy were measured on days with atypical and typical time of temperature extrema. The performance of regression models had the potential to be greatly degraded by days with atypical times of temperature extrema. The degree to which these atypical days were affected was dependent on which predictors were

included in the regression models, with the temperature extrema derived from reanalysis data playing the most important role. Implementation of the weather classification scheme also improved downscaling performances for atypical days in a number of situations. For typical days, the improvements to RMSE values were smaller and under were only present under certain predictor combinations.

1.1 Introduction

Downscaling is a process through which high resolution data is generated from low resolution data. It has many different applications in climate science research and analysis. Application of downscaling on output from lower resolution climate models has the potential to enable predictions associated with models that require higher resolution climate data for input, which may include hydrological, ecological, and crop yield models (Flint and Flint 2012; Grouillet et al. 2016; Robertson et al. 2007).

The two main types of downscaling approaches in use today are dynamical downscaling and statistical downscaling. Dynamical downscaling relies on simply increasing the resolution of climate models. Due to the extreme computational requirements of high resolution climate models, it is far more computationally expensive than statistical downscaling. The output is also limited to the variables generated by the model. Statistical downscaling relies on establishing statistical relationships between larger scale predictors and smaller scale predictands to predict values on a finer spatial scale. While computationally inexpensive, there are drawbacks associated with statistical downscaling. For one, a lengthy period of record is required for calibration. Another problematic aspect of statistical downscaling is the required assumption of stationarity with respect to climate change, which is relevant for the downscaling of climate modeling of the future (Wilby et al. 2004).

Statistical downscaling studies in recent years have largely fit into three different categories: weather classification, regression models, and weather generators (Wilby et al. 2004). Weather generators recreate the statistical properties of a climate variable without matching actual observations of that variable. Weather classification involves splitting the data into different weather “types” and is based on the theory that the statistical relationships between predictors and predictands will behave the same under similar weather conditions. The analog approach shares similarities with weather classification and has been used in numerous studies for temperature downscaling (Bettolli 2021; Bettolli and Penalba 2018; Brands et al. 2011b; Gutiérrez et al. 2013; Merkenschlager et al. 2021; Ribalaygua et al. 2013; Timbal and McAvaney 2001; Timbal et al. 2003). With the analog method, local conditions are linked with the simultaneous state of the atmosphere on a larger scale.

The predictands typically used in temperature downscaling include daily temperature extrema, daily mean temperature, or longer term mean temperature. These and other studies often compared the effectiveness of various versions of the methods using reanalysis data and a cross-validation approach (Gutiérrez et al. 2013). Downscaled output from climate models using the examined methods were sometimes included and interpreted.

The form of downscaling used in this study was multiple linear regression (MLR). Many forms of linear regression have been used to downscale temperature (Casanueva et al. 2013; Dirksen et al. 2020; Duhan and Pandey 2015; Fan et al. 2021; Goyal and Ojha

2011; Gutiérrez et al. 2013; Huth 2002; Huth 2004; Khan et al. 2006; Manzanas et al. 2018; Nojarov 2015; Radan 1999; Solman and Nuñez 1999). Nonlinear regression in the form of artificial neural networks (ANN) has also been used for temperature downscaling (Andreas and Gerd 1998; Coulibaly et al. 2005; Duhan and Pandey 2015; Gaitan et al. 2014; Goyal and Ojha 2011; Hernanz et al. 2021; Huth et al. 2007; Khan et al. 2006; Kostopoulou et al. 2007; Miksovsky and Raidl 2005; Schoof and Pryor 2001). The ANN studies typically compared the effectiveness of the ANN techniques to simpler linear regression techniques, with varying results. While regression can be simple and computationally efficient, it does have potential drawbacks. Regression models tend to produce downscaled results with lower variance than seen in observations, due to the inability of large scale predictors to fully explain all local variability (Wilby et al. 2004). Variance inflation or randomization are two methods that have been proposed as potential solutions to this problem (Huth 2002). The randomization approach consists of adding noise with the desired properties to the downscaled data set. Variance inflation is a procedure where all the downscaled anomalies are increased by the same factor to equalize the variance in the data sets. A problem with this approach is that it relies on the imperfect assumption that all the variability at the finer scale is dependent on the variability at the larger scale (Maraun 2013).

Observational data is necessary for statistical downscaling, and the degree of homogeneity in the data set used is an important factor in determining the reliability of downscaled results. Ensuring that inhomogeneities are minimized is an important factor in the development of temperature data sets (Peterson et al. 1998). These inhomogeneities can take the form of biases introduced through varying of the time-window in which the temperature extrema are recorded or in the frequency at which temperature sampling is conducted (Gough et al. 2020; Vincent et al. 2009). Station relocation and changes in observation procedures can also introduce inhomogeneities (Vincent et al. 2002). The prior examples describe inhomogeneities related to inconsistencies in measurement, but inhomogeneities can also be physical in nature. In (Žaknić-Ćatović and Gough 2021) two different regimes of temperature extrema were identified: the radiative regime and the advective regime. Temperature extrema for the radiative regime were aligned with the diurnal radiation cycle, while the corresponding extrema for the advective regime were far less influenced by the diurnal radiation cycle. Potential consequences of these different patterns of temperature extrema timing on the performance of downscaling of daily temperature extrema were the primary focus of this study. Two types of days were defined for this study, the atypical day, and the typical day. For the typical day, the temperature extrema occurred at the time of day where the extrema were most frequently seen. Radiative forcing was the driving factor in the timing of temperature extrema for the typical day. For the atypical day, the temperature extrema occurred at either the beginning or end of the day, at the side opposite the time of typical temperature extrema. For the atypical day, advective forcing was the driving factor in temperature extrema timing.

In the regression downscaling model, the larger scale state of the atmosphere is described using predictor variables. How these predictors are chosen can determine whether the model has access to information about the time of day of the temperature extrema. One key factor in the selection of predictors is the time of day at which they are taken, and whether that time of day can change dynamically. Attention to the effects of the time of day of predictors on downscaling performance has been sparse. Gutiérrez et al. 2013 offered a limited examination of the effects of shifting time-steps and found that daily minimum and maximum temperatures were affected differently, with a dynamic time-step setup performing better for downscaling maximum temperatures, and a static time-step setup performing better for downscaling minimum temperatures. The dynamic time-step setup used by Gutiérrez et al. 2013 included predictors from two time-steps, while the static time-step setup used predictors from only a single time-step. Using predictors that remain unchanging in time to predict extrema values that are not necessarily occurring near the same time every day has the potential to cause problems with downscaling performance. A potential solution examined in the study was the inclusion of modeled temperature extrema or mean daily temperature as predictors.

Another potential way of addressing the heterogeneity in time of day of temperature extrema issue was through the introduction of a weather classification technique. The weather classification technique used in this study divided the training period used for regression into partitions containing only typical or atypical days. While the weather classification process served to bolster the physical homogeneity of the days in each training period, it also potentially significantly limited the number of data points that could be used for regression. Combining weather classification with regression has been attempted before, though not in relation to heterogeneity in time of day of temperature extrema (Gutiérrez et al. 2013; Ribalaygua et al. 2013). To the author's knowledge, problems introduced to regression-based downscaling by inhomogeneities in time of day of temperature extrema has not been examined in the literature. Examining these issues and evaluating potential ways to address them were the primary objectives of this research.

1.2 Methodology

The following section describes the methodology used to conduct the study. Included are information about the data used for the study, as well as how the downscaling model was created and evaluated. The downscaling model was designed to evaluate potential effects of non-static diurnal temperature cycles on downscaling performance. This was achieved by incorporating information related to the time of day of temperature extrema through predictor variable selection and/or the implementation of a weather classification technique.

1.2.1 Data

The downscaling model used for this study relied on establishing a regression-based relationship between large scale “predictor” variables and small scale “predictands”. Predictor variables were used to describe the state of the atmosphere on a larger scale, while the predictand was what the model is designed to predict based on the atmospheric state. Data from the NARR, or North American Regional Reanalysis dataset (Mesinger et al. 2006), were used for the large scale predictor variables. The NARR set consists of gridded data covering North America at roughly 0.3 degrees or 32 km in resolution. NARR data have a temporal resolution of 3 hours, spanning 1979 to the present. Data from the standard levels used for analysis were considered, including data from the surface, and at the 850 and 500 hPa levels (Table 1.1). Meteorological variables obtained from the NARR included temperature, u wind, v wind, specific humidity at the surface and the specified pressure levels, and sea level pressure. The period of time examined for this study was the 30-year period from 1981 to 2010.

Minimum and maximum daily temperature data were obtained at 40 different stations located throughout the central United States for the same period of time (Fig. 1.1a). The stations were selected for the complete availability of data in the considered time period and for their locations in the interior of the North American continent. These temperature extrema data served as the predictands for the study.

Table 1.1 Variable Information. The times of non-daily variables were dependent on the downscaling method used.

Variable Code	Description	Levels
Ps	Sea Level Pressure	Surface
T	Temperature	850 ,500 hPa
U	U Wind	850, 500 hPa
V	V Wind	850, 500 hPa
H	Specific Humidity	850 ,500 hPa
Ts	Surface Temperature	Surface
Us	Surface U Wind	Surface
Vs	Surface V Wind	Surface
Hs	Surface Specific Humidity	Surface
Tm	Mean Daily Surface Temperature	Surface
Tn	Min Daily 3-Hourly Surface Temperature	Surface
Tx	Max Daily 3-hourly Surface Temperature	Surface

(a)

(b)

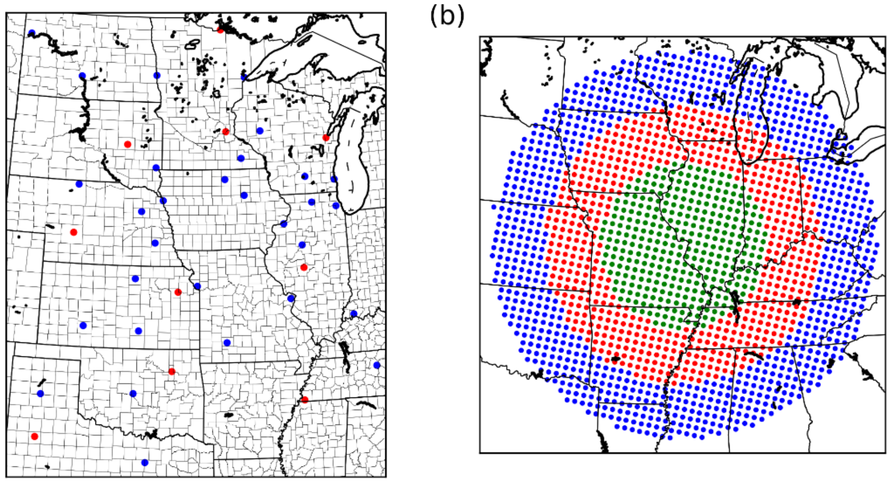

Figure 1.1 (a) Station locations across the Central United States selected for inclusion in this study. The red locations mark a subset of the data used for additional testing in the study. (b) North American Regional Reanalysis (NARR) cell locations of small (green), medium (red), and large (blue) domains associated with the station located in Saint Louis, Missouri. The focus of the study was on results for the smallest (green) domain size.

1.2.2 Predictor Variable Combinations

The variables chosen as predictors have been commonly used for downscaling purposes (Gutiérrez et al. 2013). Global Climate Models (GCMs) have been shown to skillfully reproduce these variables over southwestern Europe (Brands et al. 2011a). A subjective approach has often been taken for which combinations of predictors are tested. Correlation analysis (Khan et al. 2006) and stepwise regression (Gaitan et al. 2014) have also been used for predictor selection.

Different predictor variables combinations describe the state of the atmosphere in different ways, and in the present study the downscaling performances of several different predictor combinations were examined. Three different individual variables were tested: mean surface daily temperature, daily reanalysis maximum/minimum surface 3-hour temperature, and surface temperature from a single timestep. In this study the single timestep variables were taken at 21 UTC for daily maximum temperature and 12 UTC for daily minimum temperature. Note that the maximum/minimum surface temperatures were the highest/lowest values of the 3-hour instantaneous surface temperatures (instantaneous temperatures at 3-hour intervals) between 06 UTC one day and 06 UTC the following day and were not the actual maximum/minimums. This 06 UTC to 06 UTC period marked the boundaries of the day during which maximum/minimum temperatures were observed at all tested locations. To avoid confusion with the minimum and maximum temperature predictand variables that are being predicted, the surface maximum and minimum temperature variables extracted from the NARR data and used for predictors will be referred to as the reanalysis maximum temperature and reanalysis minimum temperature.

In addition to these three individual variables, many variable combinations were examined (Table 1.2). Single timestep values of surface temperature, u wind, v wind, sea level pressure, and specific humidity formed the most basic combination. Subsequent combinations cumulatively added temperature and wind velocity data from higher up in the atmosphere. Mean surface temperature or reanalysis maximum/minimum surface temperature were examined as add-ons to each combination. The primary goal with predictor combination selection was to evaluate performance with and without the mean surface temperature and the reanalysis maximum/minimum surface temperature variables. These two variables offered temperature information to the downscaling model that was independent of the diurnal temperature cycle, and in the case of the reanalysis maximum/minimum surface temperature described the temperature near the point in time at which the temperature extremum was reached.

Table 1.2 Tested Combinations of Predictors.

Combination Name	Predictors
CS	Ps, Ts, Us, Vs, Hs
CSm	Ps, Ts, Us, Vs, Hs, Tm
CSx	Ps, Ts, Us, Vs, Hs, Tx
CSn	Ps, Ts, Us, Vs, Hs, Tn
C8	CS, T850, U850, V850, H850
C8m	CS, T850, U850, V850, H850, Tm
C8x	CS, T850, U850, V850, H850, Tx
C8n	CS, T850, U850, V850, H850, Tn
C5	C8, T500, U500, V500, H500
C5m	C8, T500, U500, V500, H500, Tm
C5x	C8, T500, U500, V500, H500, Tx
C5n	C8, T500, U500, V500, H500, Tn
Ts	Ts
Tm	Tm
Tn	Tn
Tx	Tx

1.2.3 Downscaling Model Methodology

The regression method used for the downscaling model in this study was the linear least squares form of multiple linear regression (MLR). The MLR method was chosen for its simplicity and computational inexpensiveness.

Evaluation Methods

To evaluate the effectiveness of downscaling, an approach similar to the k-fold cross-validation technique described in (Gutiérrez et al. 2013) was used. The application of cross-validation in analysis of performance ensured that model overfitting was avoided. For this technique, the period of record examined was split into five different sections, with each section in turn being used as a validation period with the other sections used as training periods. This resulted in an 80/20% split in training/validation periods. To mitigate any influence from year-to-year trends the sections were staggered. For example, the first validation period was the following six years: 1981,1986,1991,1996,2001, and 2006. The next validation period advanced by one year: 1982,1987,1992, 1997,2002, and 2007.

Root mean square error (RMSE) and bias were evaluated using this cross-validation approach. The RMSE was given by the square root of the mean value of the squared predicted errors. Lower RMSE values were indicative of better performance, though were more affected by outliers than the mean absolute error due to the squaring of the errors. Bias described how well the model over or under predicts what the temperature will be. The RMSE and bias values were evaluated by comparing differences in the daily temperature extrema predicted from the downscaling model and the observed daily temperature extrema at the station locations.

Domains

The domains used in the downscaling model defined the region for which predictor data was confined to. The domain consisted of all NARR cell locations within a certain great circle distance of the location of the station being downscaled. Three distances were tested: 3°, 5°, and 7° of the equivalent latitudinal distance on a spherical earth. The results for this study were limited to the smallest domain size of 3°; the smallest domain tended to perform (with respect to RMSE) slightly better or similar to the larger domains under most circumstances. Figure 1.1b shows the relevant NARR cell locations of these domains for the Saint Louis station. Although the number of cells per domain was not identical across all the stations, the variation was not large; the number of cells per domain at each station stayed within roughly 6% of the mean value across all stations.

Normalization

The first step in the downscaling process was to convert the data into standardized anomalies. This procedure was performed to remove any dependence on seasonality in the model (Gutiérrez et al. 2013). This is relevant when the downscaling model is used for situations where the seasonality may change, as it may in modeling for future climates (Gutiérrez et al. 2013).

The 24-year training period was not sufficient to generate a reasonably noise-free mean and standard deviation by Julian day for the data, so a smoothing technique was employed. The Julian Day mean values were smoothed by averaging data from a window on the trailing and leading sides of the Julian Day in question. The same manipulation was performed for standard deviation after subtracting the smoothed means from the data. All variables in the study were standardized with a 20-day window (on each side) in accordance with optimal results from preliminary testing. Data from the validation periods were normalized using the means and standard deviations from the training periods. This was done to keep the validation period data completely independent of the training period data.

PCA

Principal Component Analysis (Preisendorfer and Mobley 1988) (PCA) is generally used to reduce the dimensionality of a data set while retaining most of the explained variance. The variables used for MLR were given by the principal components (PCs) of normalized data from the selected domain. The dominance method of PC selection was used, with the total number of variables retained capped in proportion to the length of the training period. Prior studies have indicated no loss of performance with higher order PCs that would typically be considered noise (Huth 2002; Radan 1999). These were done with a much simpler data set, where the ratio of the length of the training period to the number of variables in the model was not an issue. The maximum number allowed by the length of the training period was given by the ratio of training days to total variables to be entered into the model. Thirty-five different ratios were tested, ranging from a minimum of 4 to a maximum of 8000, and the optimal value (minimum RMSE) for that particular station/variable combination was chosen. This threshold was more relevant for the methods that relied on weather classification, due to the shortened training periods. If any ratios were sufficiently high to completely eliminate a variable from the model, the corresponding results were not computed.

The principal component analysis (PCA) was performed on data from the training period and data from the validation period was projected onto the PCA axes generated from the training period. This was again done to keep the validation period data completely independent of the training period data.

<u>Time of Day of Reanalysis Maximum/Minimum Temperature</u>

Two of the three versions of the downscaling model developed for the present study used a combined weather classification and regression approach. The categories of weather in this case were the varying times of day that the reanalysis maximum and minimum temperatures were estimated to have taken place using the surface temperature predictor data set. The reanalysis temperature data were calculated for each station using spatial bilinear interpolation. The estimated time of reanalysis maximum and minimum temperature for every day was calculated based on these interpolated values, at the 3-hour temporal resolution of the NARR data.

All stations were in the central time zone of the United States, where the beginning and end of each day occurs at 06 UTC. Figures 1.2a and 1.2b show the annual time of day of reanalysis maximum and minimum temperature distribution across all 40 stations. The median number of days where the reanalysis minimum temperature occurred at 06E UTC (E here refers to end of the day) was roughly 25% of the total number of days (Fig. 1.2a). The median number of days where the reanalysis maximum temperature occurred at 06B UTC (B refers to beginning of day) was slightly over 5% (Fig. 1.2b). The corresponding values for the times of reanalysis minimum and maximum temperature (12 UTC for minimum, 21 UTC for maximum) were about 80% and 50%, respectively.

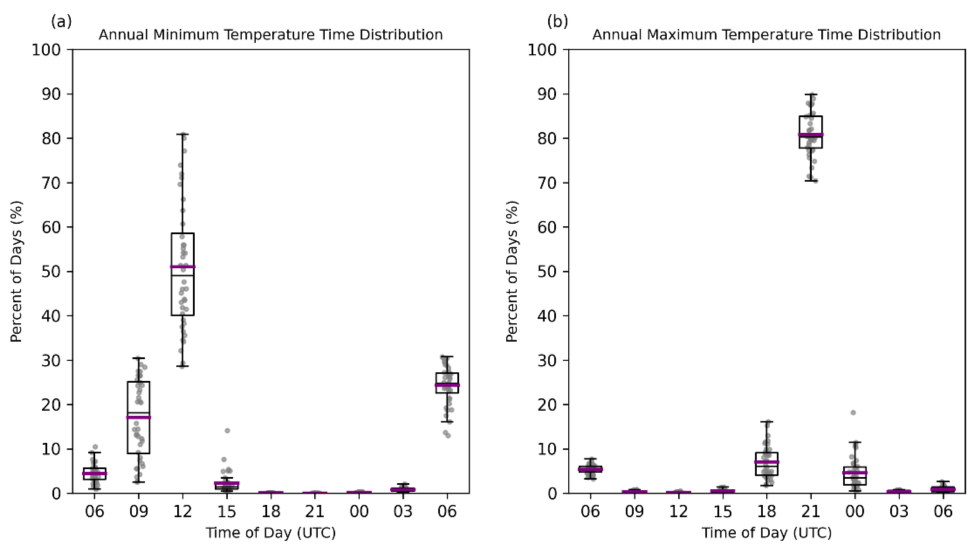

Figure 1.2 Box and whisker plots of percent of days for which minimum temperature (a) and maximum temperature (b) occurred at a given UTC time on an annual basis. Median values are given by the thin black lines and mean values by the thick purple lines.

Figure 1.3 shows the seasonal breakdown of the distributions for reanalysis minimum temperature shown in Figure 1.2. The highest percentage of days with the reanalysis minimum temperature at 06B was winter, and the lowest in summer (Fig. 1.3a, 1.3c). A greater number of days with time of reanalysis minimum temperature at 03 UTC occurred in summer, potentially due to the greater day lengths (Fig. 1.3c). Figure 1.4 shows the corresponding seasonal breakdown for maximum temperature. Maximum temperatures at the beginning of the day occurred predominately in winter and were almost non-existent in summer (Fig. 1.4a, 1.4c). The higher percentages in winter for atypical time of expected temperature extrema were not unexpected, as they are generally associated with cold air advection (Fig. 1.4a). Cloud cover can also potentially play a role in the time of day for temperature extrema.

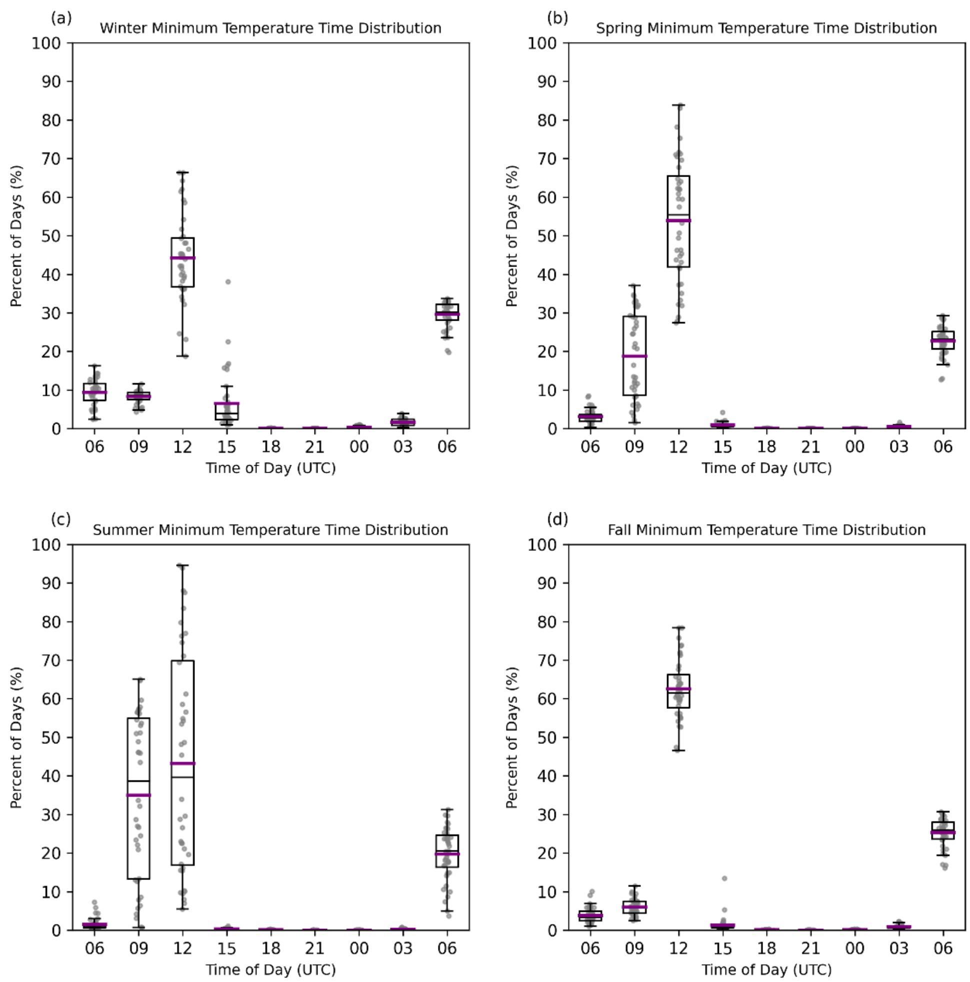

Figure 1.3 Box and whisker plots indicating the frequency at which the minimum temperature occurred at the specified times of day in winter (a), spring (b), summer (c), and fall (d) at all 40 station locations. Median values are given by the thin black lines and mean values by the thick purple lines.

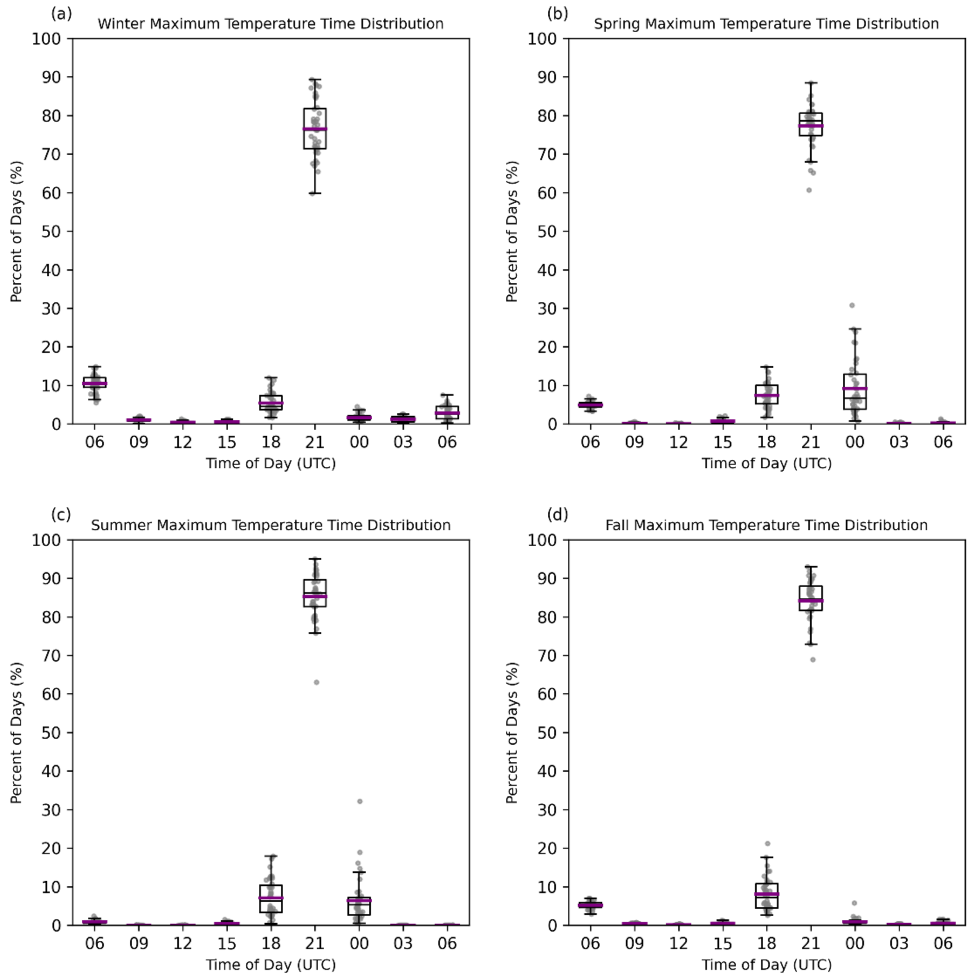

Figure 1.4 Box and whisker plots indicating the frequency at which the maximum temperature occurred at the specified times of day in winter (a), spring (b), summer (c), and fall (d) at all 40 station locations. Median values are given by the thin black lines and mean values by the thick purple lines.

The focus of this study was on downscaling performance for days that were typical and days that were atypical with respect to time of temperature extrema. The typical days were those for which the minimum and maximum temperatures were expected to occur at 12 UTC and 21 UTC, respectively. Atypical days were chosen as the days where the maximum and minimum temperatures were at 06B UTC and 06E UTC, respectively. These times were chosen because they were the farthest from the typical times and had a sufficient number of days to be used as data points. As mentioned in the introduction, for the typical days radiative forcing was the dominant factor in the time at which the extrema occurred, while for the atypical days advective forcing was the dominant factor in the timing of extrema.

Atypical/typical days for minimum temperature were defined independently from those days for maximum temperature, meaning that an atypical day for minimum temperature could potentially be a typical day for maximum temperature and vice versa.

Versions of the Downscaling Model

Three versions of the downscaling model were tested. The first version was designated the standard model (M1 method), where the entire training period was used. The second version used the previously described combined weather classification and regression approach (M2 method). For the second version, atypical and typical days from the training period in the model were used to downscale the temperature for the corresponding atypical and typical days from the validation period.

The third version was identical to the second, except variables with dynamic times were used (M3 method). For the M3 method, all single time-step variables were adjusted based on whether the atypical or typical days were being used for training. This meant that when the atypical days were used for training, the single time-step variables with the M3 method were set at 06B UTC for maximum temperature downscaling and 06E UTC for minimum temperature downscaling. Single time-step variables for training with typical days with the M3 were set at the standard 21 UTC for maximum temperature downscaling and 12 UTC for minimum temperature downscaling.

With the weather classification approach, the training periods were adjusted so that no inhomogeneity existed in the time of day of temperature extrema within the training period. Results using the original and modified training periods could then be compared to evaluate potential effects of the inhomogeneity.

<u>Choice of Time-step for Static Predictors</u>

As previously noted, the time-steps of 21 UTC and 12 UTC were chosen for the static time-step of predictor variables used for the prediction of maximum daily temperature and minimum daily temperature, respectively. The objective was to select the times for prediction of maximum and minimum temperature that broadly performed the best across all predictor combinations. The 21 UTC and 12 UTC selections were logical choices due to their being the most frequent time of reanalysis maximum and minimum temperature at every station. Results from relevant preliminary testing are shown in Figure 1.5 (Minimum) and Figure 1.6 (Maximum). The stations used for the preliminary testing are marked in red in Figure 1.1a. These figures show the performances for selected predictors combinations as the static time-step was adjusted. For prediction of minimum temperature daily temperature, all variable combinations, including those with the mean daily temperature and the reanalysis minimum daily temperature, overall performed best with the static predictors set at 12 UTC (Fig. 1.5a). For atypical days, all predictor combinations examined trended toward improved RMSE performance as the time-step used for predictors got later in the day, suggesting that time-step of the static predictors was important for predicting minimum temperatures on the days where the reanalysis minimum temperature occurred at the end of the day (Fig. 1.5b). The performance for prediction of maximum daily temperature was not quite so uniform, with the optimal time-step varying with the variable combination used (Fig. 1.6a). This was particularly true for the combinations with the reanalysis maximum temperature included. As was expected, the combinations that included either reanalysis minimum or maximum daily temperature saw much less variation with change in time-step of static predictors.

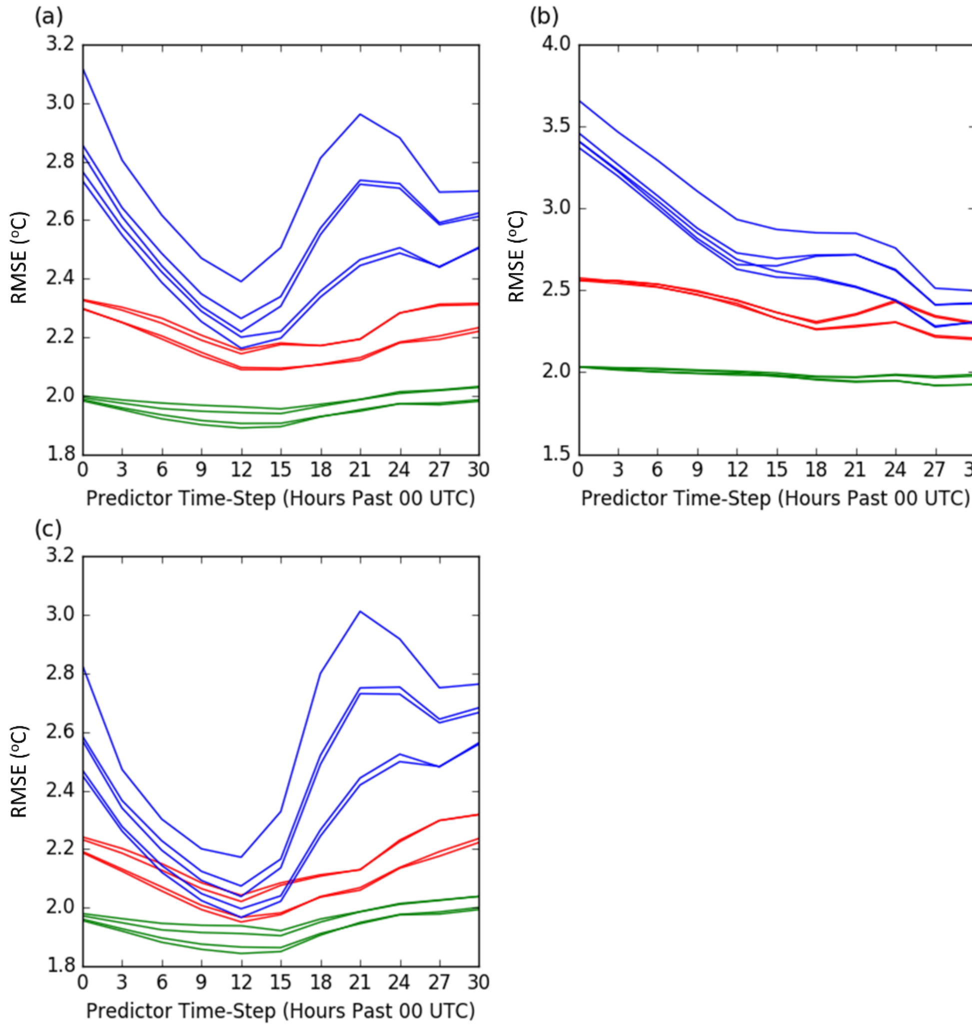

Figure 1.5 RMSE (°C) performance for downscaling of minimum daily temperature using a subset of data locations (marked in red in Fig. 1.1a) and allowing the time of single time-step predictors to vary. The combinations were split into three categories: combinations where only the single time-step predictors were included (blue, combinations included CS, C8, and versions of those with specific humidity excluded), combinations where the mean daily surface temperature predictor was included in addition to the single time-step variables (red, combinations included CSm, C8m, and versions of those with specific humidity excluded), and combinations where the estimated minimum daily surface temperature predictor was included in addition to the single time-step predictors (green, combinations included CSn, C8n, and versions of those with specific humidity excluded). (a) shows RMSE performance for all days, (b) shows performance for atypical days (minimum temperature at 06E UTC), and (c) shows performance for typical days (minimum temperature at 12 UTC).

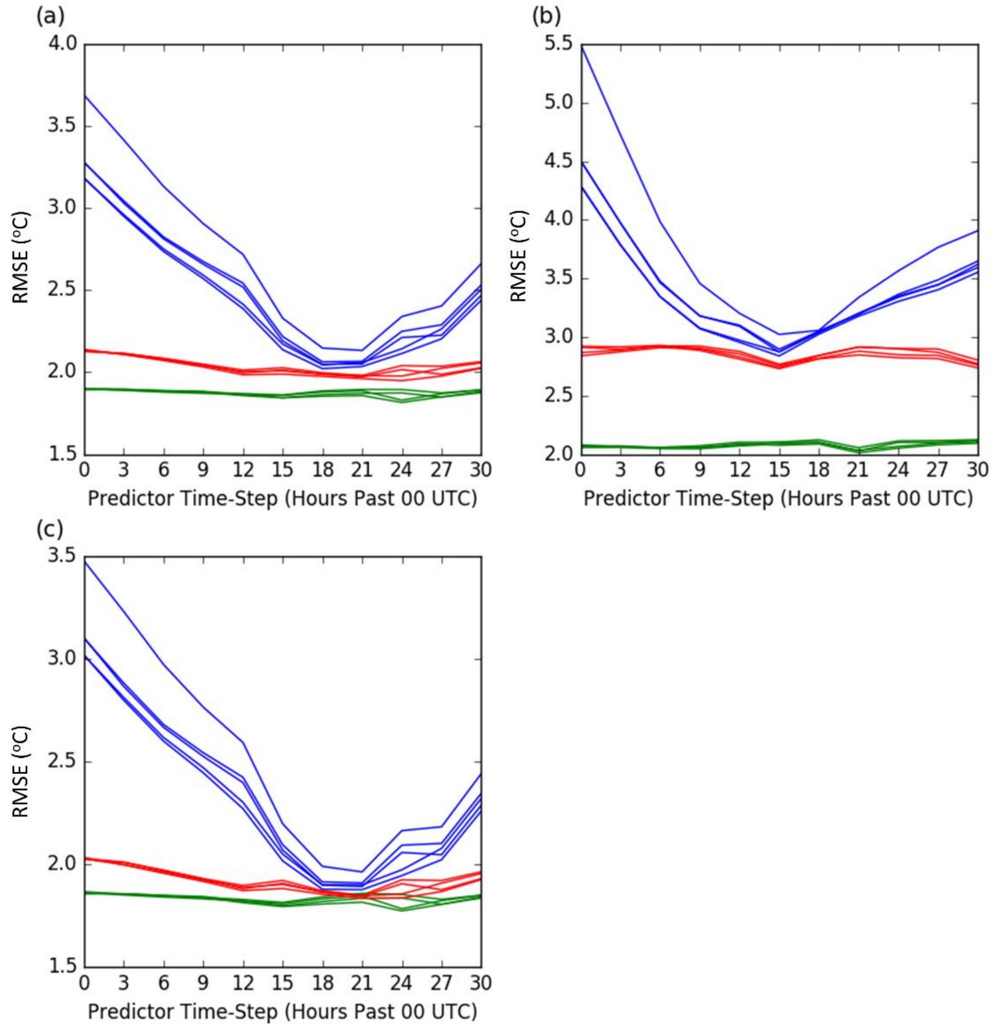

Figure 1.6 RMSE (°C) performance for downscaling of maximum daily temperature using a subset of data locations (marked in red in Fig. 1.1a) and allowing the time of single time-step predictors to vary. The combinations were split into three categories: combinations where only the single time-step predictors were included (blue, combinations included CS, C8, and versions of those with specific humidity excluded), combinations where the mean daily surface temperature predictor was included in addition to the single time-step variables (red, combinations included CSm, C8m, and versions of those with specific humidity excluded), and combinations where the estimated maximum daily surface temperature predictor was included in addition to the single time-step predictors (green, combinations included CSx, C8x, and versions of those with specific humidity excluded). (a) shows RMSE performance for all days, (b) shows performance for atypical days (maximum temperature at 06B UTC), and (c) shows performance for typical days (maximum temperature at 21 UTC).

1.3 Results and Discussion

The following portion of the study contains the results of the experiments conducted for the study as well as some discussion related to the results. Included are RMSE and bias results for minimum and maximum temperature for both atypical and typical days. When analyzing the results, the effects of predictor variable selection and type of weather classification used on downscaling performance were the primary focus. These were the two factors that determined what information about the diurnal temperature cycle was available to the downscaling model.

1.3.1 Daily Minimum Temperature – Atypical Days

The performance of techniques downscaling daily minimum temperatures when the reanalysis minimum temperature occurred at an atypical time (06E UTC, referencing the end of the calendar day) is discussed in this section. Box and whisker plots describing the results at the 40 site locations using the indicated measure were generated. All results in Section 3 used the smallest domain (3^{o} "radius") for the predictors, with training and validation data sets using the same domain.

RMSE

Figure 1.7a illustrates the results for RMSE using the M1 method (single time-step predictors, except for the temperature extrema and mean temperature variables where indicated). A large increase in performance for the selected days occurred when the information about the mean daily temperature (MT combinations, includes Csm, C8m, and C5m), or the reanalysis minimum daily temperature (ET combinations, includes Csx, C8x, C5x, Csn, C8n, and C5n) was included in the predictor combination, with the inclusion of the reanalysis daily minimum temperature resulting in the largest increase in performance. Less clear were any effects of adding additional variables at different levels in the atmosphere. The biggest increase in performance with respect to variables from upper levels in the atmosphere came from the inclusion of 850 hPa information to the single time-step) combinations (ST combinations, includes Cs, C8, and C5). Results for the single variable (SV combinations, includes Ts, Tm, and Tx) combinations indicated that using only the reanalysis minimum or mean daily temperatures as predictors produced only slightly poorer performance than using the corresponding MT and ET combinations. The performance decrease from using only the single time-step surface temperature (compared to the ST combinations) was much larger.

With the M2 method (single time-step predictors and weather classification), the range of differences in the RMSE of the ST, MT, and ET combinations narrowed (Fig. 1.7b), implying that the advantage from the inclusion of the reanalysis temperature extrema as predictors was reduced through the inclusion of weather classification. This result was seen in the percent changes in RMSE obtained at each tested location by switching from the M1 to M2 method (Fig. 1.8a). Decreases in RMSE were seen across the board, with the smallest decreases seen in the ET combinations.

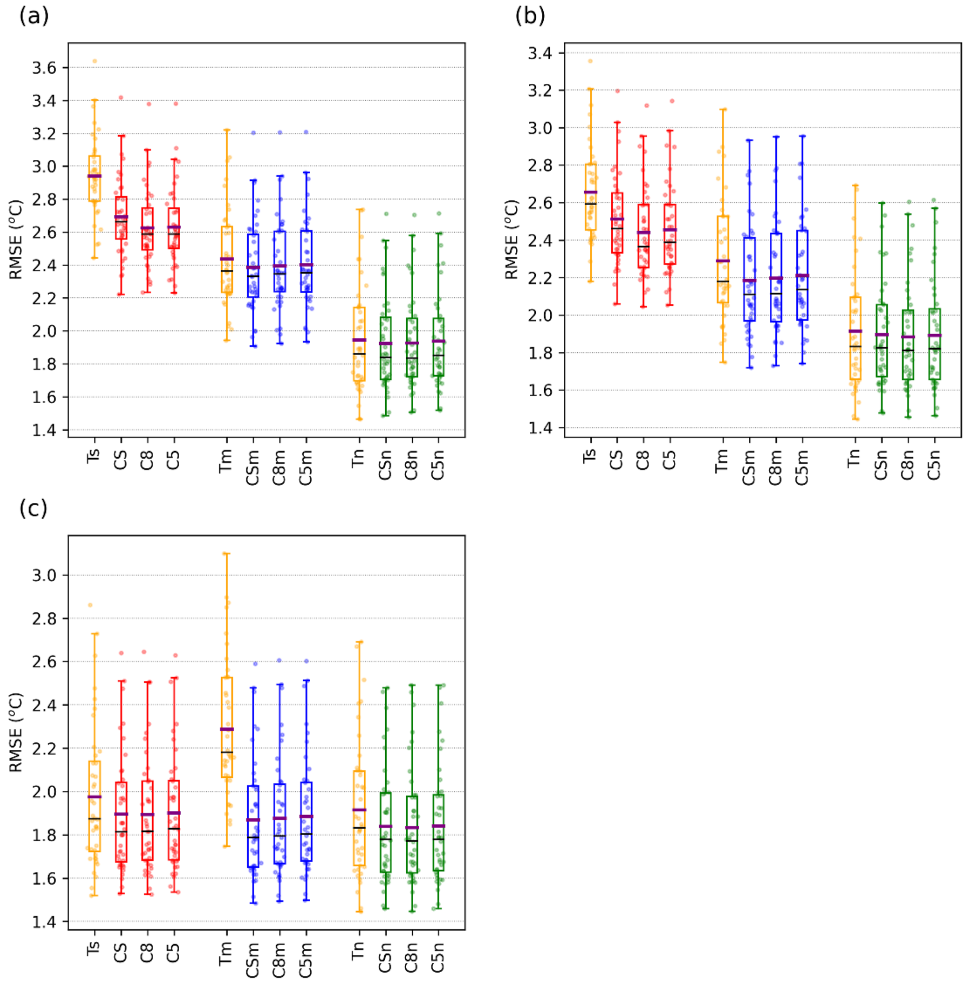

Figure 1.7 RMSE results at all 40 locations for downscaling of minimum daily temperature *only on atypical days* using the M1 method (a), the M2 method (b) and the M3 method (c). The ST (red) combinations used only single time-step data for predictors, the MT (blue) combinations included the Tm predictor in addition to the single time-step predictors, the ET (green) combinations included the Tn predictor in addition to the single time-step predictors, and the SV (yellow) combinations used only the single predictor indicated. Median values are given by the thin black lines and mean values by the thick purple lines.

(a)

(b)

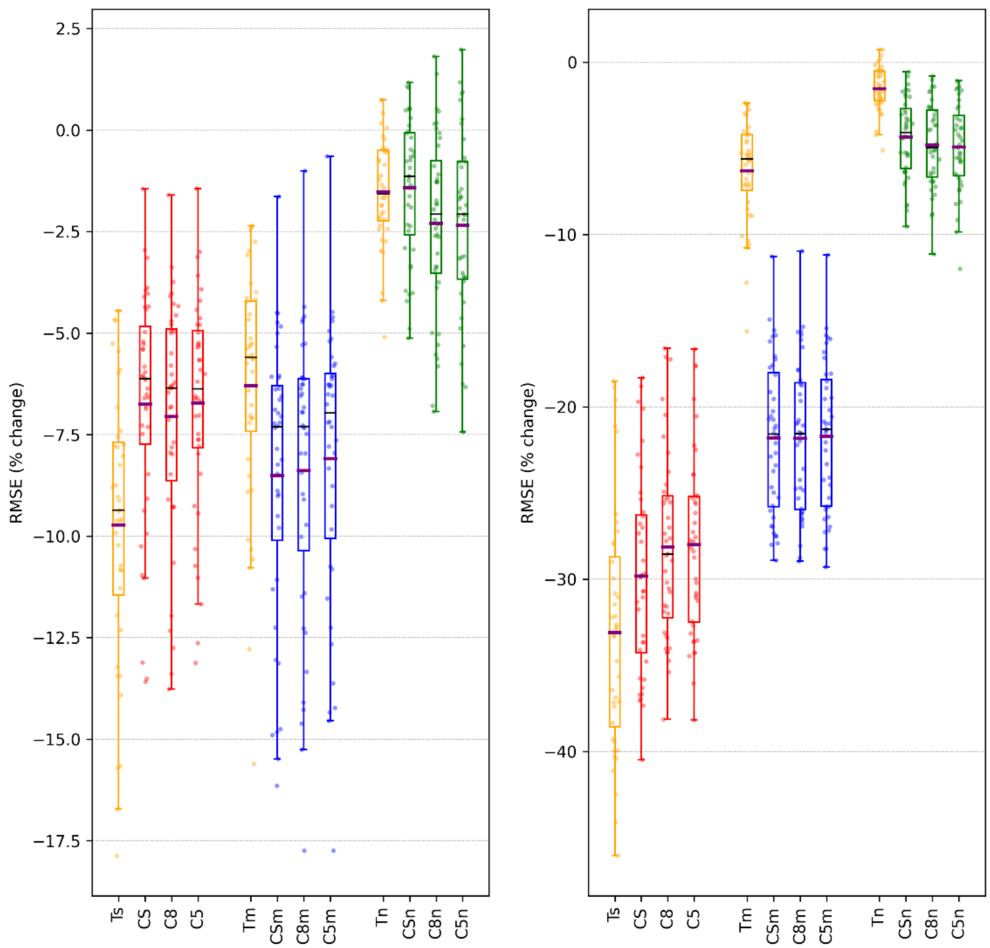

Figure 1.8 (a) The percent changes in RMSE resulting from switching from the M1 method to the M2 method for the results from Figure 1.7. (b) The percent change in RMSE resulting from switching from the M1 method to the M3 method for the results shown in Figure 1.7. Predictor combination information is the same as described in Figure 1.7. Note the scale changes in the ordinate between figures. Median values are given by the thin black lines and mean values by the thick purple lines.

For the M3 method (dynamic time-step predictors and weather classification) differences in RSME values among the ST, MT, and ET combinations decreased substantially, with the ET combinations still performing the best (Fig. 1.7c). Use of the single time-step temperature predictor (Ts) resulted in a very large decrease in RMSE with the M3 method. Figure 1.8b shows the changes in RMSE at all stations between the M1 to M3 methods.

For atypical minimum daily temperature, the overall best RMSE performances among the combinations examined were seen with the ET combinations using the M3 method, with all ET combinations having very similar results.

Bias

For the M1 method, the ST and MT combinations exhibited a tendency to substantially overestimate temperatures for atypical days (Fig. 1.9a). The ET combinations, while still overestimating temperature, did so at a smaller magnitude. Biases using the M2 and M3 methods were still positive (Figs. 1.9b, 1.9c), but greatly reduced compared to the M1 method (cf. Fig. 1.9a with Figs. 1.9b, 1.9c). The ET combinations produced the smallest biases for the M1 and M3 methods, with the pattern less clear for the M2 method.

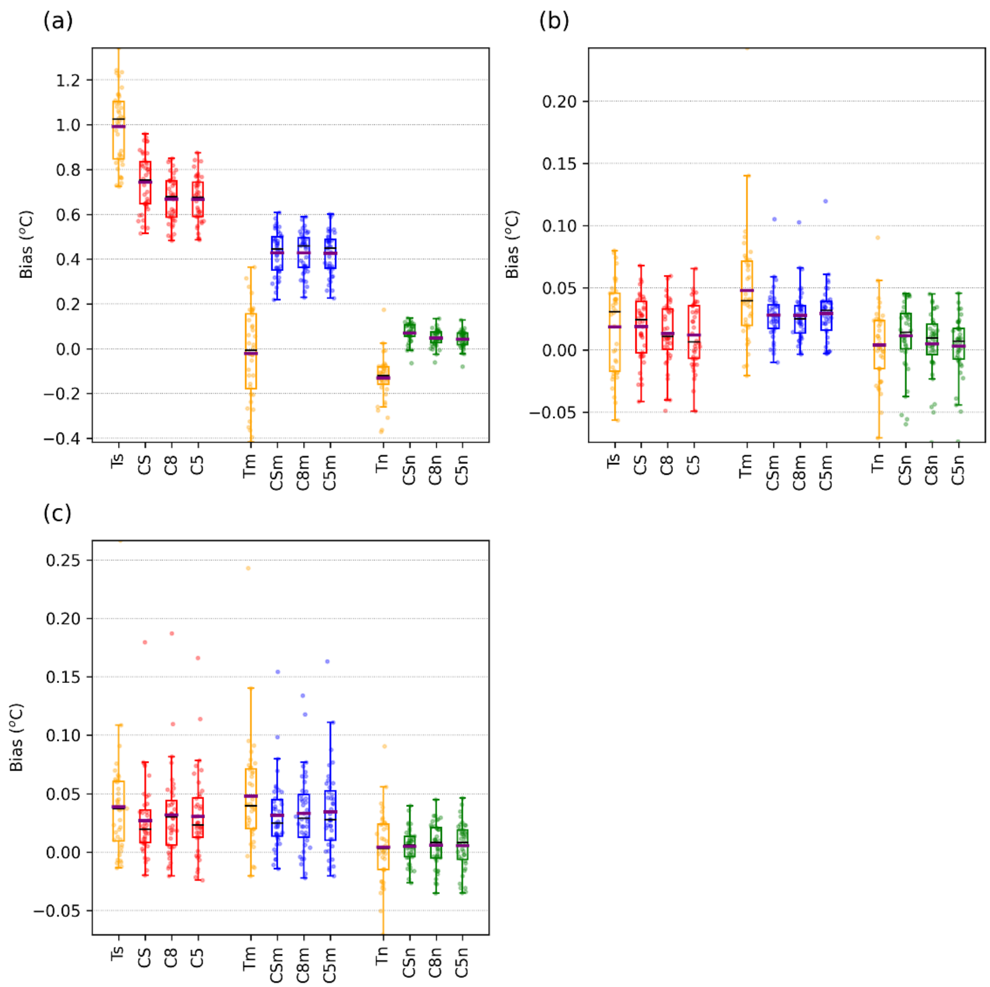

Figure 1.9 (a) Bias (°C) results at all 40 locations for downscaling of minimum daily temperature *only on atypical days* using the M1 method (a), the M2 method (b) and the M3 method (c). Predictor combination information is the same as described in Figure 1.7. Note the scale changes in the ordinate among figures. Median values are given by the thin black lines and mean values by the thick purple lines.

1.3.2 Daily Minimum Temperature – Typical Days

The performance of downscaling of daily minimum temperatures when the reanalysis minimum temperature occurred at the typical time (12 UTC) is discussed in this section. The primary objective of this exercise was to determine whether the inclusion of atypical days in the training portion of the standard M1 method had the potential to degrade the performance of the downscaling model for typical days.

RMSE

RMSE performance for the M1 method on the typical days is shown in Figure 1.10a. As was the case for the atypical days, the ET combinations performed best but with a smaller margin than that seen with the atypical days. The SV methods tended to perform poorly with respect to RMSE (Fig. 1.10a). RMSE differences between the ST, MT, and ET combination types were very small for the M3 method on typical days (Fig. 1.10b). Results for the M2 method were identical to those for the M3 method for typical days and thus are not shown.

The change in RMSE at each tested location by switching from the M1 to the M3 method is shown in Figure 1.10c. The dynamic predictor time/weather classification method (M3) produced slightly increased RMSE for the ET combinations, while the ST and MT methods exhibited modest decreases in RMSE. These results were evidence that the inclusion of atypical days in the training period did degrade performance for the predictor combinations where the reanalysis minimum temperature was not included.

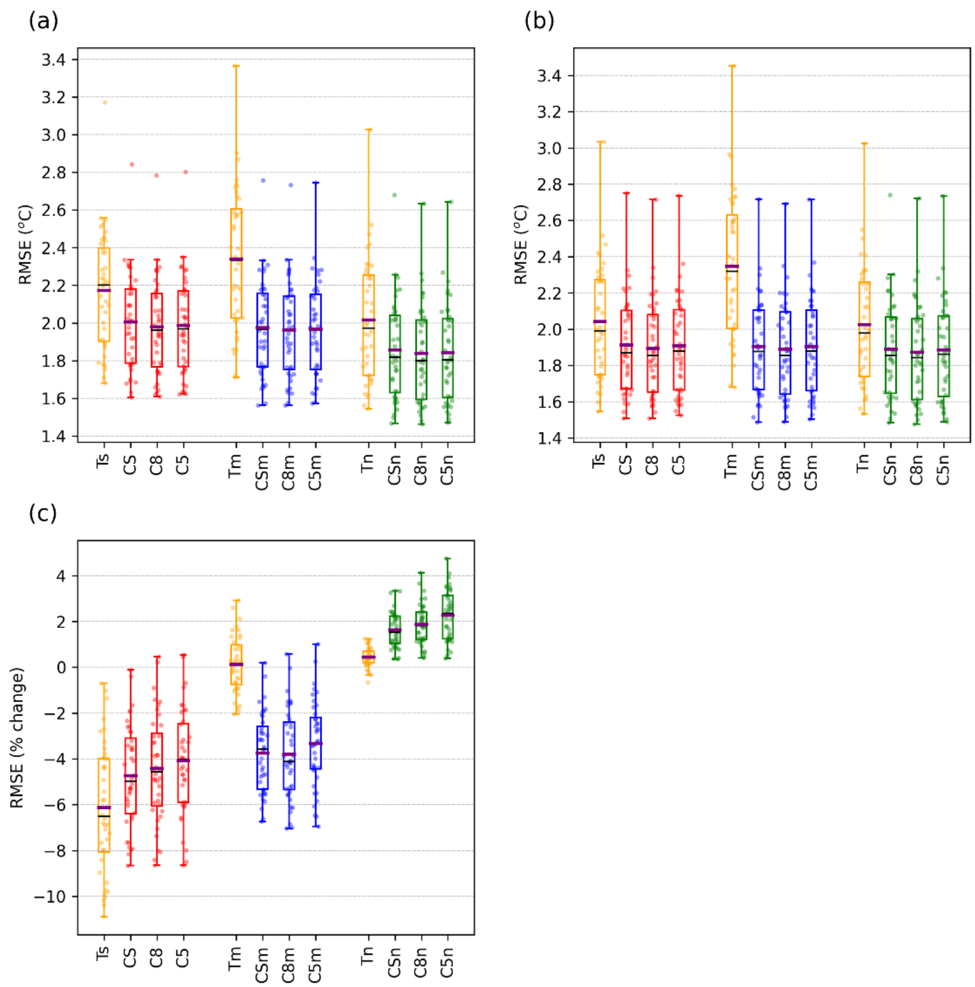

Figure 1.10 RMSE (°C) results at all 40 locations for downscaling of minimum daily temperature *only on typical days* using the M1 method (a), the M3 method (b). The percent changes in RMSE resulting from switching from the M1 method to the M3 method are shown in (c). Predictor combination information is the same as described in Figure 1.7. Median values are given by the thin black lines and mean values by the thick purple lines.

For typical minimum daily temperature, the overall best RMSE performances among the combinations examined occurred with the ET combinations using the M1 method, with the ET combinations again having very similar results.

Bias

Bias results for the M1 method on typical days (Fig. 1.11a) indicated the ST and MT combinations tended to underestimate daily minimum temperature, but not to the degree that they overestimated temperature on atypical days (cf. Fig. 1.9a and Fig. 1.11a). Conversely, the ET combinations slightly overestimated daily minimum temperature. Bias for the M3 method on typical days is shown in Figure 1.11b. The magnitude of biases seen using the M3 method were smaller than those seen with the M1 method for the ST and MT combinations, and reversed in sign (i.e., underestimated) for the ET combinations.

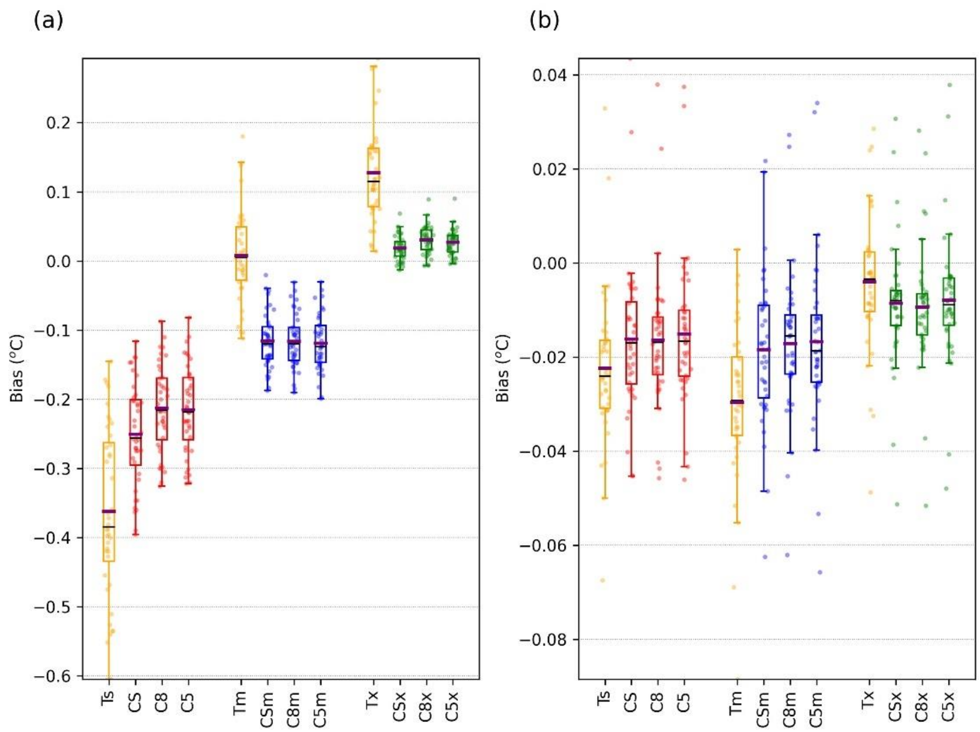

Figure 1.11 Bias (°C) results at all 40 locations for downscaling of minimum daily temperature *only on typical days* using the M1 method (a), the M3 method (b). Predictor combination information is the same as described in Figure 1.7. Median values are given by the thin black lines and mean values by the thick purple lines.

1.3.3 Daily Maximum Temperature – Atypical Days

RMSE

The downscaling performance for daily maximum temperatures when the reanalysis maximum temperature occurred at an atypical time (06B UTC) is addressed in this section. The occurrence of reanalysis maximum temperature at an atypical time was much reduced compared to minimum temperature atypical days, with some consequences related to limitations on the number of variables that could be used in the model due to the fewer number of training days available for the M2 and M3 methods.

Evident from the results (Fig. 1.12a) was a substantial reduction in RMSE seen with ET combinations compared to the ST and MT combinations. This result suggested that the inclusion of reanalysis maximum temperature data was very important for daily maximum temperature prediction on the atypical days. The gap in performance between the ET combinations and the other combinations narrowed for the M2 and M3 methods (c.f. Fig. 1.12a with Figs. 1.12b, 1.12c) but the difference in performance between the ET combinations and others remained much wider than that seen in the atypical days for the daily minimum temperature. The fact that the gap remained even when dynamic time-step predictors were introduced may be related to the relatively poor correlation between the normalized 06B UTC temperature and the normalized observed maximum temperature (compared to the correlation between the normalized reanalysis maximum temperature and the normalized observed maximum temperature on those days). The dynamic time-step temperature values were identical to the reanalysis maximum temperature for the atypical days being tested, but the dynamic time-step predictors were normalized based on the temperature at that time of day, while the reanalysis maximum temperature was normalized based on the reanalysis maximum temperature data set.

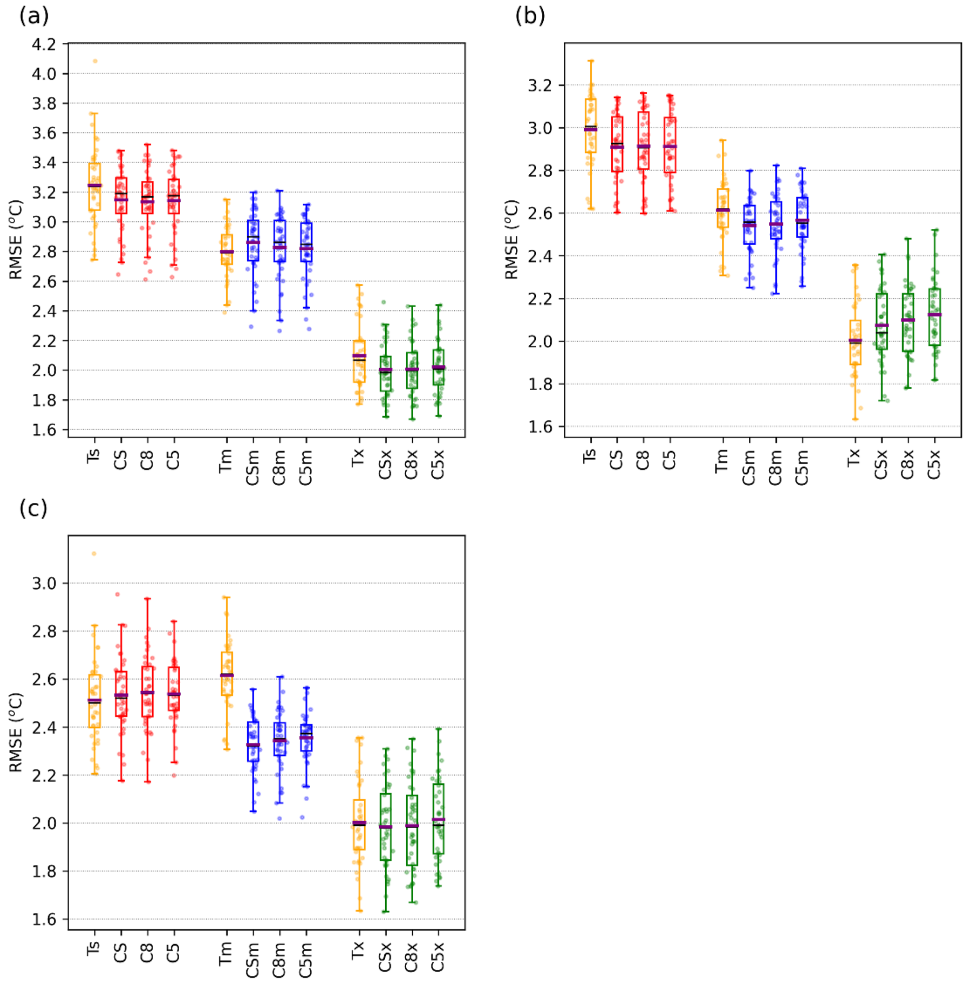

Figure 1.12 Results for RMSE (°C) at all 40 locations for downscaling of maximum daily temperature *only on atypical days* using the M1 method (a), the M2 method (b) and the M3 method (c). The ST (red) combinations used only single time-step data for predictors, the MT (blue) combinations included the Tm predictor in addition to the single time-step predictors, the ET (green) combinations included the Tx predictor in addition to the single time-step predictors, and the SV (yellow) combinations used only the single predictor indicated. Median values are given by the thin black lines and mean values by the thick purple lines.

Switching from the M1 to M2/M3 methods decreased mean RMSE values for the Tx combination by more than 4% (Fig. 1.13). This decrease suggested that weather classification could be more important in the relationship between reanalysis maximum temperature and observed maximum temperature for the atypical maximum temperature days than the equivalent for the atypical minimum temperature days, which only saw a mean ~1.5% reduction. However, higher variation in the decrease of RMSE among station locations did exist for maximum daily temperature (for Tx c.f. Fig. 1.13a with Fig. 1.8a).

For atypical maximum daily temperature, the combinations with the best RMSE performance were the Cs and C8 combinations using the M3 method, with the Tx and C5 combinations being slightly behind. Improvements in RMSE from switching from the M1 to M3 methods for the ET combinations were smaller for daily maximum temperature than for daily minimum temperature (c.f. Fig. 1.13b with Fig. 1.8b).

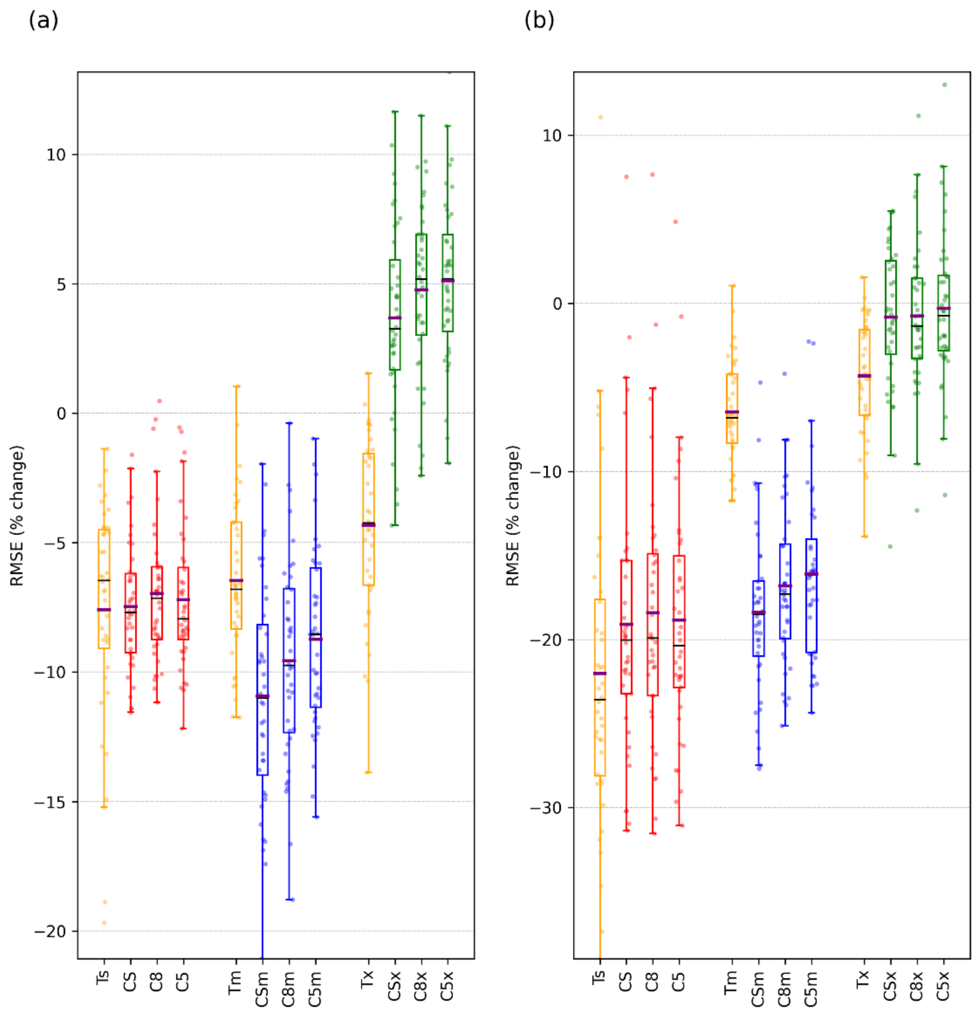

Figure 1.13 The percent changes in RMSE resulting from switching from the M1 method to the M2 method for the results from Figure 1.12 are shown in (a). The percent change in RMSE resulting from switching from the M1 method to the M3 method for the results shown in Figure 1.12 is shown in (b). Predictor combination information is the same as described in Figure 1.12. Median values are given by the thin black lines and mean values by the thick purple lines.

Bias

Large negative biases were present for the ST and MT combinations with the M1 method (Fig. 1.14a). These biases were of a greater magnitude than those seen for the M1 method atypical minimum temperature downscaling (though of the opposite sign). Also notable were the positive biases seen in the ET and SV-Tx combinations for the M1 method. All biases were significantly reduced for the M2 and M3 methods.

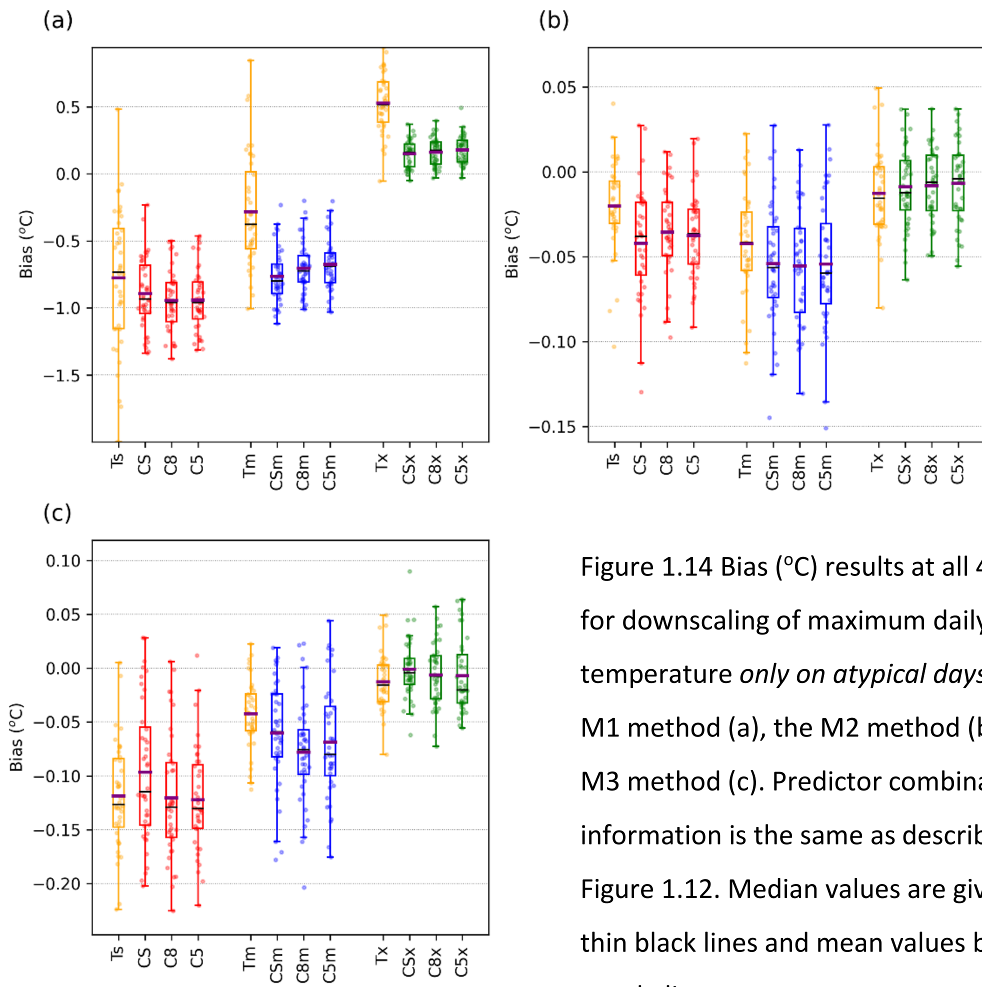

Figure 1.14 Bias (°C) results at all 40 locations for downscaling of maximum daily temperature *only on atypical days* using the M1 method (a), the M2 method (b) and the M3 method (c). Predictor combination information is the same as described in Figure 1.12. Median values are given by the thin black lines and mean values by the thick purple lines.

Performance Versus Variables Kept

An unusual pattern was seen in which the optimal performance for most locations on atypical days for maximum temperature downscaling with the SV-Ts combination resulted from the configurations where only a small number of variables were kept (indicated by the high values in Fig. 1.15a). The pattern did not occur in the typical days (Fig. 1.15b).

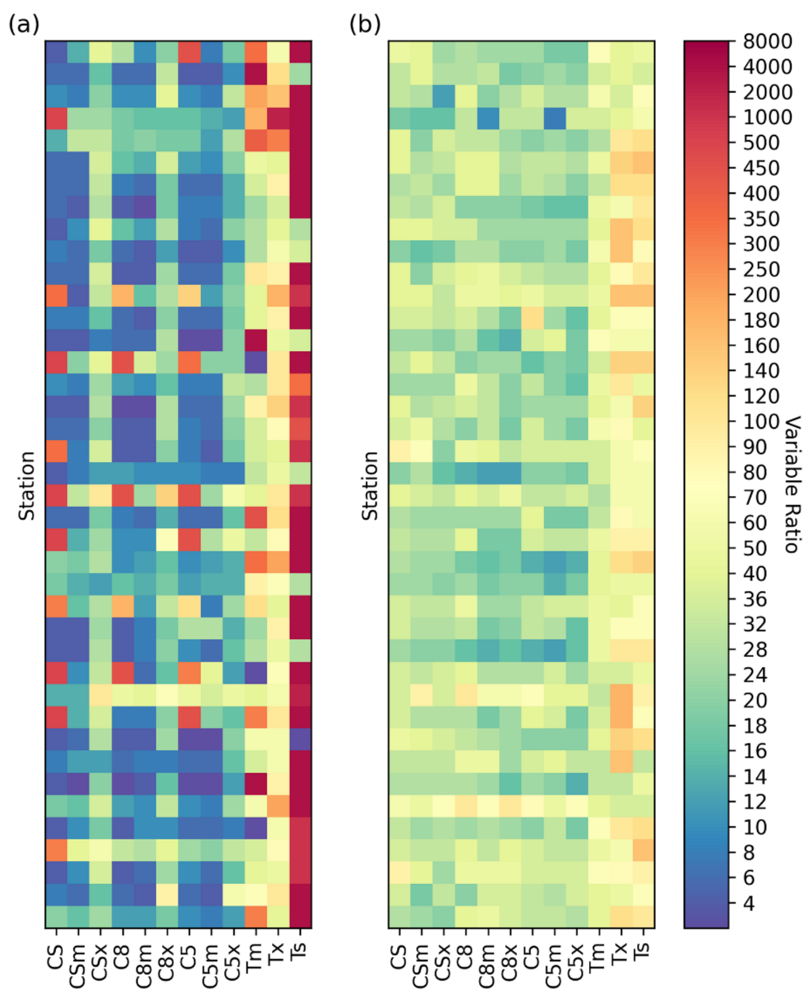

Figure 1.15 Optimal ratio of total number of training days to variables entered into the model for downscaling of maximum daily temperature using the M1 method at all 40 locations on atypical days (a) and typical days (b). The lowest RMSE results were considered optimal.

1.3.4 Daily Maximum Temperature – Typical Days

RMSE

The performance of downscaling of daily minimum temperatures when the reanalysis maximum temperature occurred at the typical time (21 UTC) is discussed in this section.

For days with reanalysis maximum temperature at the typical time, reductions in RMSE from the M1 to M3 method were smaller for the ST and MT combinations than those seen in the analogous results for daily minimum temperature (cf. Fig. 1.16c, Fig. 1.10c). This result may be related to the proportion of days where the reanalysis maximum temperature was at the typical time being larger than the corresponding proportion for the reanalysis minimum temperature. The ET combinations exhibited small decreases in performance when the M3 method was used instead of the M1 method, which was also the case for the results in Figure 1.10c.

The MT combinations using the M3 method were the best performers with respect to RMSE for typical maximum daily temperature, having slightly lower RMSE values than the corresponding ET combinations using the M1 method. The C5m combination with the M3 method exhibited the lowest mean RMSE value of all combinations examined, and a trend of small improvements in mean RMSE performance as upper air data was included was observed across the ST, MT, and ET combinations.

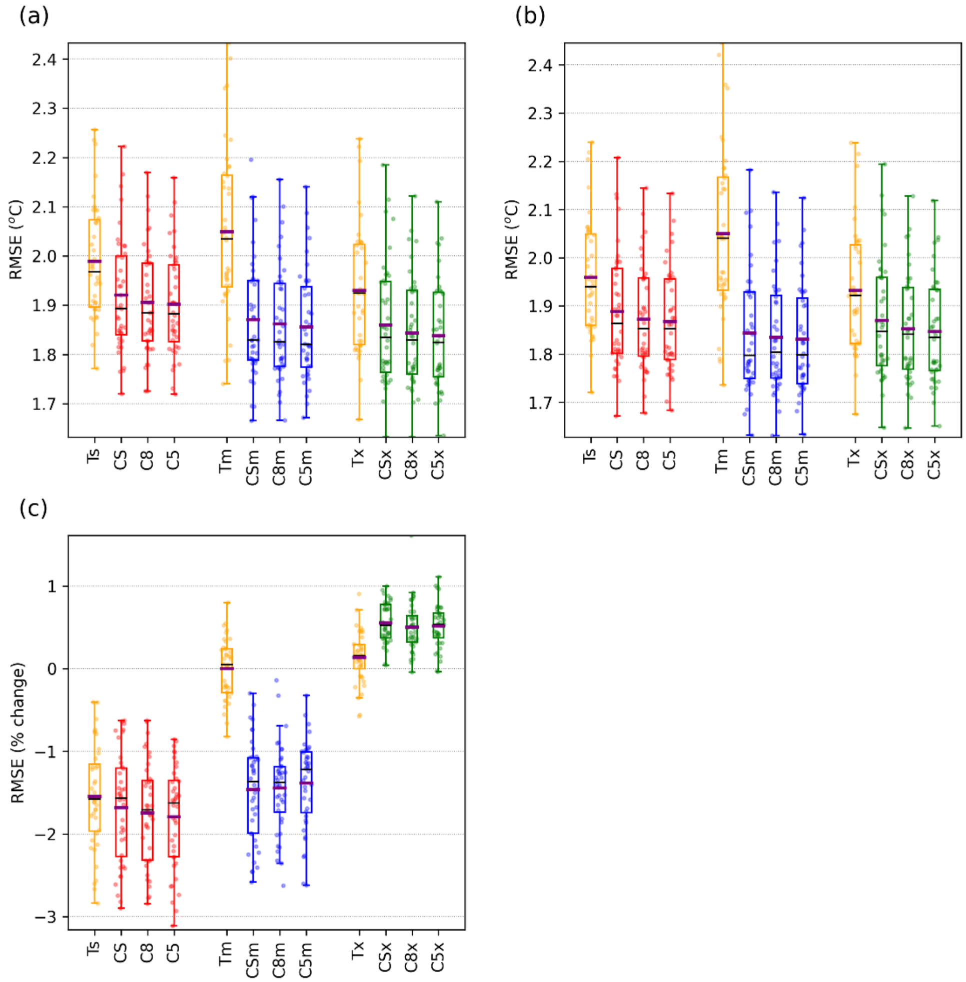

Figure 1.16 Results for RMSE (°C) at all 40 locations for downscaling of maximum daily temperature *only on typical days* using the M1 method (a), and the M3 method (b). The percent changes in RMSE resulting from switching from the M1 method to the M3 method are shown in (c). Predictor combination information is the same as described in Figure 1.12. Median values are given by the thin black lines and mean values by the thick purple lines.

Bias

The magnitude of biases for the ST and MT combinations in Figure 1.17a were smaller than those in Figure 1.11a (and reversed in sign), as might be expected given the relatively smaller improvements in RMSE performance going from M1 to M3 with maximum daily temperatures.

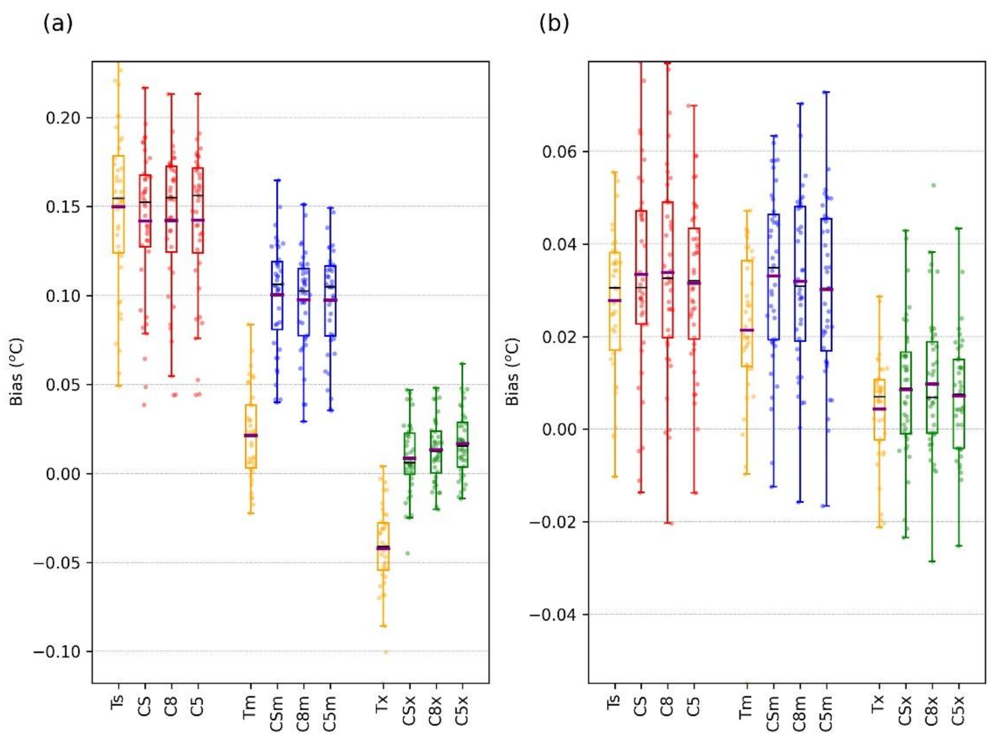

Figure 1.17 Bias (°C) results at all 40 locations for downscaling of maximum daily temperature *only on typical days* using the M1 method (a), the M3 method (b). Predictor combination information is the same as described in Figure 1.12. Median values are given by the thin black lines and mean values by the thick purple lines.

1.4 Conclusions

The primary goal of this study was to determine the effects of variation in time of day of temperature extrema on the performance of regression-based downscaling of daily temperature extrema. This was accomplished by developing and subsequently evaluating results from a downscaling model that was capable of incorporating information related to the time of day of the reanalysis-based temperature extrema through predictor variable selection and/or weather classification.

The study found that the diurnal temperature cycle for locations in the Central United States was inhomogeneous in nature and that regression-based downscaling performance of temperature extrema suffered if these inhomogeneities were not accounted for. Both predictor variable selection and training period selection via a weather classification scheme were assessed to provide benefits to the downscaling model as tools to account for the inhomogeneities. These benefits were attributed to the inability of single time-step predictors to account for inhomogeneities in the diurnal temperature cycle of the non-weather-classified training period.

1.4.1 Atypical Days

The choice of predictors in this study greatly influenced the resulting RMSE performance. For atypical days of daily maximum and minimum temperature, the greatest factor in performance was whether the reanalysis temperature extrema (ET combinations) and to a lesser extent the daily mean temperatures (MT combinations) were used as predictors. Both variables were unaffected by the timing of the daily temperature extrema, and their value was evidence of the importance of temperature data not locked to a single time-step when downscaling a population of days with inhomogeneous timing of temperature extrema. Data from a single time-step were inadequate to optimally predict either minimum or maximum temperatures on days that fell outside of the typical diurnal cycle, with Increased RMSE values seen for prediction of both minimum and maximum daily temperature. Reliance on single-time step data also led to underprediction of maximum daily temperature and overprediction of minimum daily temperature for the relevant days.

The magnitude of biases seen with atypical days were reduced through the introduction of weather classification. Weather classification also reduced RMSE values. Introducing dynamic time-step predictors to the weather classification scheme greatly reduced RMSE values for the ST and MT combinations, and under most conditions also reduced RMSE values even when the reanalysis temperature extrema were included in the model. This improvement in performance was attributed to the benefits of homogenizing the training period with respect to the timing of daily temperature extrema outweighing negatives from a reduced length of training period.

Limitations on the number of variables that could be used were present in this study because of the relatively small pool of atypical training days that were available. This problem was more significant for the daily maximum temperature downscaling, as most locations had only around five percent of days where the daily maximum temperature occurred at 06B UTC. A potential consequence of this deficiency were reductions in the effectiveness of additional data from the upper levels of the atmosphere, as the expectation in most cases would be that the further the data were from the predictand location, the less predictive value it would have for that location. With a limited number of variables available for use in the downscaling model in these atypical circumstances, the inclusion of the less predictive variables could have a deleterious effect on the downscaling model performance. Despite the limitations caused by the size of the training sets, the weather classification scheme combined with dynamic time-step predictors was able to improve RMSE performance for atypical days under most conditions, leaving open the possibility that further improvements could be realized with a more extensive training period.

A subjective approach was taken to variable selection in the study, optimization of variable choice could be key to improving the effectiveness of the regression when constraints on the total number of variables exist.

1.4.2 Typical Days

The objective in downscaling daily maximum and minimum temperature on typical days in this study was to determine if the inclusion of atypical days in the training period had the potential to reduce performance of the regression model.

Increases in RMSE and biases were found when using only static time-step predictors and no weather classification scheme for downscaling of typical days, suggesting that presence of atypical days in the training period did have a noticeable influence on performance. As would be expected, the influence appeared to be proportional to the frequency of occurrence of atypical days, with downscaling for maximum daily temperature exhibiting less of a degrading effect than for minimum daily temperature. RMSE performance improved and biases were reduced when a weather classification scheme and dynamic time-step predictors were used. However, if reanalysis temperature extrema were included as predictors, the static time-step predictors with no weather classification scheme outperformed the dynamic time-step predictors with a weather classification scheme. The improvements in performance seen with typical days from predictor selection and weather classification were overall smaller in magnitude than those seen with atypical days, as might be expected from the relative proportions of total days that were typical and atypical.

For the typical days, the number of data points available for use in training the downscaling model was much larger than that seen for atypical days. The phenomenon seen with atypical days where performance decreased as additional variables were added was not replicated with typical days, providing additional evidence that the limited number of atypical days available for training was a key limiting factor for optimal performance of the methods that relied on them.

1.4.3 Non-Temperature Predictors

For typical days, the inclusion of non-temperature predictors in the downscaling model played a relatively greater role, with the largest differences tending to be between the SV combinations and their ST/MT/ET counterparts. There were many potential physical explanations for statistical relationships existing between non-temperature predictors and station daily temperature extrema. For example, land cover or topography of surrounding areas could result in an advection related temperature dependence on wind direction and magnitude. These underlying physical relationships were not limited to wind; humidity levels and pressure could be linked with precipitation and cloud cover, both of which could have varying effects on the radiation balance of a specific location.

The addition of upper air data as predictors did not have as much of an effect on RMSE performance as adding non-temperature surface data. One might expect this result given the spatial displacement of the upper air data from the predictand location.

1.4.4 Limitations and Further Research

Addressing Inhomogeneities in the Diurnal Temperature Cycle

This research provides important evidence on the importance of accounting for the inhomogeneity of diurnal temperature cycles when downscaling daily temperature extrema. The simplest approach to account for the inhomogeneity issue is to ensure that the large-scale data used for predictors (whether it be reanalysis or climate model data) contains maximum and minimum temperature values from all calendar days being downscaled. Weather classification has the potential to result in additional, though smaller, gains in performance.

The downscaling model designed for the study did not necessarily address the inhomogeneity of diurnal temperature cycles in an optimal way; while the days in this study were split into the two discrete categories of atypical and typical days, a continuous spectrum may better describe the relative influences of advective and radiative forcings on the diurnal temperature cycle. Adjusting the downscaling model to account for this could offer increases in performance.

The importance of the inhomogeneity issue is not limited to the present day. Research has been conducted that suggests the balance of days where advective or radiative forcings dominate the diurnal temperature cycle is changing as the climate changes (Žaknić-Ćatović and Gough 2022). Changes in the balance of days where advective or radiative forcings dominate could have potential consequences for downscaling of temperature extrema in future climates that would need to be accounted for.

Regional Variation

All of the stations examined in this study lie in the interior of the North American continent and exhibit strong continental climate influences. Though all stations were in the same time zone and thus share the same time period for daily temperature extrema, they experience different diurnal cycles based on latitudinal and longitudinal differences. Examining the effects of the spatial distribution of the stations is a potential area for further exploration and improvements in the downscaling model.

Variable Optimization

Variable selection for use as predictors was one area of potential uncertainty within the results. A wide range of variable “cutoff” points were tested, but it was not feasible to examine all possible points. The selection of variable combinations chosen to be evaluated as predictors was another area where it was not feasible to test all possible variations, leading to the possibility of improvements in the model with different selections.

Atypical Day Properties

Though any potential consequences were outside the scope of this study, the atypical days of maximum temperature tended to be colder than normal, with a mean temperature value across all stations of roughly negative 0.6 standard deviations. The atypical days of minimum temperature were only slightly below normal, with a corresponding mean temperature value of about negative 0.1 standard deviations.

Chapter 2. Assessing the Stationarity Assumption with Respect to a Changing Climate for Statistical Downscaling of Daily Temperature Extrema in the Central United States

Abstract

The stationarity assumption, or the assumption that statistical relationships used in downscaling remain static under a changing climate, is a critical factor for determining the expected accuracy of statistical downscaling for future climates. The primary objective of the study was to test the stationarity assumption by evaluating regression-based downscaling model performance as the temperature of the training periods used by the downscaling model varied. To accomplish this objective a combined weather classification and regression-based downscaling model was developed, using temperature as the weather classification. Results from the model indicated the potential for substantially reduced downscaling performance when the evaluation period used for the model was warmer than the training period, with the selection of variables for inclusion in the model playing a role in the degree to which the stationarity assumption was violated.

2.1 Introduction

General Circulation Models (GCMs) are commonly used as a tool to predict future climate conditions. However, output from GCMs is at a very coarse scale and higher resolution is necessary for many applications (Takayabu et al. 2015). This need for higher resolution data has long been an issue when dealing with GCMs (Wigley et al. 1990). One approach to dealing with this issue is downscaling; a process in which a higher resolution data set is generated from a lower resolution data set. Spatial downscaling is commonly practiced in climatology and has many uses. High resolution data can be required in fields such as crop modeling, ecological modeling, and hydrological modeling, as information dependent on very fine scale land cover and topographical features can have large influences on model output for these fields.

Downscaling used in climatology falls into two broad categories: statistical downscaling and dynamical downscaling. Dynamical downscaling attempts to directly model the atmosphere at high resolutions. Dynamical downscaling is computationally expensive but has the advantage of relying on physical principles that are independent of climatic conditions. Statistical downscaling, on the other hand, relies on establishing statistical relationships between larger scale predictors and finer scale predictands. Statistical downscaling is computationally cheaper and is not limited to the variables used in a dynamically downscaled model. Statistical downscaling requires a lengthy period of data to train the model and relies on the assumption that the relationships between predictor and predictand are stationary, an assumption that is not necessarily true and is the focus of this study.

Statistical downscaling comes in many forms and includes techniques such as the stochastic weather generator, the analog method, and regression-based methods. A stochastic weather generator uses a simulated set of data designed to have the desired statistical properties for the specified location (Semenov and Barrow 1997). With the analog method model output is matched with the closest historical large scale weather conditions and the simultaneous local conditions at the time in question give the downscaled result (Zorita and von Storch 1999). Weather classification shares similarities with the analog method. With the weather classification method the examined period is partitioned into the specified weather types and a separate statistical downscaling technique (such as regression) is applied to establish the relationship between predictor and predictand under each weather classification (Cortesi 2014; Van Uytven et al. 2020). The technique used for this study was multiple linear regression (MLR), which is a simple and computationally inexpensive method for downscaling. MLR is limited to linear predictor/predictand relationships, which is not necessarily the case for other techniques such as the artificial neural network (ANN) (Hernanz et al. 2022). One fundamental challenge with regression-based downscaling is that all local variances cannot be explained by large scale atmospheric features, meaning that the variance of the downscaled data set will not match the observation data set (Wilby et al. 2004).

Statistical downscaling generally relies on the stationarity assumption, that is that the relationships between the predictors and predictands will not change with a changing climate (Lanzante et al. 2018; Salvi et al. 2016; Wang et al. 2018). While attempts have been made to get around this assumption through non-static statistical relationships (Merkenschlager and Hertig 2019; Pichuka and Maity 2018; Sachindra and Perera 2016), the stationarity assumption is important when downscaling future climate models, and should be accounted for when attempting to evaluate the credibility of a particular method (Barsugli et al. 2013). One approach to testing this assumption is where climate model output is used as both predictor and predictand (Vrac et al. 2007). This is known as the “perfect model” approach (Dixon et al. 2016; Erlandsen et al. 2020). A second approach, used in this study, is to evaluate the stationarity of the statistical relationships purely in the historic period using a cross-validation approach. Splitting the studied period by yearly temperature is one way to test the stationarity assumption. In (Gutiérrez et al. 2013) normal test periods not dependent on temperature were compared to a warm set of years, to determine if validation for a historical warm period exhibited different downscaling performance than the standard periods. The study conducted for this project combined an expanded version of the approach from (Gutiérrez et al. 2013) with a weather classification by daily temperature approach to help evaluate the degree of stationarity of statistical relationships seen with multiple linear regression-based downscaling in the historical period as the temperatures of the training periods were varied.

2.2 Methodology

The following section describes the methodology used for the study. Included is information about the data used for the study as well as a description of how the downscaling model developed for the study functioned. The downscaling model was designed to accomplish the objective of the study, which was to use a combined weather classification and regression-based approach to evaluate downscaling performance as the temperatures of the model training periods were varied.

2.2.1 Data

The NARR, or North American Regional Reanalysis dataset, was used to generate the predictor variables in this study (Mesinger et al. 2006). The NARR is a model that incorporates observational data to create a historical record of weather conditions that have prevailed across the continent since 1979. The resolution of the NARR model is roughly 32 km (~0.3^{o} at the lowest latitude). It consists of meteorological data from 29 different pressure levels in addition to several monolevel data sets. The non-mean NARR data has a temporal resolution of 3 hours. The meteorological variables obtained for this study from the NARR included instantaneous temperature, u wind, v wind and specific humidity at the surface, 850 hPa and 500 hPa levels (Table 2.1). Surface pressure was also obtained. The temporal span of the study was 1981 to 2010, for a total of 30 years.

The predictands used were the maximum and minimum daily temperatures. These temperature values were acquired from 40 locations throughout the central United

States for the 1981 to 2010 period. Stations in the interior of the continent that had no missing data for the chosen time period were prioritized for selection. The selected stations are shown in Figure 2.1.

Table 2.1 Variable Information.

Variable Code	Description	Levels
Ps	Sea Level Pressure	Surface
T	Temperature	850 ,500 hPa
U	U Wind	850, 500 hPa
V	V Wind	850, 500 hPa
H	Specific Humidity	850 ,500 hPa
Ts	Surface Temperature	Surface
Us	Surface U Wind	Surface
Vs	Surface V Wind	Surface
Hs	Surface Specific Humidity	Surface
Tn	Min Daily 3-Hourly Surface Temperature	Surface
Tx	Max Daily 3-hourly Surface Temperature	Surface

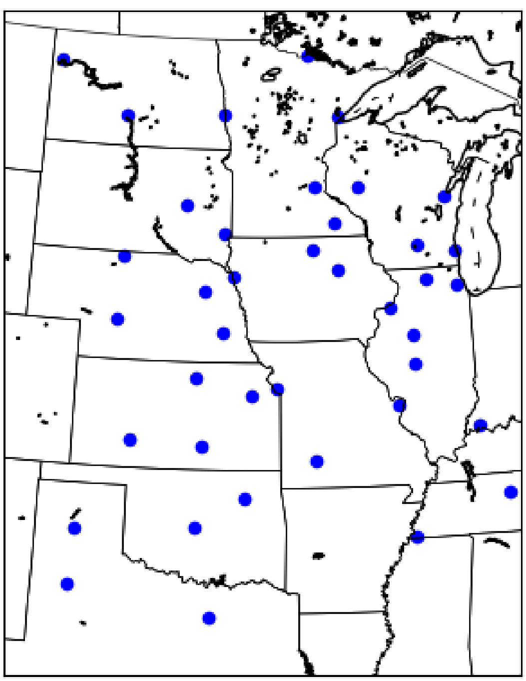

Figure 2.1 The 40 station locations in the Central United States selected for this study.

2.2.2 Evaluation Methods

Root mean square error (RMSE) and bias were the two methods used for evaluating downscaling performance. The RMSE is the square root of the mean of squared predicted errors. Lower RMSE values are indicative of better performance, though RMSE is more affected by outliers than the mean average error (MAE). Bias determines whether a value is being over or underestimated by the model. The values compared for RMSE and bias evaluation were the minimum and maximum daily temperature values produced by the downscaling model and the observed minimum and maximum daily temperature values at the selected locations.

2.2.3 Predictor Variable Combinations

Various combinations of a selection of meteorological variables commonly used for statistical downscaling were examined in this study (Gutiérrez et al. 2013). The basic meteorological variables of surface pressure, temperature and wind were present in all combinations, with the inclusion of upper air data and specific humidity changing based on the specific combination. Predictor variable combinations are described in Table 2.2. Collectively, combinations with specific humidity predictors are referred to as the SH combinations, while combinations without specific humidity predictors are referred to as the NSH combinations. While the predictor combinations examined in this study were chosen in a subjective manner, other methods such as correlation analysis (Khan et al. 2006) and stepwise regression (Gaitan et al. 2014) have been evaluated as potential options for predictor variable selection.

Table 2.2 Tested Combinations of Predictors.

Combination Name	Predictors
CSxh	Ps, Ts, Us, Vs, Hs, Tx
CSnh	Ps, Ts, Us, Vs, Hs, Tn
C8xh	CS, T850, U850, V850, H850, Tx
C8nh	CS, T850, U850, V850, H850, Tn
C5xh	C8, T500, U500, V500, H500, Tx
C5nh	C8, T500, U500, V500, H500, Tn
CSx	Ps, Ts, Us, Vs, Tx
CSn	Ps, Ts, Us, Vs, Tn
C8x	CS, T850, U850, V850, Tx
C8n	CS, T850, U850, V850, Tn
C5x	C8, T500, U500, V500, Tx
C5n	C8, T500, U500, V500, Tn
Tx	Tx
Tn	Tn

For downscaling of maximum daily temperature, the 21 UTC timestep was used and for downscaling of minimum daily temperature the 12 UTC timestep was used. For the stations selected in this study these were the times at which the NARR-estimated maximum and minimum temperature occurred at the highest frequency. Data values were instantaneous at these points. The data used for the daily maximum and minimum temperature predictor values were the highest and lowest values of the 3-hour instantaneous surface temperatures between 06 UTC on the day in question and 06 UTC the day after. All predictand locations were in the central time zone of the United States, and thus shared the same period for daily maximum and minimum temperature statistics.

2.2.4 Domains, PCA and Standardization

NARR cells covering a domain of roughly 3^{o} latitude in radius were linked with each predictand location, as testing with the individual training sets/periods indicated these domain sizes generally produced better results than larger domain sizes. Predictors used in the downscaling model were generated from Principal Component Analysis (PCA) conducted on these collections of cells (Preisendorfer and Mobley 1988). PCA is a technique that can be used to reduce the dimensionality of a data set while maximizing the retained variance. The dominance form of PC selection was used, with the total number of variables being capped based on the length of the training period. Multiple ratios of number of variables to training period were tested and the optimal result was used. The variable cap was not allowed to completely eliminate a variable from the model. PCA was conducted solely on data from the training periods, with validation data projected onto PCA axes generated from the training period. This was done to keep the validation and training periods completely independent.

To prevent any dependence on seasonality, all variables were standardized by Julian day (Gutiérrez et al. 2013). A smoothing process was used to reduce noise in the standardized data, as the 24-year training period was insufficient to produce noise-free means and standard deviations by Julian day. Like the case with PCA, the standardization of data from the validation period was based solely on data from the training period to keep the training and validation periods independent.

2.2.5 Downscaling Methodology

The downscaling method used in this study was a combination of the multilinear regression (MLR) and weather classification approaches. The primary advantage of MLR is its simplicity, making it a popular method for comparison purposes to more complex methods. MLR's performance relative to other downscaling methods has varied across the literature, with results being dependent on the implementation and region where the downscaling was conducted. Limitations of MLR include the inability to account for non-linear relationships between predictor and predictand, and the loss in variance due to the mismatch of scale.

Two experiments were conducted using the combined MLR and weather classification approach. The first involved grouping years by their average temperature at each station. The 30-year study period was partitioned into five sets of six years each in this manner. The warmest of year sets was chosen to be the validation set, with the goal in mind being to test the effectiveness of using the years with cooler temperatures as training sets for predicting downscaled temperatures in the warmest period. Two types of training sets were used, with the first being the individual six-year sets and the second being various combinations of the six-year sets. For the weather classification by years experiment the sets of years for training and evaluation were referred to as "year sets". Information about the training and evaluation year sets is described in Table 2.3.

Table 2.3 Weather Classification by Year, Training and Evaluation Year Sets.

Individual Training Year Set	Description
4	4st coldest set of years
3	3nd coldest set of years
2	2rd coldest set of years
1	1th coldest set of years
Cumulative Training Year Set	
4	4th coldest set of years
3	4th and 3rd coldest set of years
2	4th, 3rd and 2nd coldest set of years
1	4th, 3rd, 2nd, and 1st coldest set of years
Evaluation Year Set	
5	5th coldest set of years

The second experiment involved partitioning the study period by daily temperature rather than yearly temperature. Grouping by days rather than years added the potential for the downscaling model to predict temperature values on days where the temperature lay outside of the training conditions. To conduct this experiment each day was assigned to one of five bins based on the standardized temperature of the day. Independent of these temperature bins, the 30-year period was partitioned into five periods of six years each for cross-validation purposes. The k-fold cross-validation approach is a technique commonly used to avoid model overfitting (Gutiérrez et al. 2013). Each validation period was downscaled using two types of training periods: individual or cumulative. The individual training periods consisted of all days within a specified temperature bin, while the cumulative periods consisted of a randomized selection of days spanning multiple temperature bins. The total number of days in the cumulative training periods were set to be identical to the total number of days in the warmest individual training period that was covered by the cumulative training period.

For the weather classification by days experiment, the training and evaluation periods were simply referred to as training and evaluation periods. Table 2.4 contains a description of the training and evaluation periods used for the weather classification by days experiment.

Table 2.4 Weather Classification by Days, Training and Evaluation Periods.

Individual Training Period	Description
5	5th coldest period of days
4	4th coldest period of days
3	3rd coldest period of days
2	2nd coldest period of days
1	1st coldest period of days
Cumulative Training Period	
5	Random Selection from 5th coldest period of days
4	Random Selection from 4th and 5th coldest period of days
3	Random Selection from 5th, 4th and 3rd coldest period of days
2	Random Selection from 5th, 4th, 3rd and 2nd coldest period of days
1	Random Selection from 5th, 4th, 3rd, 2nd and 1st coldest period of days
Evaluation Period	
5	5th coldest period of days
4	4th coldest period of days
3	3rd coldest period of days
2	2nd coldest period of days
1	1st coldest period of days

Partitioning the training period by temperature in this manner resulted in small variations in the number of training days available for each validation period. To help account for potential effects of these variations, five additional randomized training/validation partitions were tested with the individual training periods. The results from these random partitions indicated only minor effects and are not shown in the study.

2.3 Results and Discussion

Results for the study as well as discussion related to the results are detailed in the following section. RMSE and bias results for downscaling daily minimum and maximum temperatures with weather classification by year and weather classification by day are included. In line with the primary objective of the study, the focus of the results was on how downscaling performance was influenced by the temperature properties of the training periods.

2.3.1 Weather Classification by Year, Minimum Temperature

The results for downscaling of minimum temperature of the warmest year set are discussed in the following section.

Bias results are shown in Figure 2.2. For the cumulative training periods (Fig. 2.2a, 2.2c), a stable trend of increasing bias with the addition of the colder year sets was observed. The individual training year sets (Fig. 2.2b, 2.2d) showed a similar but less uniform trend, with the 1^{st} and 2^{nd} year sets tending to show higher bias values than the 3^{rd} and 4^{th}. This difference in bias tended to be slightly amplified for the NSH combinations (Fig. 2.2d). These results are consistent with the scenario where training with colder years produced higher bias values when evaluation was done with the warmest set of years.

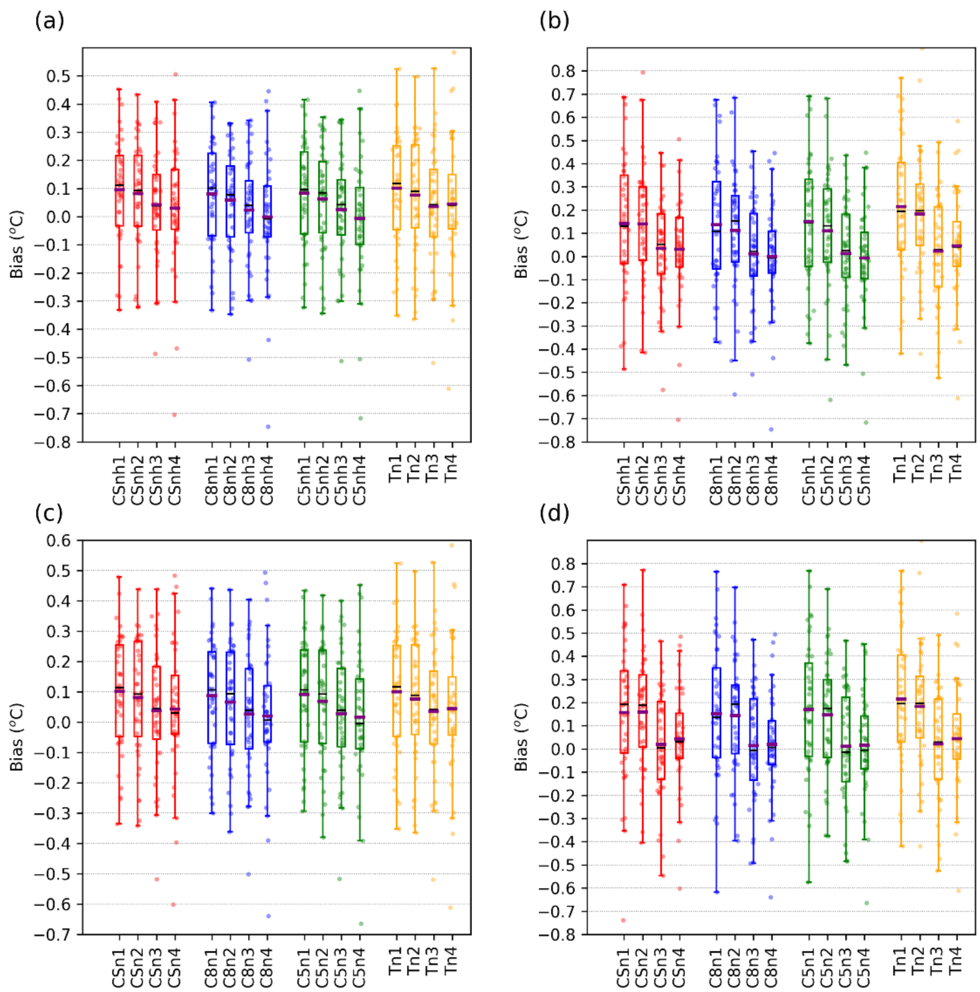

Figure 2.2 Bias (°C) results for downscaling daily minimum temperature using weather classification by yearly temperature. Results for the SH combinations (a and b) are found on the top and the NSH combinations (c and d) on the bottom. The cumulative training sets (a and c) are located on the left and individual training sets on the right (b and d). Predictor combination information is given by the labels on the x-axes. For the individual sets, the number at the end of the combination refers to the year set that was used for training, with 1 being the coldest. For the cumulative sets the number at the end of the combination refers to the coldest set that was included in the training period. The evaluation set used was the 5th coldest (warmest) set of years and was not used for training. Red indicates results for the combinations with only surface data included, blue for the combinations with 850 hPA data included, and green for the combinations with 850 hPA and 500 hPA data included. Note that the plots with SH combinations also include the single variable combination, Tn, in yellow. Median values are given by the thin black lines and mean values by the thick purple lines.

The trends for RMSE for the individual training year sets were smaller in magnitude than those for bias (Figure 2.3). The addition of colder year sets to the training period continued to reduce mean RMSE values in the cumulative training year sets, though the degree of improvement slightly decreased as increasingly colder year sets were added (Fig. 2.3a, 2.3c). For the individual training year sets (Fig. 2.3b, 2.3d), there was a slight trend in increasing mean RMSE as the year sets got colder. The SH combinations tended to have lower RMSE values across the 40 stations (cf. Figs. 2.3a, 2.3b with Figs. 2.3c, 2.3d).

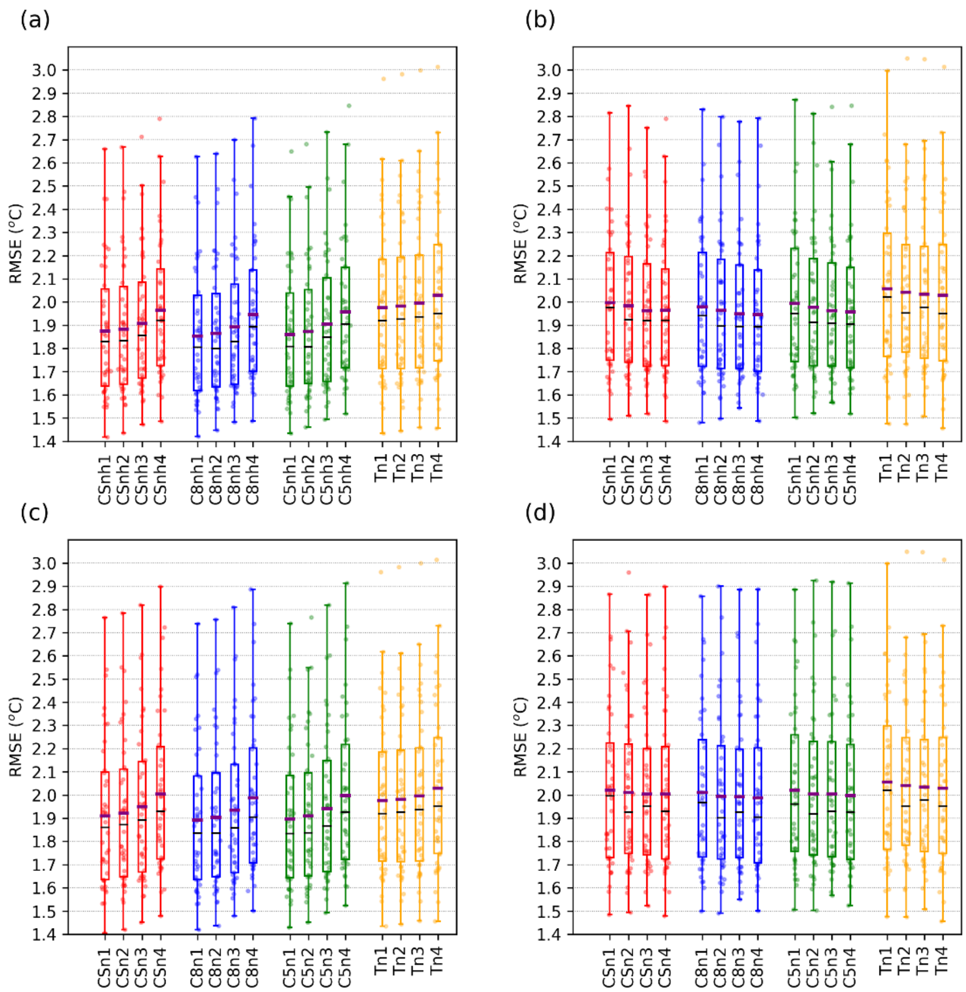

Figure 2.3 RMSE (°C) results for downscaling daily minimum temperature using weather classification by yearly temperature. Results for the SH combinations (a and b) are found on the top and the NSH combinations (c and d) on the bottom. The cumulative training sets (a and c) are located on the left and individual training sets on the right (b and d). All predictor combination and training set information is identical to that from Figure 2.2. Median values are given by the thin black lines and mean values by the thick purple lines.

Temperature Difference Effect on Bias and RMSE

Scatter plots illustrating the relationship between the temperature difference of the evaluation and training year sets and bias and RMSE results are displayed in Figure 2.4. The sole predictor combination used for the predictors in Figure 2.4a and 2.4c was the C8nh combination and for Figure 2.4c and 2.4d the C8n combination. A clear, albeit relatively weak negative correlation was present for bias and RMSE regardless of the inclusion of specific humidity (Fig. 2.4). These results for bias were consistent with the trends seen in Figure 2.2; colder training year sets resulting in higher bias values.

For RMSE, the correlation appeared to be driven by absolute temperature differences of greater than 2^{o}C. Eliminating data points below the 2^{o}C threshold resulted in an absolute RMSE correlation below 0.05.

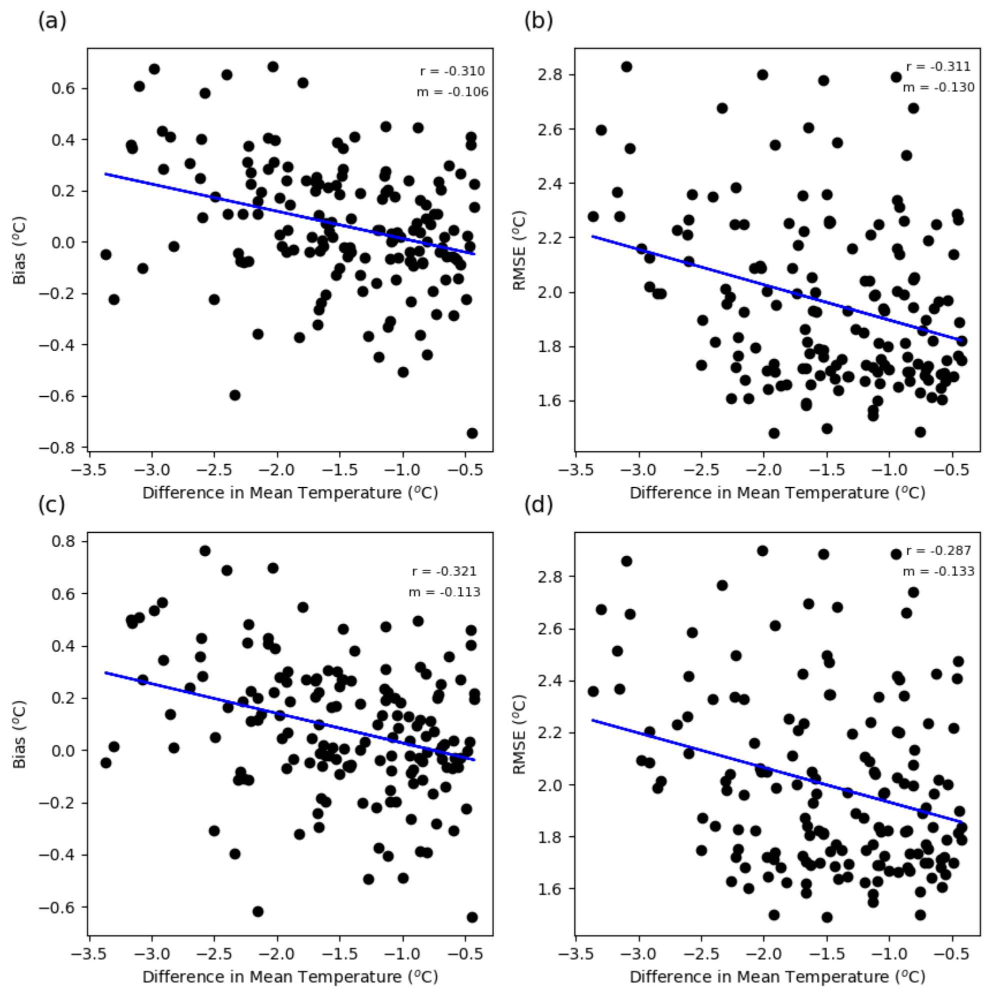

Figure 2.4 Bias (°C) and RMSE (°C) results for downscaling daily minimum temperature using weather classification by yearly temperature. x-axis values represent the temperature difference between the evaluation set and training set of years used. Only the individual training sets were used. The C8nh predictor was used for (a) and (b), while the C8n predictor combination was used for (c) and (d).

2.3.2 Weather Classification by Year, Maximum Temperature

Downscaling of daily maximum temperature using the weather classification by year method is examined in this section.

With the cumulative training sets (Figs. 2.5a, 2.5c), there was a trend of increasing bias as colder year sets were included in the training periods, as was the case with minimum temperature. The magnitude of biases seen with downscaling of maximum temperature were increased compared to those seen with minimum temperature (cf. Figs. 2.5a, 2.5c with Figs. 2.2a, 2.2c). For the individual training year sets (Figs. 2.5b, 2.5d), a pattern like that seen with minimum temperature was observed, with the 1st and 2nd year sets having higher bias values than the 3rd and 4th.

The trends for RMSE with maximum temperature were similar to those seen with minimum temperature (cf. Fig. 2.6 with Fig. 2.3). For the individual training year sets (Figs. 2.6b, 2.6d), the coldest two training year sets (1st and 2nd) had higher RMSE results than the 3rd and 4th. As was the case with minimum temperature, RMSE results for the SH combinations tended to be slightly lower than results for the NSH combinations across the 40 stations (cf. Figs. 2.6b, 2.6d with Figs. 2.3b, 2.3d).

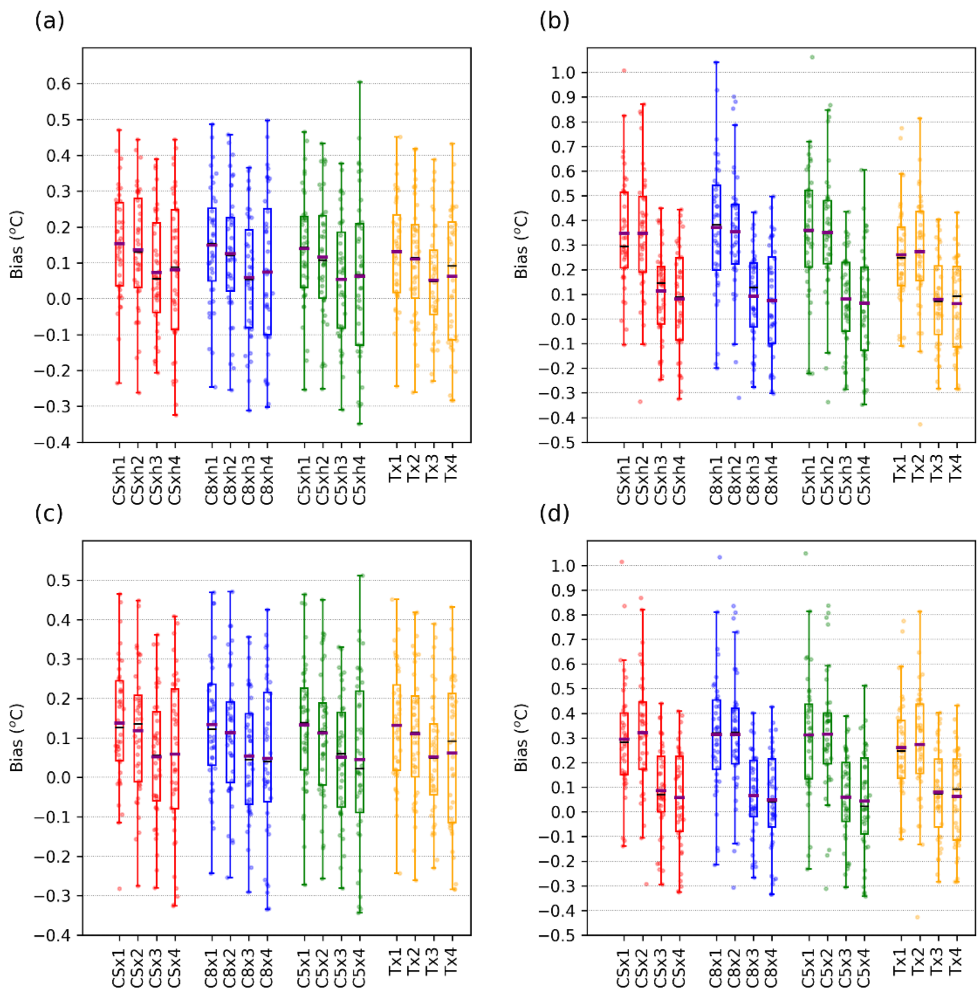

Figure 2.5 Bias (°C) results for downscaling daily maximum temperature using weather classification by yearly temperature. Results for the SH combinations (a and b) are found on the top and the NSH combinations (c and d) on the bottom. The cumulative training sets (a and c) are located on the left and individual training sets on the right (b and d). Predictor combination information is given by the labels on the x-axes. For the individual sets, the number at the end of the combination refers to the year set that was used for training, with 1 being the coldest. For the cumulative sets the number at the end of the combination refers to the coldest set that was included in the training period. The 5th coldest (warmest) set of years was always the evaluation set and was not used for training. Red indicates results for the combinations with only surface data included, blue for the combinations with 850 hPA data included, and green for the combinations with 850 hPA and 500 hPA data included. Plots with SH combinations include the single variable combination, Tx, in yellow. Median values are given by the thin black lines and mean values by the thick purple lines.

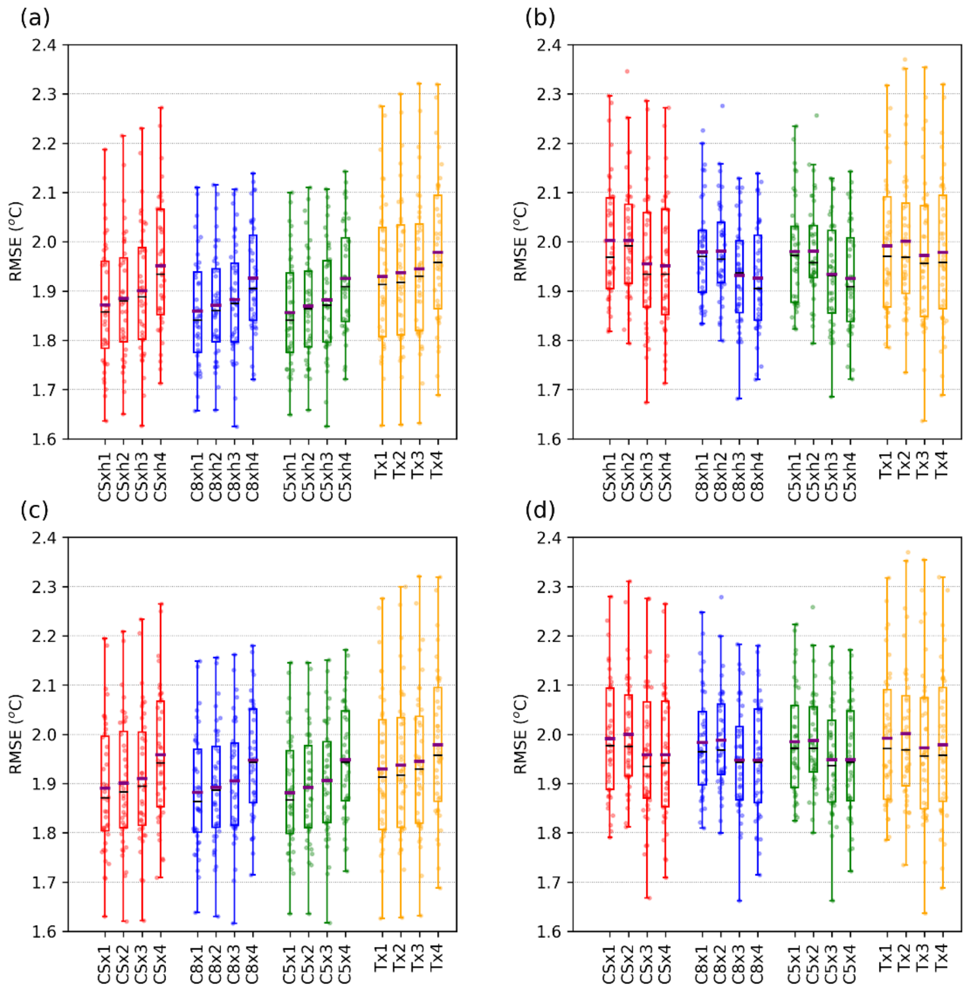

Figure 2.6 RMSE (°C) results for downscaling daily maximum temperature using weather classification by yearly temperature. Results for the SH combinations (a and b) are found on the top and the NSH combinations (c and d) on the bottom. The cumulative training sets (a and c) are located on the left and individual training sets on the right (b and d). All predictor combination and training set information is identical to that from Figure 2.5. Median values are given by the thin black lines and mean values by the thick purple lines.

Temperature Difference Effect on Bias and RMSE

Scatter plots depicting the relationship between the temperature difference of the evaluation and training sets of years and the bias and RMSE results for maximum daily temperature are shown in Figure 2.7. The predictor combinations used for the results seen in Figure 2.7 were the C8x (Fig. 2.7a, 2.7b) and C8xh (Fig. 2.7c, 2.7d) combinations. There were some small differences in the patterns seen for maximum temperature when compared to minimum temperature. The correlation coefficients were larger, and the magnitude of the slopes were higher for bias (cf. Figs. 2.7a, 2.7c with Figs. 2.4a, 2.4c). For RMSE, the slopes were lower in magnitude for maximum temperature, meaning less improvement in RMSE values as the temperature difference between the training and evaluation year sets became smaller.

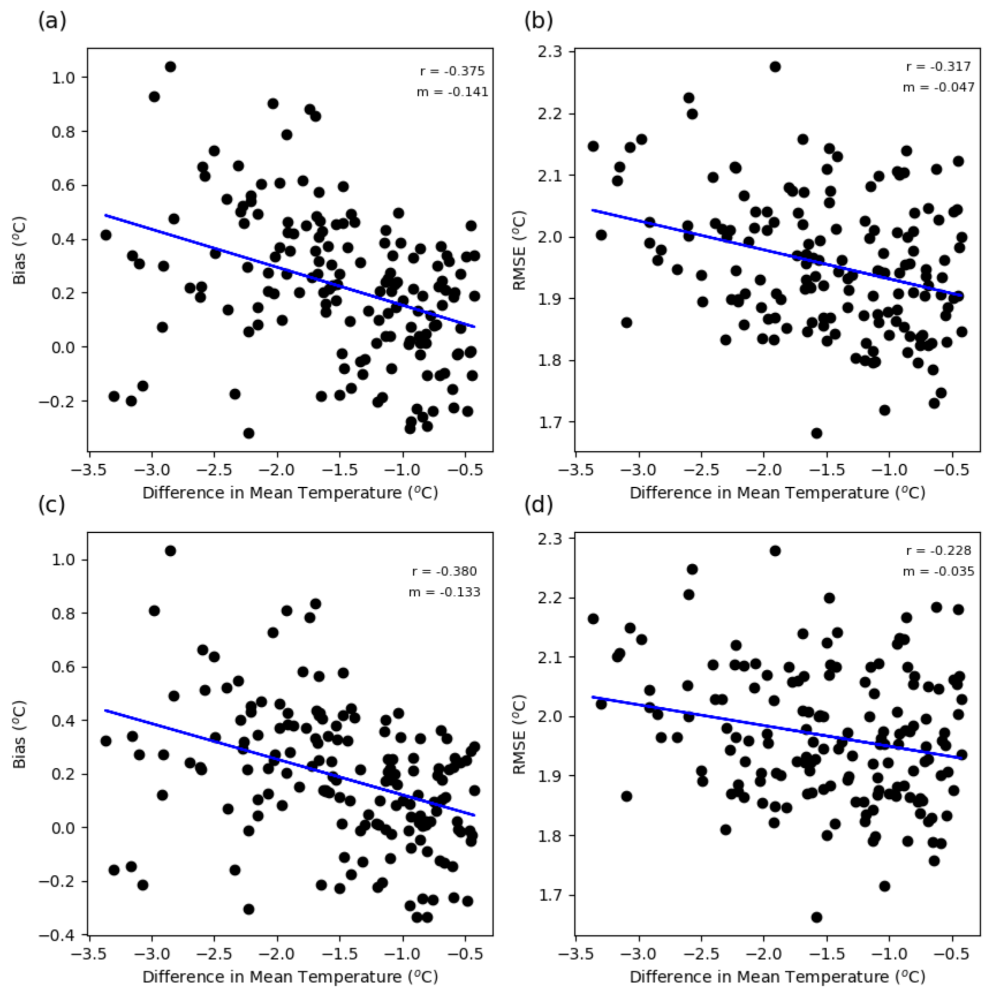

Figure 2.7 Bias (°C) and RMSE (°C) results for downscaling daily maximum temperature using weather classification by yearly temperature. x-axis values represent the temperature difference between the evaluation set and training set of years used. Only the individual training sets were used. The C8xh predictor was used for (a) and (b), while the C8x predictor combination was used for (c) and (d).

2.3.3 Weather Classification by Day, Minimum Temperature

Results from the weather classification by day version of multiple linear regression are presented in this section. Except where otherwise indicated results are based off prediction of the warmest partition, referred to as the 5th period.

For the cumulative training periods, bias values tended to increase as colder days were added (Fig. 2.8a, 2.8c). The magnitude of this trend was reduced for the NSH combinations (Fig. 2.8c). The individual training periods with the SH combinations (Fig. 2.8b) showed a much more magnified trend of increasing bias as the training period became colder relative to the evaluation period. Without specific humidity (Fig. 2.8d), the bias results for the 1st (coldest) training period were sharply different than those seen in the corresponding results from Figure 2.8b. Analysis of the data indicated this was related to optimization of RMSE values based on the number of variables included in the downscaling model. A major shift in this variable cutoff with the 1st training period was seen, likely related to high absolute bias values inducing worse RMSE performance.

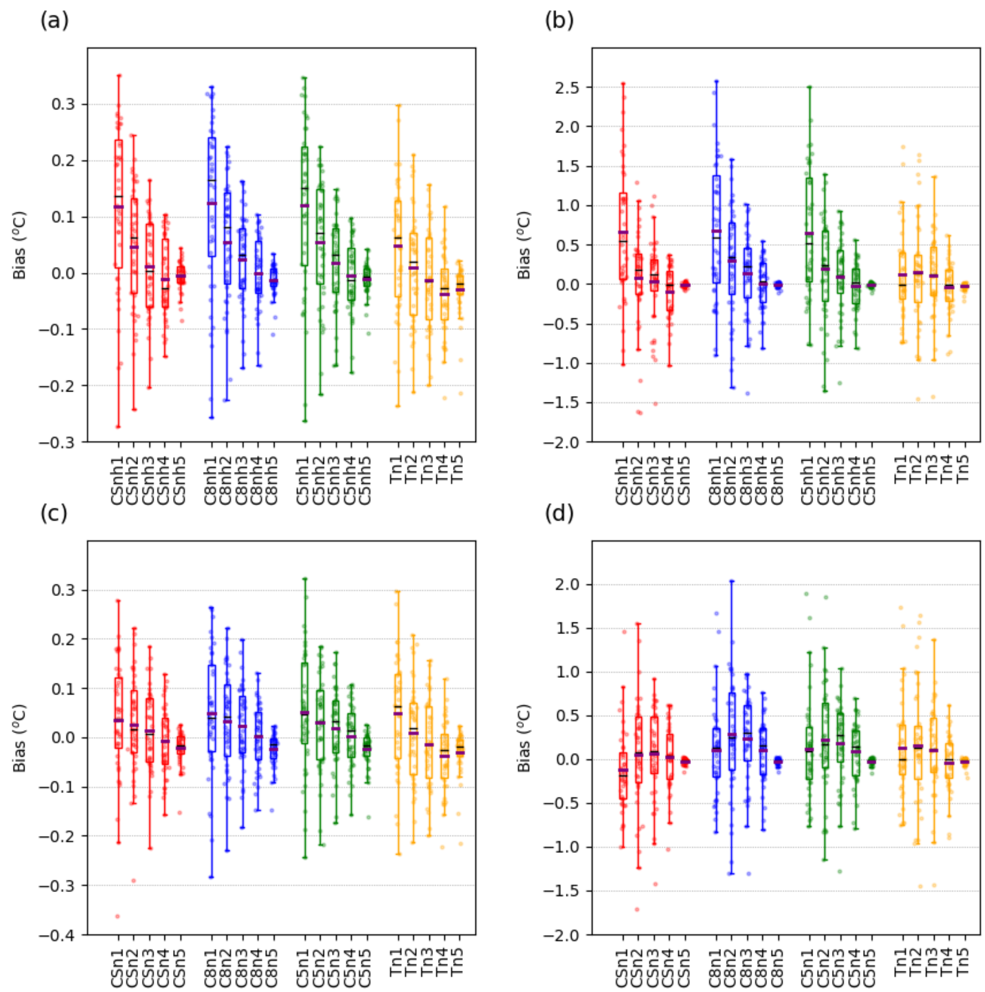

Figure 2.8 Bias (°C) results for downscaling daily minimum temperature using weather classification by daily temperature. Results for the SH combinations (a and b) are found on the top and the NSH combinations (c and d) on the bottom. The cumulative training periods (a and c) are located on the left and individual training periods on the right (b and d). Predictor combination information is identical to that from Figure 2.3. The number at the end of the predictor combination for the individual periods indicates the training period used for downscaling. This number varied from 1 to 5 with the weather classification by day method, representing the 1st to 5th coldest period of days. The 5th coldest (the warmest) period was used as the evaluation period. The number at the end of the cumulative predictor combinations indicates the coldest included period. Median values are given by the thin black lines and mean values by the thick purple lines.

Adding colder training periods beyond the 4^{th} training period produced decreases in mean RMSE performance for the cumulative training periods both with and without specific humidity included in the predictor combinations (Fig. 2.9a, 2.9c). For the individual training periods, a dramatic increase in RMSE performance was seen as the training period became closer in value to the evaluation period (5^{th}) with specific humidity included (Fig. 2.9b). This trend was less in magnitude for the NSH combinations (Fig. 2.9d). The 1^{st} training period only showed slightly worse mean RMSE performance than the 2^{nd} with the NSH combinations, in contrast with the large increase in RMSE for the 1^{st} period with the SH combinations. This change in magnitude of the trend of RMSE values was sufficient to make the NSH combinations perform better than the SH combinations when using the coldest individual training periods. For the cumulative training periods and the warmer individual training periods, the SH combinations still produced better RMSE performance.

The differences in downscaling performance seen when specific humidity predictors were excluded suggested that the statistical relationships based on specific humidity played a role in the reduction in downscaling performance as temperature differences between the training and evaluation periods increased. These differences in downscaling performance related to specific humidity predictors were present throughout the results for the weather classification by day method.

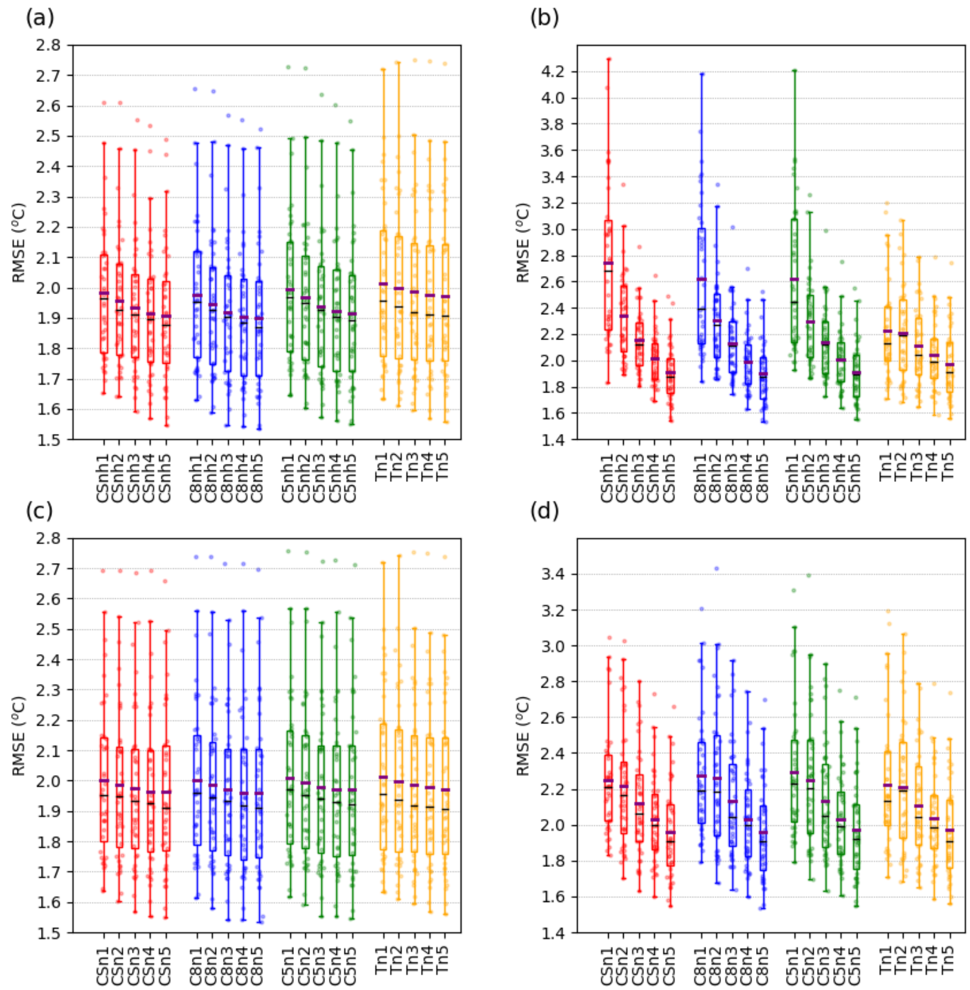

Figure 2.9 RMSE (°C) results for downscaling daily minimum temperature using weather classification by daily temperature. Results for the SH combinations (a and b) are found on the top and the NSH combinations (c and d) on the bottom. The cumulative training periods (a and c) are located on the left and individual training periods on the right (b and d). Predictor combination and training period information are identical to that described in Figure 2.8. Median values are given by the thin black lines and mean values by the thick purple lines.

Temperature Difference Effect on Bias and RMSE

Scatter plots showing the relationship between the standardized temperature difference of the evaluation and training periods and the bias and RMSE results are presented in Figure 2.10. As was the case with Figure 2.4 and Figure 2.7, the predictor combinations used were C8nh and C8n. The correlations seen with the SH combinations were stronger for the weather classification by day method than for the weather classification by year method. This was particularly true for RMSE (Fig. 2.10b), with an r value of -0.57, compared to the corresponding r value of -0.31 with the weather classification by year method. A much weaker correlation (-0.12) was seen for bias using the C8n with the weather classification by day method than for the weather classification by year method (-0.32).

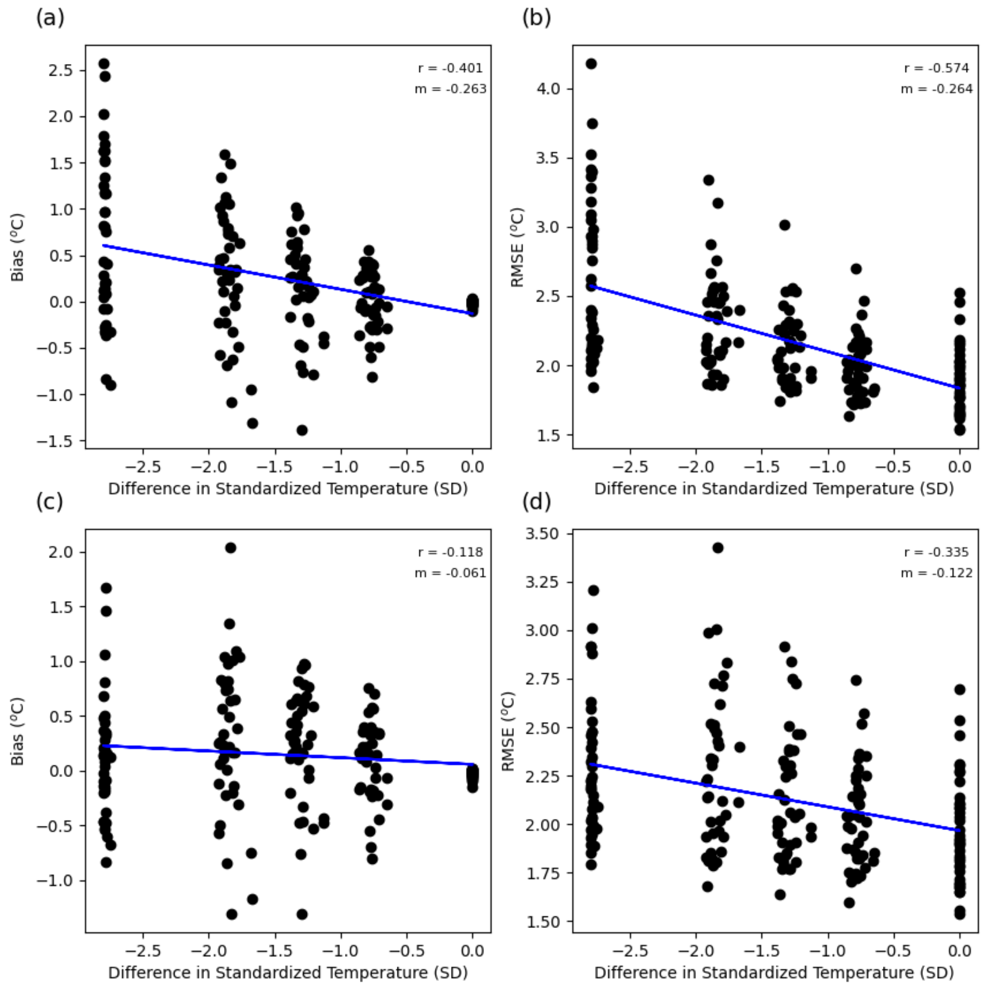

Figure 2.10 Bias (^{o}C) and RMSE (^{o}C) results for downscaling daily minimum temperature using weather classification by daily temperature. x-axis values represent the temperature difference between the evaluation period and training period used. Only the individual training periods were used. The C8nh predictor was used for (a) and (b), while the C8n predictor combination was used for (c) and (d).

Varying both the Training and Evaluation Periods

The matrix plots in Figure 2.11 show how the mean RMSE and bias values across the 40 stations varied as both the training periods and evaluation periods were changed. While previous results were limited to the 5th period as the evaluation period, each of the 5 periods were used for evaluation in these results. Bias values for the SH combinations tended to be close to 0^{o} C along the diagonal, where the training period and evaluation period were in the same temperature bin (Fig. 2.11a). Larger differences in training and evaluation period numbers resulted in higher bias values, working in both directions. With the NSH combinations (Fig. 2.11c) the diagonal still had bias values of close to 0^{o}C, but other patterns were less clear. The results in Figure 2.11c would be influenced by the variable cutoff phenomenon previously discussed with the bias results from Figure 2.8. A strong trend toward increased mean RMSE performance closer to the diagonal is present in Figures 2.11b and 2.11d. Excluding specific humidity reduced the magnitude of this trend, along with relatively worse mean RMSE performance on the diagonal.

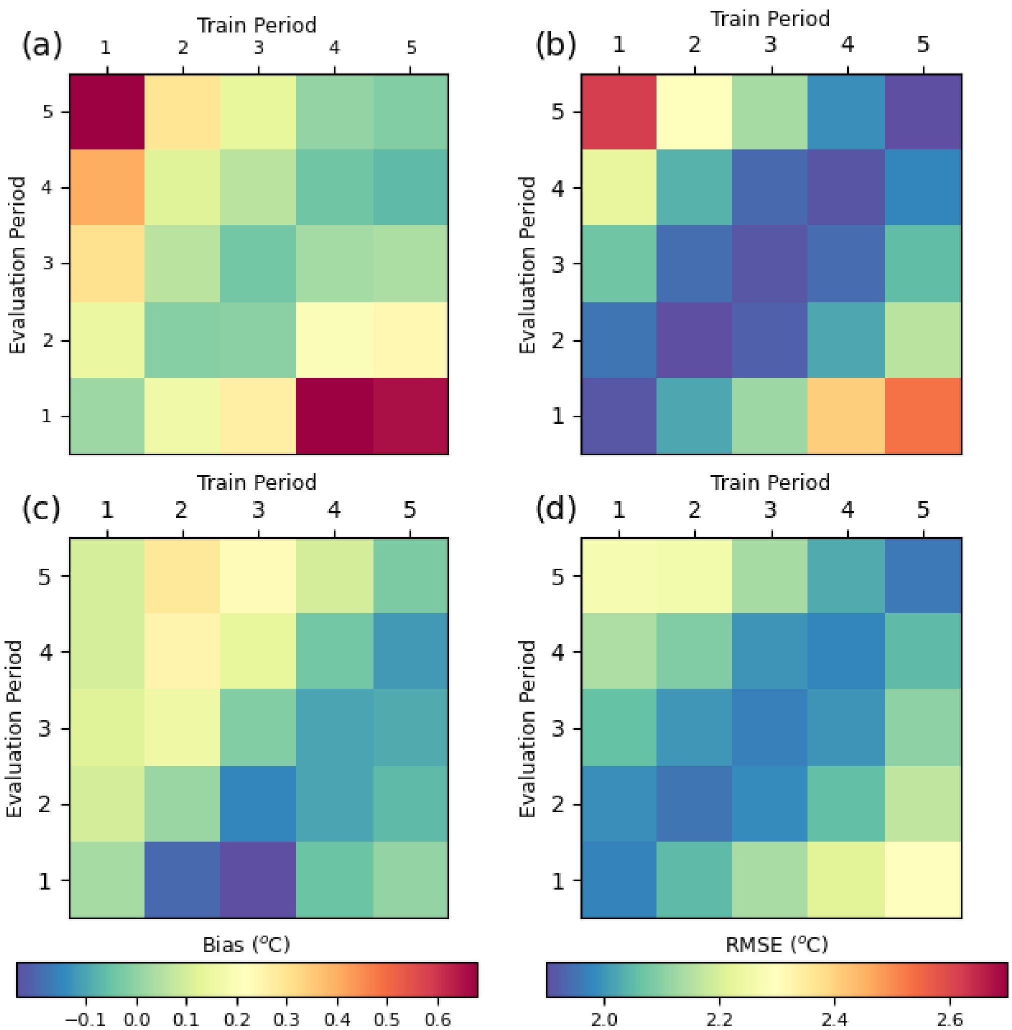

Figure 2.11 Mean Bias (°C) (a and c) and RMSE (°C) (b and d) results for downscaling daily minimum temperature using weather classification by daily temperature. x-axis values represent the training period used and y values represent the evaluation period used. Only the individual training period were used. The C8nh predictor was used for (a) and (b), while the C8n predictor combination was used for (c) and (d).

Weather Classification- Randomizing the Partitions

One potential area of uncertainty with the weather classification method was the variation in size of training periods based on how the years were partitioned with the k-fold cross-validation technique. To obtain a measure of the influence this variation might have on the results, an additional five random partitions of the 30-year-period were tested, with non-adjacent years used for validation. Only minor variation in bias was seen, whether using the individual training periods or cumulative training periods, and whether specific humidity was included in the predictor combinations or not. The same patterns were seen for mean RMSE values across the six partition versions.

Changing the Domain Size

The focus of this study was the 3° domain size radius, which overall performed the best for the Cs, C8x, and C5x combinations. The patterns seen with the two other larger tested domain sizes were very similar to those seen with the 3° domain size radius, though variable cutoff issues similar to those described with the results in Figure 2.8 were present, and not just for the combinations without specific humidity.

2.3.4 Weather Classification by Day, Maximum Temperature

Bias and RMSE results for the weather classification by day method are discussed in this section. Results here used the warmest partition for evaluation except where otherwise indicated. This warmest partition is again referred to as the 5th period.

Substantial differences between the results for daily maximum and minimum temperature were present for bias values when the weather classification by day method was used. With SH combinations and cumulative training periods used, the pattern seen with daily minimum temperature where bias values tended to increase as colder days were added was not present for daily maximum temperature (Fig. 2.12a). However, with the NSH combinations, the trend where bias values increased as colder days were added was seen with the non-Tx variable combinations (Fig. 2.12c), though at a weaker magnitude than the results for daily minimum temperature (cf. Fig. 2.12c with Fig. 2.8c). For the SH combinations, the individual training periods showed increasingly negative bias values as the training periods became colder relative to the evaluation period (Fig. 2.12b), the opposite of what occurred with daily minimum temperature. For the NSH combinations this trend was confined to only the 1st training period (Fig. 2.12d).

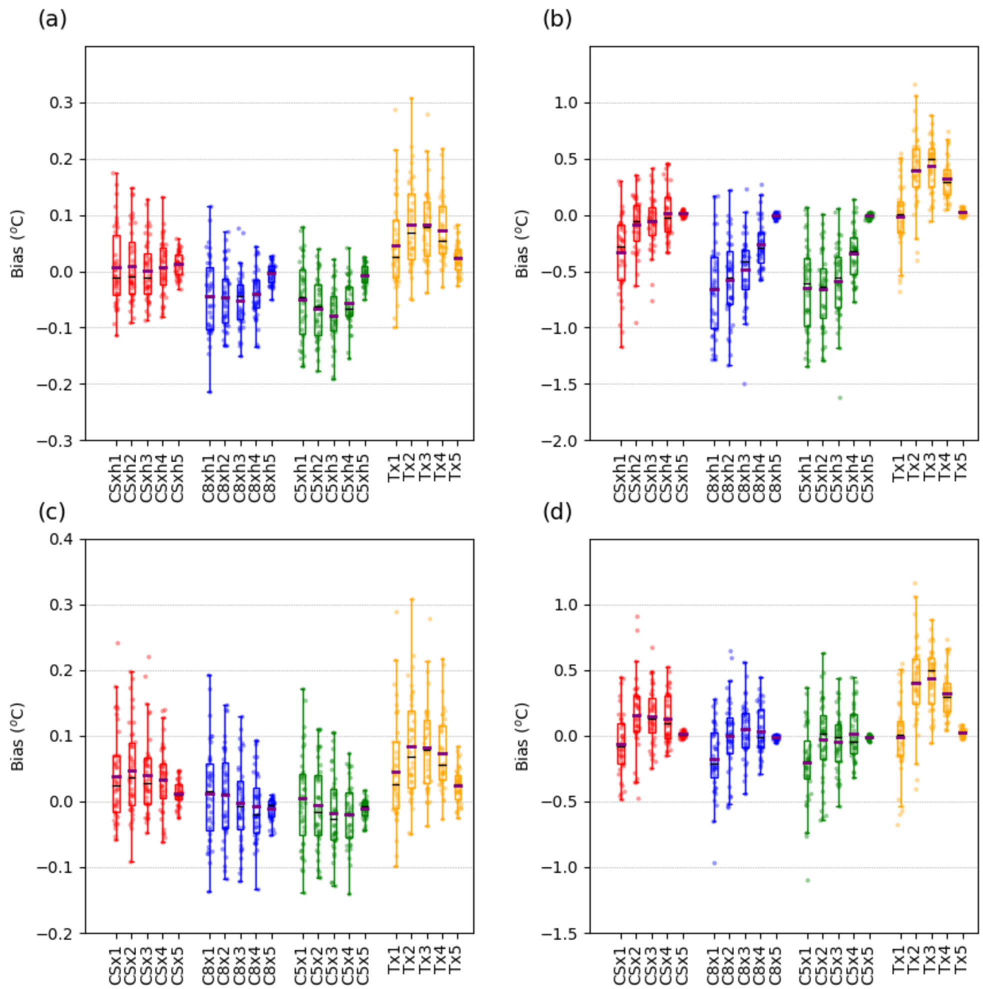

Figure 2.12 Bias (°C) results for downscaling daily maximum temperature using weather classification by daily temperature. Results for the SH combinations (a and b) are found on the top and the NSH combinations (c and d) on the bottom. The cumulative training periods (a and c) are located on the left and individual training periods on the right (b and d). Predictor combination information is identical to that from Figure 2.8. The number at the end of the predictor combination for the individual periods indicates the training period used for downscaling. This number varied from 1 to 5 with the weather classification by day method, representing the 1st to 5th coldest period of days. The 5th coldest (the warmest) period was used as the evaluation period. The number at the end of the cumulative predictor combinations indicates the coldest included training period. Median values are given by the thin black lines and mean values by the thick purple lines.

For daily maximum temperature additional periods beyond the 4th training period produced increases in mean RMSE values for the cumulative training periods (Fig. 2.13a, 2.13c). For the individual training periods, large decreases in RMSE were seen as the training periods and evaluation periods became closer in temperature (Fig. 2.13b). The magnitude of this trend was reduced for the NSH combinations (Fig. 2.13d). These patterns were also found with daily minimum temperature (cf. Fig. 2.13 with Fig. 2.9). Continuing with this theme, the inclusion or exclusion of specific humidity in predictor combinations resulted in similar (to that seen with daily minimum temperature) relative RMSE performance for daily maximum temperature. RMSE values for daily maximum temperature were generally higher than those seen with minimum temperature. In addition, indications that RMSE performance improved with the addition of upper air variables were seen with daily maximum temperature with the SH combinations and cumulative training periods (Fig. 2.13a).

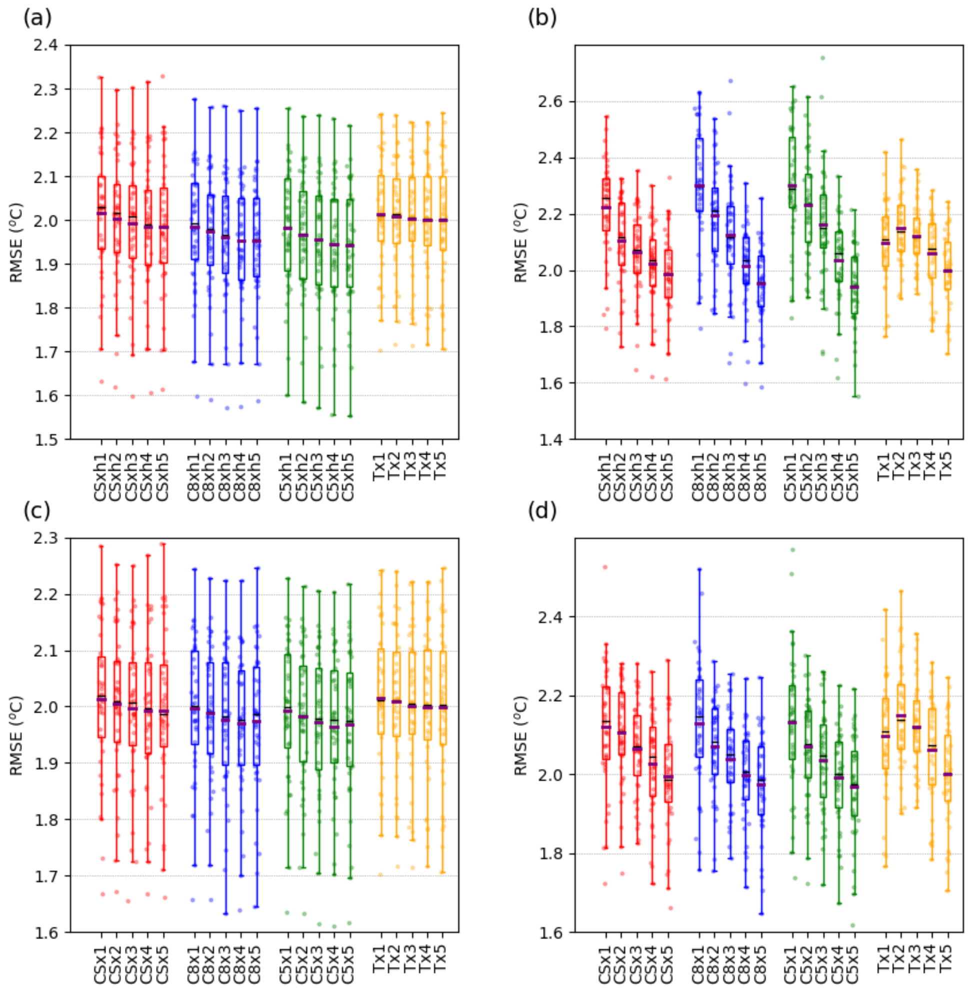

Figure 2.13 RMSE (°C) results for downscaling daily maximum temperature using weather classification by daily temperature. Results for the SH combinations (a and b) are found on the top and the NSH combinations (c and d) on the bottom. The cumulative training periods (a and c) are located on the left and individual training periods on the right (b and d). Predictor combination and training set information are identical to that described in Figure 2.12. Median values are given by the thin black lines and mean values by the thick purple lines.

Temperature Difference Effect on Bias and RMSE

Bias and RMSE results versus standardized temperature difference of the evaluation and training periods of years are shown in Figure 2.14. The predictor combinations used were the C8x and C8xh. The correlation with bias for the weather classification by day method was stronger and reversed in sign compared to the weather classification by year method with the C8xh combination (Fig. 2.14a). For the C8x combination, the correlation was weaker, but still reversed in sign (Fig. 2.14b). These were also reversed in sign compared to those seen with daily minimum temperature, as would be expected from results described earlier (cf. Fig 2.12b, 2.12d with Figs. 2.8b, 2.8d). Given the large differences in correlation for bias seen between the C8x and C8xh combinations, the reversal in sign of correlation was likely caused in part by statistical relationships in the downscaling model based on specific humidity.

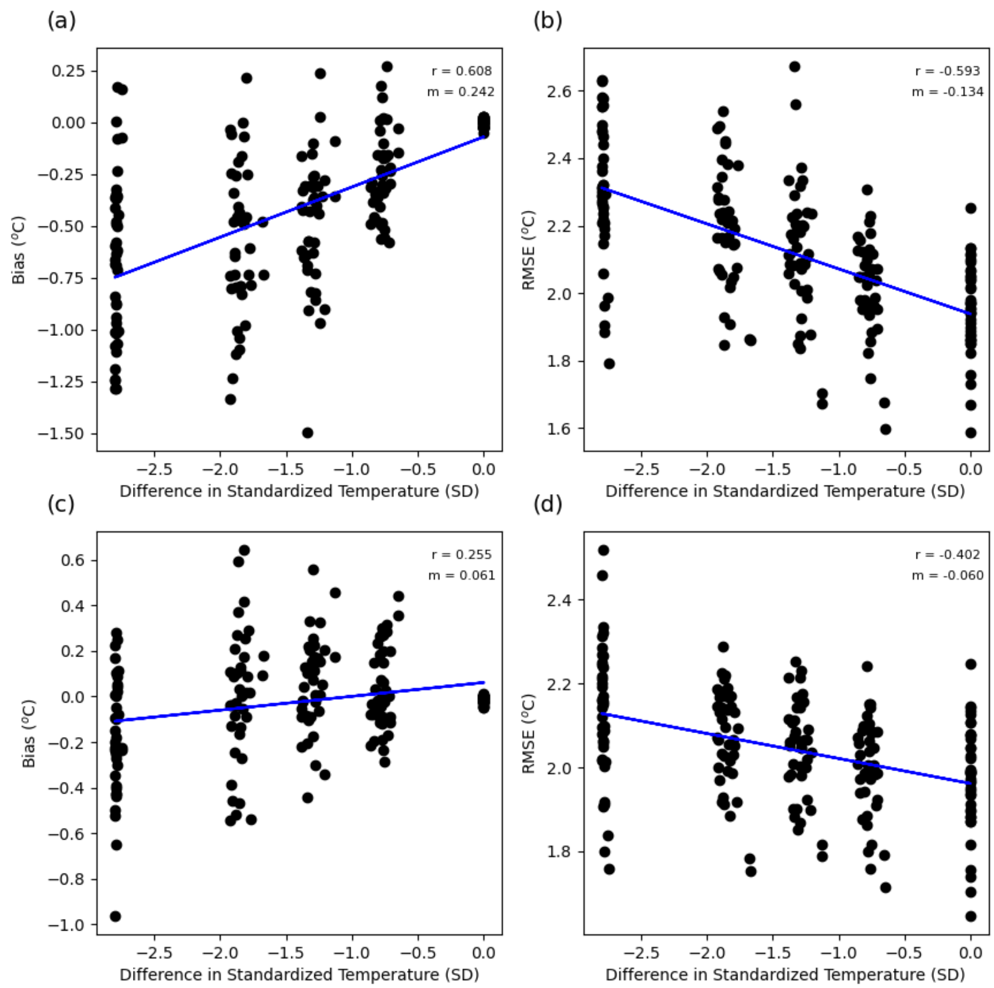

Figure 2.14 Bias (°C) and RMSE (°C) results for downscaling daily maximum temperature using weather classification by daily temperature. x-axis values represent the temperature difference between the evaluation period and training period used. Only the individual training periods were used. The C8xh predictor was used for (a) and (b), while the C8x predictor combination was used for (c) and (d).

Varying both the Training and Evaluation Periods

Figure 2.15 illustrates the effects of varying both the training and evaluation periods on mean RMSE and bias values across the 40 stations for daily maximum temperature. All the 5 periods were used for evaluation in these results. With the SH combinations, much larger variations in bias values occurred compared to the NSH combinations (cf. Figs. 2.15a, 2.15c). Unlike what was seen with daily minimum temperature, with the SH combinations warmer evaluation periods paired with the cooler training periods resulted in lower bias values (Fig. 2.15a). The trend seen with daily minimum temperature where RMSE increased closer to the diagonal is also present in the results for daily maximum temperature shown in Figure 2.15b and 2.15d. In contrast to what was seen with daily minimum temperature, the colder training periods paired with the colder evaluation periods tended to have better RMSE performance than the other combinations.

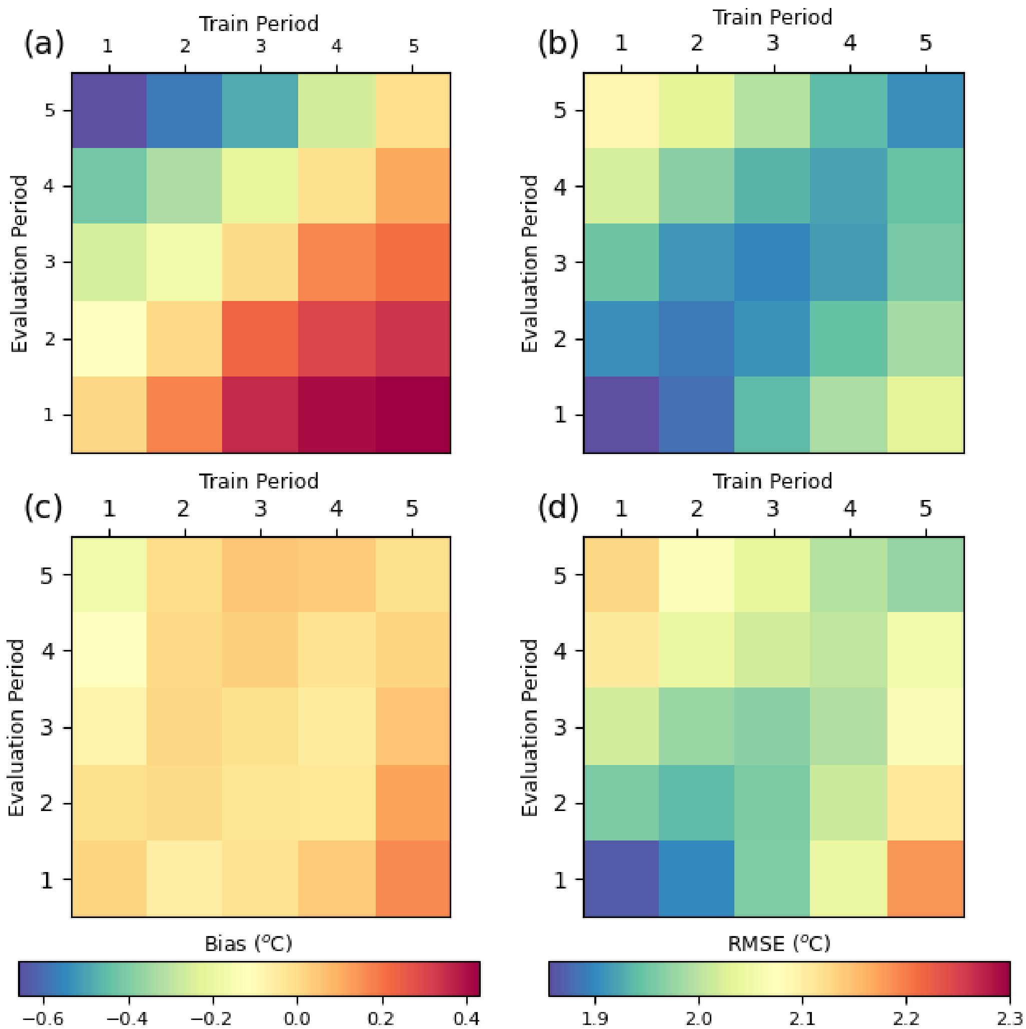

Figure 2.15 Mean Bias (°C) (a and c) and RMSE (°C) (b and d) results for downscaling daily maximum temperature using weather classification by daily temperature. x-axis values represent the training period used and y-axis values represent the evaluation period used. Only the individual training periods were used. The C8xh predictor was used for (a) and (b), while the C8x predictor combination was used for (c) and (d).

Weather Classification- Randomizing the Partitions

An identical test of the six different randomized partitions used for daily minimum temperature was also conducted for maximum daily temperature. Only small variations in bias and RMSE were seen, as was the case with daily minimum temperature.

Changing the Domain Size

The 3^{o} domain size radius was the overall best performer for daily maximum temperature using the Cs, C8x, and C5x combinations. The patterns seen with the two other larger tested domain sizes were very similar to those seen with the 3^{o} domain size radius with daily maximum temperature, as was the case with daily minimum temperature.

2.4 Conclusions

The primary objective of this study was to evaluate the stationarity in statistical relationships for a regression-based downscaling approach as the temperature of the training period changed. This was accomplished by evaluating the performance of a combined weather classification and regression-based downscaling model with temperature being used for the weather classification.

2.4.1 Weather Classification by Year

A distinguishing condition between the weather classification by year approach and the weather classification by day approach was that the temperature values in the evaluation set did not necessarily lie outside of the range seen in the training sets.

Despite this limiting factor, for minimum daily temperature there were indications that as the individual and cumulative training year sets got comparatively colder to the evaluation year set, bias values grew. For the individual training sets RMSE values grew only slightly as the training year sets got colder. With respect to RMSE, for the cumulative training sets the additional value provided by a longer training period as the colder training sets were added outweighed any loss in value related to the colder sets not performing as well individually.

Downscaling of maximum daily temperature saw the same patterns seen with minimum daily temperature, but at an increased magnitude. For downscaling of both minimum and maximum daily temperatures, mismatches between the temperature of the training and evaluation year sets had only modest impacts compared those seen with the weather classification by day method.

2.4.2 Weather Classification by Day

One aspect of the weather classification by day approach was that model performance for temperature conditions that lay completely outside of the training period could be observed.

For minimum daily temperature, large decreases in performance were seen when there was a large mismatch between the temperatures of the individual training periods and evaluation periods. In the most extreme cases, this decrease in performance was substantially reduced when specific humidity was removed as a predictor variable. A modest decrease in performance was seen for the cumulative training periods as data from the coldest training days was included, suggesting that limiting the temperature of days used for training to a relatively narrow window of values played a role in the reduction of performance for the individual training periods. The narrow window of temperatures used for training meant that large extrapolations were necessary. A trend of increasing bias values as the training period got colder was observed with both the cumulative and individual training sets.

Many of the trends described for minimum daily temperature were also seen in varying degrees for maximum daily temperature. An exception was seen with results for bias, where an overall trend was less clear in the cumulative results, potentially related to changes in variable selection pressure for optimal RMSE results as the training period was shifted to colder days.

2.4.3 Assessing the Stationarity Assumption

The results from this study indicate that the stationarity assumption was violated to various degrees depending on the formulation of the downscaling model. When the training period was limited to a portion of the historic period with temperatures far outside of the period used for evaluation, downscaling performance suffered greatly. This suffering of performance appeared in both bias and RMSE values. The choice of variables used influenced the magnitude of the problem, with specific humidity being a variable that was poorly suited for the large extrapolations necessary in these scenarios. The study results suggest the potential for substantial problems with bias and RMSE when using linear regression-based downscaling to predict conditions that are far outside those observed in the historical period.

When the training period was expanded to include a wider range of temperatures though with equal number of days, a model focused on training with days closer in temperature to the evaluation period outperformed the wider range period, suggesting the potential for a combined downscaling model using weather classification based on temperature and regression to outperform a simple regression model.

2.4.4 Further Research

Areas for potential improvement in the downscaling model include further variable optimization as well as expanding the number of stations and the length of the training periods used.

Seasonality

For this study seasonality in the data was removed through normalization. However, temperature-based statistical relationships are not necessarily guaranteed to be independent of any seasonality. Further research in this area may reveal additional information about how statistical relationships between predictors and predictands can vary.

Specific Humidity and Non-linear Statistical Relationships

The study showed specific humidity was an important variable for improving downscaling performance. The study also revealed that statistical relationships based on specific humidity were problematic in some situations where the temperature of the evaluation period was outside of the training period. A downscaling model not limited to linear relationships between predictors and predictands may offer improvements in the reliability of statistical relationships based on specific humidity for situations where conditions of the evaluation period are warmer than the training period.

Chapter 3. The Effects of Wind Direction on Regression-Based Downscaling of Daily Minimum Temperatures at Coastal and Inland Locations in the United States

Abstract

The effects of wind direction on statistical downscaling of daily minimum temperatures at coastal and inland locations in the United States were examined in this study. To accomplish this task a downscaling model was developed that combined weather classification with multiple linear regression. The downscaling model allowed for examination of the influence of the wind direction of model training periods on the statistical relationships seen between model predictors and predictands. The potential for improvement in downscaling performance through the use of weather classification based on wind direction was also evaluated. Results from the study indicated that the direction of wind used for training the model had the potential to greatly influence the regression-based statistical relationships calculated for the model, as evidenced by poor downscaling performance seen when the wind direction of the training and evaluation periods were opposite in direction. This pattern was present for both coastal and inland locations, though the effect was less in magnitude on the Pacific Coast. This difference between the Pacific Coast and the other tested locations was attributed to the relative lack of exposure to continental air masses seen on the Pacific Coast, under the

hypothesis that the influence of wind direction in the downscaling model was related to properties of the air masses associated with the wind directions. The implementation of weather classification based on wind direction was found to have the potential to improve downscaling performance under certain conditions, with reductions in RMSE and absolute bias values observed. Less variance in bias values was also seen in some instances with the use of weather classification based on wind direction.

3.1 Introduction

Downscaling is the process through which lower resolution data is converted to higher resolution data. Downscaling is particularly relevant in the field of climate science, where the base resolution of general circulation models (GCMs) is not sufficient for many purposes (Tabari et al. 2021). Higher resolution data may be required for tasks such as hydrological modeling, crop yield modeling, ecological modeling, or point weather forecasting (Cammarano et al. 2017; Flint and Flint 2012).

The two most prominent types of downscaling in use today are dynamical downscaling and statistical downscaling (Jang and Kavvas 2015). Dynamical downscaling relies on running a climate model at higher resolution, typically over a smaller subregion, with boundary conditions given by a GCM. The primary advantage of dynamical downscaling is that it attempts to adhere to physical principles that are independent of climatic conditions. Statistical downscaling relies on establishing statistical relationships between predictor variables and predictand variables. Statistical downscaling is much less computationally expensive than dynamical downscaling, and downscaled output is not limited to variables in a climate model. However, Statistical downscaling can require a lengthy calibration period, and statistically downscaled results are not necessarily bound by the physical principles that govern dynamically downscaled data.

Commonly used techniques in statistical downscaling include the stochastic weather generator (Dabhi et al. 2021; Kim et al. 2020; Vesely et al. 2019), the analog method

(Bettolli 2021; Cortesi 2014; Timbal and McAvaney 2001), weather classification methods (Camus et al. 2016; Lin et al. 2017), and regression-based methods (Gutiérrez et al. 2013). The stochastic weather generator involves generating a data set with the desired statistical properties. The analog method shares similarities with weather classification; with the analog method local conditions are linked with concurrent large scale weather patterns. Weather classification involves a two-part process where the examined period is first split into different weather regimes. For the second step, a downscaling method is selected and used to downscale using the different weather regimes for training. For this study, a combined weather classification and multiple linear regression (MLR) approach was used. MLR has the advantage of being simple and computationally inexpensive, with the downside of potential poor representation of non-linear statistical relationships. More complex techniques like the artificial neural network (ANN) have the potential to better account for non-linearities in these relationships (Hernanz et al. 2021).

A key challenge for regression-based downscaling is that all local variances cannot be explained by lower resolution atmospheric data, resulting in a mismatch of variance for the observed and downscaled data set (Wilby et al. 2004). Variance inflation and randomization are two techniques that offer solutions to this problem (Huth 2002). Variance inflation involves scaling the variance of the downscaled data upward to match the desired level, while randomization involves adding random noise to the data with the desired statistical properties.

Statistical downscaling relies on the stationarity assumption, that is that the statistical relationships between predictors and predictands are unchanging (Lanzante et al. 2018; Pichuka and Maity 2018; Pichuka et al. 2022). The stationarity assumption is important when assessing the reliability of downscaling for future climates, due to potential alterations in statistical relationships as climate changes. The assumption of stationarity of statistical relationships can play a role in downscaling for current climates as well, as any failure of the stationarity assumption can result in degraded downscaling performance.

Wind direction is an important meteorological variable that can be indicative of the relative level of influence exerted by maritime or continental air masses on observed local temperatures for both coastal and inland locations. The objective of this study was to examine the potential for wind direction to influence statistical relationships between predictors and predictands in a downscaling model with a combined regression and weather classification approach. This was examined by evaluating downscaling performance based on the weather classifications of the training and evaluation periods used. Two types of locations were examined for the study: coastal locations and inland locations. The effects of wind direction relative to the axis of the coast were evaluated for the coastal locations, while the effects of cardinal wind direction were examined for the inland locations.

3.2 Methodology

The methodology is described in the following section. Included is information about the data used for the study, as well as how the downscaling model developed for the study functioned. The downscaling model was designed to accomplish the primary objectives of the study, which were to evaluate potential effects of wind direction on statistical relationships between predictors and predictands in the downscaling model and to determine whether a weather classification scheme based on wind direction could be used to improve downscaling performance.

3.2.1 Data

The North American Regional Reanalysis (NARR) data set was used to generate the predictor variables in this study (Mesinger et al. 2006). The NARR uses a hybrid approach that incorporates observational data with modelling to create a historical record of weather conditions across the North American continent. The NARR data are generated at 32 km in resolution, which is roughly equivalent to 0.3^{o} for the lowest latitudes. Table 3.1 describes the meteorological variables obtained from the NARR for this study, including instantaneous temperature, u wind, v wind and specific humidity at the 850 hPa and 500 hPa levels. Values for these variables at the surface were also obtained, along with surface pressure. The period of time examined for the study was from 1981 to 2020, for a total of 40 years.

Table 3.1 Variable Information.

Variable Code	Description	Levels
Ps	Sea Level Pressure	Surface
T	Temperature	850 ,500 hPa
U	U Wind	850, 500 hPa
V	V Wind	850, 500 hPa
H	Specific Humidity	850 ,500 hPa
Ts	Surface Temperature	Surface
Us	Surface U Wind	Surface
Vs	Surface V Wind	Surface
Hs	Surface Specific Humidity	Surface
Tn	Min Daily 3-Hourly Surface Temperature	Surface

The predictands used for the study were minimum daily temperatures. The stations chosen for predictands were divided among coastal and inland locations. The total number of stations examined was 36, with 12 inland locations and 24 coastal locations. The coastal locations were divided between the Pacific Coast, the Gulf Coast, the Atlantic Coast, and the Great Lakes, while the inland locations were in the center of the continent. Figure 3.1 depicts the selected locations.

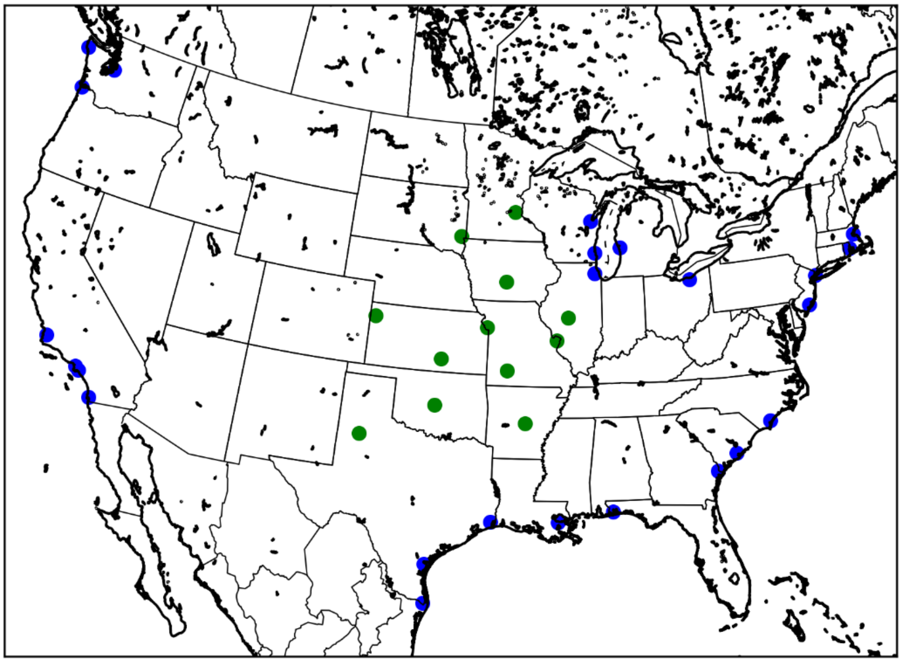

Figure 3.1 Station locations for the 24 coastal locations (blue) and 12 inland locations (green).

3.2.2 Predictor Variable Combinations

The variables used for this study have been commonly used in temperature downscaling studies (Gutiérrez et al. 2013). Several different combinations of predictor variables were tested for the study; explanations for the variable combinations examined are described in Table 3.2. Though several different combinations of predictor variables were tested for the study, the primary focus of the results was on the C8 combination, which proved to have similar performance to the C5 combination and better performance than the CS combination. Also examined in the results was the simplest combination, which included the reanalysis minimum temperature as the sole meteorological variable. While for this study the predictor combinations were chosen with a subjective approach, other techniques such as correlation analysis (Khan et al. 2006) and stepwise regression (Gaitan et al. 2014) have been used for predictor selection.

Table 3.2 Tested Combinations of Predictors.

Combination Name	Predictors
CS	Ps, Ts, Us, Vs, Hs, Tn
C8	CS, T850, U850, V850, H850, Tn
C5	C8, T500, U500, V500, H500, Tn
Tn	Tn
CS_{nw}	Ps, Ts, Hs, Tn
$C8_{nw}$	CS_{nw}, T850, H850, Tn
$C5_{nw}$	$C8_{nw}$, T500, H500, Tn

The stations used for the study were distributed across three different time zones on the North American continent. For each time zone, an estimate of the most frequent time for minimum daily temperature was used as the time step for instantaneous predictor variables. It is important to note that these predictor variables were not necessarily temperature related, but the time at which they were measured was. For the Eastern Time Zone (ETZ) the 9 UTC timestep was used, and for the Pacific Time Zone (PTZ) and the Central Time Zone (CTZ) the 12 UTC timestep was used. For locations near the edge of the ETZ and the CTZ, the minimum temperature occurred in similar amounts at 9 and 12 UTC, and for simplicity's sake stations close to that edge were simply grouped with their zone.

The actual estimated daily minimum temperature predictor values were based on the minimum values of surface temperature seen within the calendar day defined by the time zone the relevant station as located in. Temporal interpolation was used when the boundaries of the calendar day did not match with the 3-hour temporal resolution of the NARR. These temperatures are referred to in this study as the reanalysis minimum temperatures.

3.2.3 Evaluation Methods

To evaluate model performance, the root mean square error (RMSE) and bias metrics were used. Lower RMSE values indicate better performance. RMSE performance is more affected by outliers than the mean absolute error (MAE). Bias determines whether the model is either overestimating or underestimating temperatures in its predictions.

3.2.4 PCA and Standardization

Principal Component Analysis (PCA) was performed on the predictor variables as a means to reduce the dimensions of the data (Preisendorfer and Mobley 1988). The dominance method of PC selection was used to determine which PCs were kept, and a wide variety of variance thresholds were tested to determine an optimal cutoff point. If the cutoff point resulted in a predictor variable being completely eliminated from the model the results were discarded. PCA was evaluated on the data from the training period, and data from the evaluation period was projected onto those axes generated from the training period. This was done to maintain the independence of the PC results in the evaluation period from the training period.

All variables were standardized by Julian Day (JD) to mitigate potential effects of seasonality (Gutiérrez et al. 2013). To reduce noise in the data a smoothing process was used. The smoothing process involved calculating means and standard deviations for a collection of data temporally adjacent to each JD in question.

3.2.5 Downscaling Methodology

The primary downscaling method used for this study was a combination of the weather classification and multiple linear regression (MLR) methods. The main advantage provided by the MLR method is its simplicity and computational inexpensiveness, making it particularly useful for comparison purposes to other downscaling methods. MLR does come with drawbacks in that it cannot account for non-linear relationships between predictors and predictands and is unable on its own to provide variance that is lost from the mismatch in scale.

A k-fold cross-validation technique similar to that described in (Gutiérrez et al. 2013) was used to evaluate performance in the study. The time period in question (1981-2020) was split into five different periods of eight years. Each set of eight years was then used as an evaluation period, with the other four periods serving as training periods. Evaluating model performance in this manner served to address problems related to model overfitting.

3.2.6 Weather Classification

The weather classifications used for partitioning the training period in the study were solely based off the NARR-estimated wind direction generated from the mean u and v components over the 6 hours preceding the NARR-estimated daily minimum temperature at each of the station locations. Temporal interpolation was necessary when the estimated minimum temperature did not occur at the 3-hour standard NARR timestep. Each day was assigned to the closest of four directions. For the coastal locations the four directions were based on the wind direction relative to the axis of the coast (perpendicular over water, perpendicular over land, or parallel), while for the inland stations the wind directions were based on the four cardinal directions. The downscaling model was run using training periods based on all four directions. For comparison purposes, additional training periods containing either all available days or random selections of days with size equivalent to each of the directional training periods were also tested. For the random selections, six different sets of randomly selected days for every needed directional equivalent size were all evaluated, and the results were then pooled. Table 3.3 contains an overview of the training and evaluation periods used in the study.

Table 3.3 Training and evaluation period descriptions. Wind direction, where applicable, designates the direction source of the wind.

Training Period	**Wind Direction**
T_W	Water
T_L	Land
T_{S1}	Parallel to coast
T_{S2}	Parallel to coast, reverse of TS1
T_N	North
T_S	South
T_{We}	West
T_E	East
T_R	Random selection of days with equivalent length to the indicated direction (for example T_{RE} would be a random set of days with equal size of the T_E training period
T_A	All days
Evaluation Period	**Wind Direction**
V_W	Water
V_L	Land
V_{S1}	Parallel to coast
V_{S2}	Parallel to coast, reverse of TS1
V_N	North
V_S	South
V_{We}	West
V_E	East

Breakdowns of model performance were generated based on wind speed, and wind direction. The methodology described in this portion of the study allowed for comparisons of downscaling performance at inland and coastal locations under different wind direction training conditions as well as different evaluation conditions. In addition, potential effects of the domains used for predictor generation under the same training and evaluation conditions could also be examined.

3.2.7 Domains

Five different domains were tested with every configuration. Each of these domains defined which NARR cells were used for generation of the predictors. The domains were based off distance from a predetermined point that marked the center of the domain. The domains at the coastal stations and inland stations were defined in a similar way to how their wind directions were defined. Domains for the coastal stations were dependent on their positions relative to the axis of the coast, while for the inland stations the domain positions were based on the four cardinal directions. For each of the coastal stations, a water domain (D_W) was created with a center point roughly 140 kilometers from the station location, in a direction perpendicular to the coast. A land domain (D_L) was then created with a center point at the same distance 180^o in direction from the water domain's center point. Two additional domains were then defined at the 90^o directions with center points roughly parallel to the coast (D_{S1} and D_{S2}). The final domain was centered on the station itself (D_C). For the inland stations, the domain centers were set at locations roughly 140 kilometers from each station along the four cardinal directions (D_N, D_S, D_W, and D_E). A domain centered on the station was also used with each of the inland stations (D_C).

NARR cells within roughly 195 kilometers of the domain center, using the great circle distance, were included with the domain. Figure 2.2 illustrates an example of the relative positions of the domains to the stations as well as their NARR cell compositions.

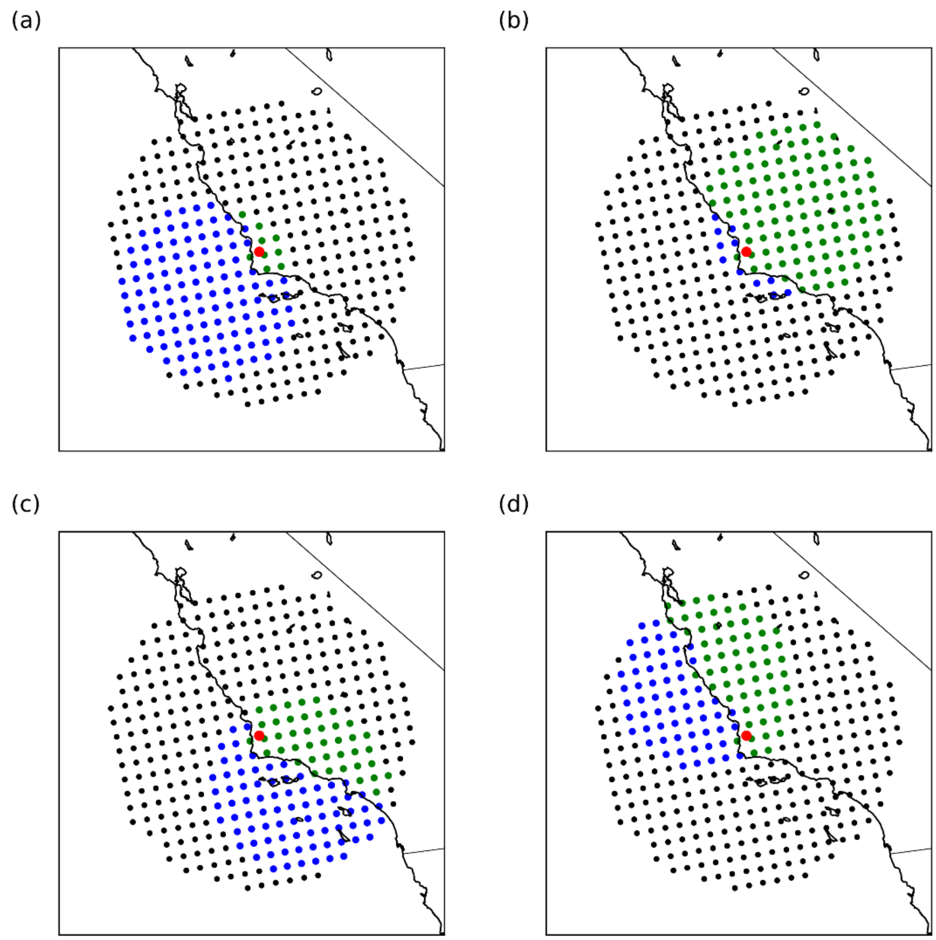

Figure 3.2 NARR cells covering the D_W (a), D_L (b), D_{S1} (c), and D_{S2} (d) domains for the Santa Maria Public Airport station in California. The blue dots are water cells in the NARR land/sea mask while the green dots are land cells. The black dots are all NARR cell locations within roughly 335 km of the station location.

3.3 Results and Discussion

The focus for the results of the study was on the potential role wind direction played in the statistical relationships between predictors and predictands, and whether weather classification based on wind direction could be useful for mitigating any problems arising from changes in the statistical relationships.

The following section contains downscaling performance results as well as discussion of these results. Included are box plots detailing RMSE and bias performance at the 24 coastal locations and 12 inland locations under various training period and evaluation period conditions. Unless otherwise indicated in this section, the domain used for predictor generation in the downscaling model was centered on the station location, and the C8 combination defined the selection of variables used for predictor generation.

3.3.1 Coastal Locations

Results for downscaling performance of daily minimum temperature at the coastal locations are discussed in this section. Unless otherwise indicated in this section, all 24 coastal locations were included in the results. The different training periods used for the results in this section consisted of the T_A, T_R, T_W, T_L, T_{S1}, and T_{S2} periods. The T_A training period included every day that could potentially be used for training the downscaling model, while the T_R training periods consisted of a random selection of days with size equivalent to one of the weather classification training periods. The four weather classification training periods (T_W, T_L, T_{S1}, and T_{S2}) were made up of the days where the mean wind direction came from the specified direction. The standard evaluation periods used for the results covered the same days that the four weather classification training periods covered (V_W, V_L, V_{S1}, and V_{S2}). Additional evaluation periods included further limiting the days used for evaluation based on mean wind speed.

Standard Evaluation Conditions

RMSE results for the standard training and evaluation periods are shown in Figure 3.3. Mean RMSE performance was easily the best when the training period was matched to the evaluation period (T_W, T_L, T_{S1}, T_{S2} in Fig. 3.3a, 3.3b, 3.3c, 3.3d respectively). Mean RMSE values were highest when the wind directions of the training and evaluation periods were opposite in direction (T_L, T_W, T_{S2}, T_{S1} in Fig. 3.3a, 3.3b, 3.3c, 3.3d respectively); RMSE performance was exceptionally poor when the training was done with T_W and the evaluation with V_L. Matched training/evaluation periods outperformed the T_R training periods for every wind direction with respect to mean RMSE and were close in mean RMSE values to the T_A training periods.

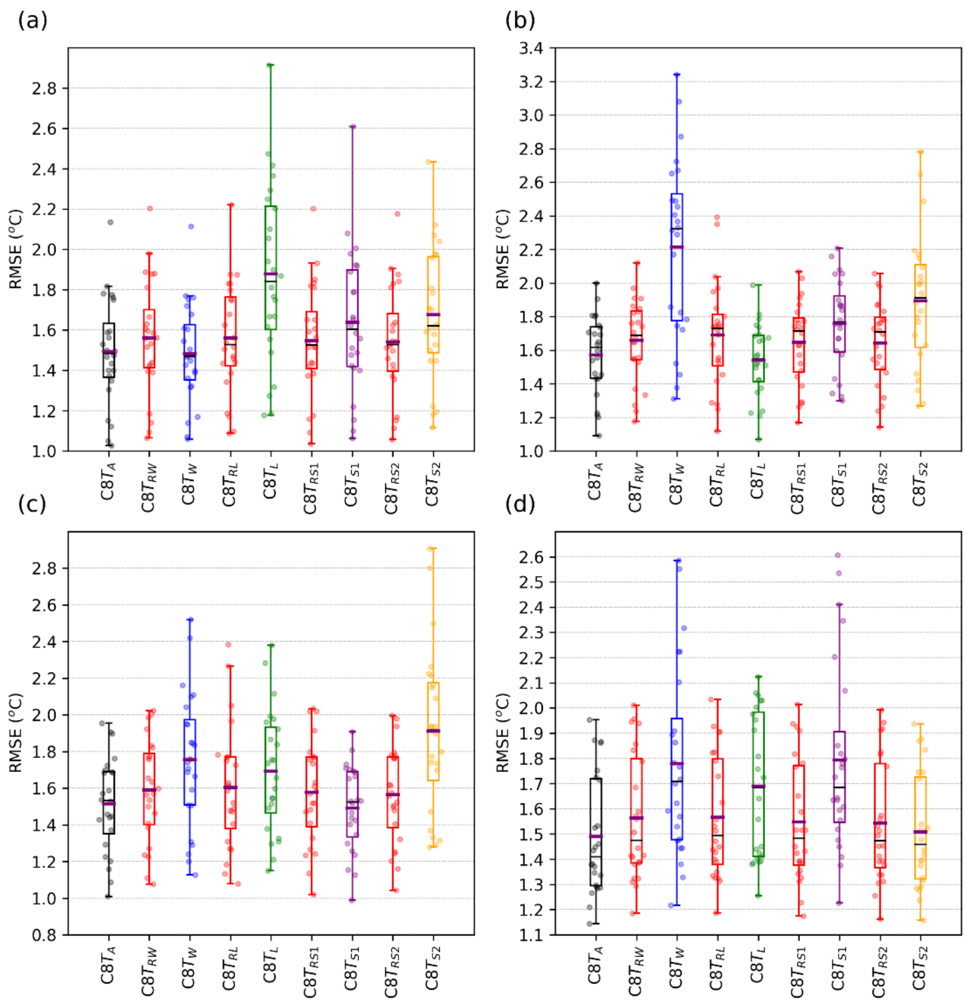

Figure 3.3 RMSE results using the C8 predictor combination for the 24 coastal locations with various training periods. Results are shown for the V_W evaluation period (a), V_L evaluation period (b), V_{S1} evaluation period (c), and V_{S2} evaluation period (d). The training periods are described on the x-axes. The T_A and T_R training periods are shown in black and red, respectively. The T_W, T_L, T_{S1}, and T_{S2} are shown in blue, green, purple, and orange, respectively. Median values are given by the thin black lines and mean values by the thick purple lines.

Bias results for the standard evaluation and training periods are illustrated in Figure 3.4. Many patterns seen with bias in Figure 3.4 were like those seen with RMSE in Figure 3.3. Unmatched training/evaluation periods exhibited negative mean bias values that were higher in absolute magnitude than the matched periods. When the wind direction of the training period was opposite in direction from that of the evaluation period, the absolute magnitude of the mean bias was highest. Mean bias values for all matched training/evaluation periods were very close to 0 ^{o}C, which was not the case with the T_R and T_A training periods. The variance of bias values seen with the matched training/evaluation periods across the 24 stations was also smaller than that seen with the T_R and T_A training periods.

These results for RMSE and bias indicated that the statistical relationships between predictors and predictands were strongly dependent on the wind direction of the training period.

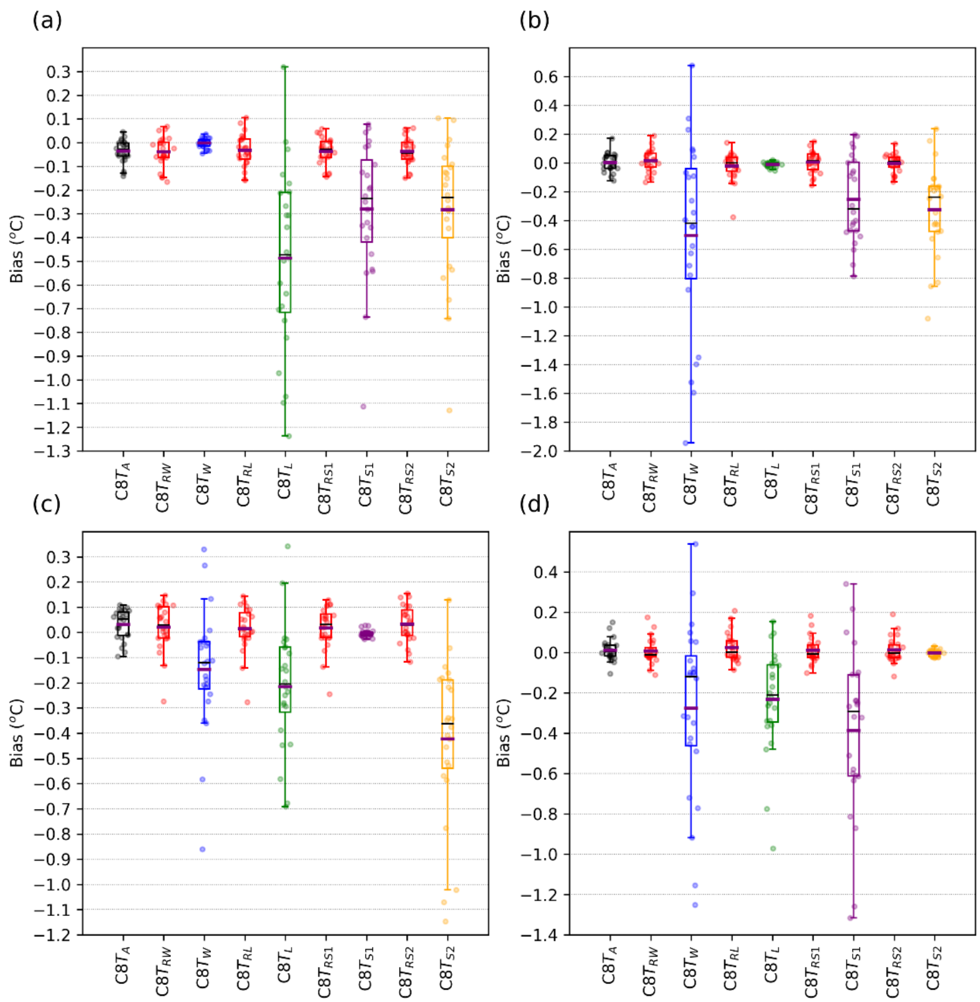

Figure 3.4 Bias results using the C8 predictor combination for the 24 coastal locations with various training periods. Results are shown for the V_W evaluation period (a), V_L evaluation period (b), V_{S1} evaluation period (c), and V_{S2} evaluation period (d). The training periods are described on the x-axes. The T_A and T_R training periods are shown in black and red, respectively. The T_W, T_L, T_{S1}, and T_{S2} training periods are shown in blue, green, purple, and orange, respectively. Median values are given by the thin black lines and mean values by the thick purple lines.

Wind Speed Threshold

The evaluation periods for the results described in this section have been limited to days where the mean wind speed was greater than 4 m/s. RMSE results with the wind speed threshold are described in Figure 3.5. The mean RMSE values of the matched training/evaluation periods were lower relative to the mean RMSE values for the T_R and T_A training periods when the wind speed threshold was included, while the opposite was the case with the unmatched training/evaluation periods (c.f. Fig. 3.5 with Fig. 3.3).

Bias results with the 4 m/s wind speed threshold are shown in Figure 3.6. For the unmatched training/evaluation periods, bias values were of larger absolute magnitude with the wind speed threshold (c.f. Fig. 3.6 with Fig. 3.4). For the V_W and V_L evaluation periods, the matched training/evaluation periods produced bias values of a smaller absolute magnitude than seen with the T_R and T_A training periods (Fig. 3.6a, 3.6b).

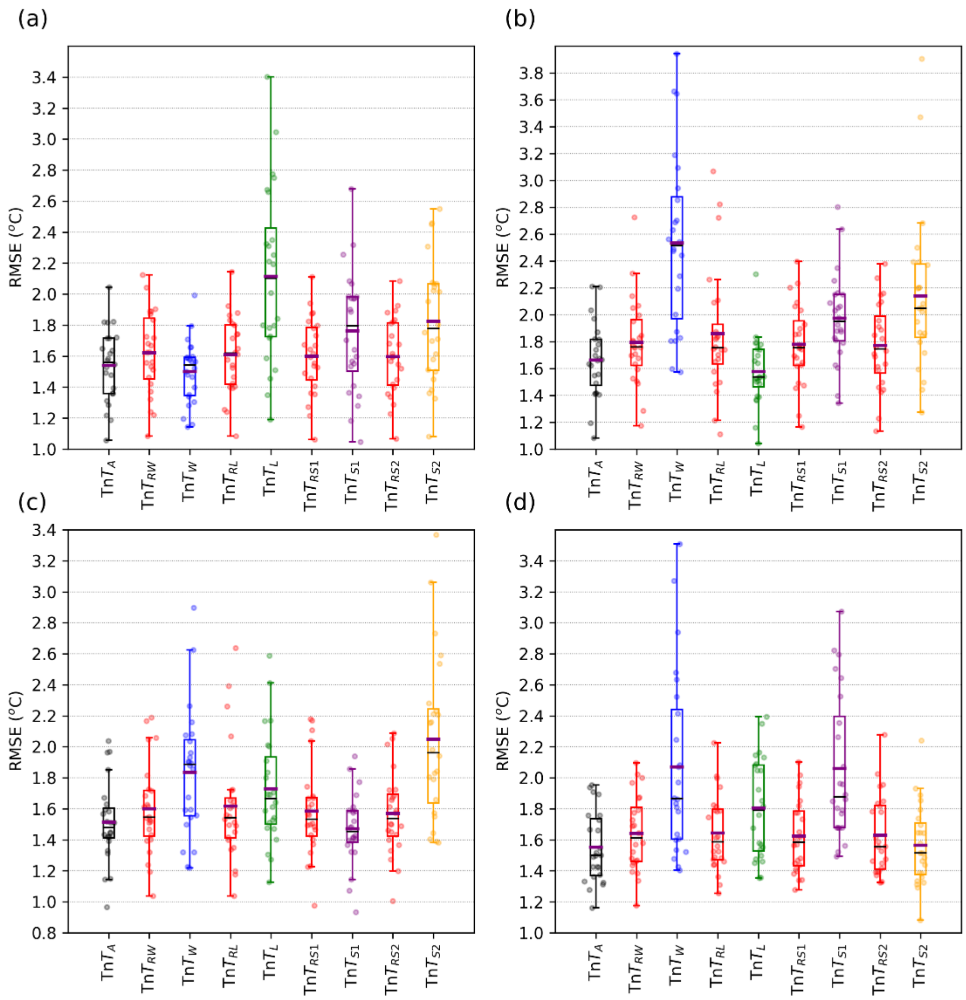

Figure 3.5 RMSE results using the C8 predictor combination for the 24 coastal locations with various training periods. Results for this figure are limited to days where the mean wind speed for the relevant time was greater than 4 m/s. The wind speed threshold limited the number of days available for evaluation. Results are shown for the V_W evaluation period (a), V_L evaluation period (b), V_{S1} evaluation period (c), and V_{S2} evaluation period (d). The training periods are described on the x-axes. The T_A and T_R training periods are shown in black and red, respectively, while the T_W, T_L, T_{S1}, and T_{S2} training periods are shown in blue, green, purple, and orange, respectively. Median values are given by the thin black lines and mean values by the thick purple lines.

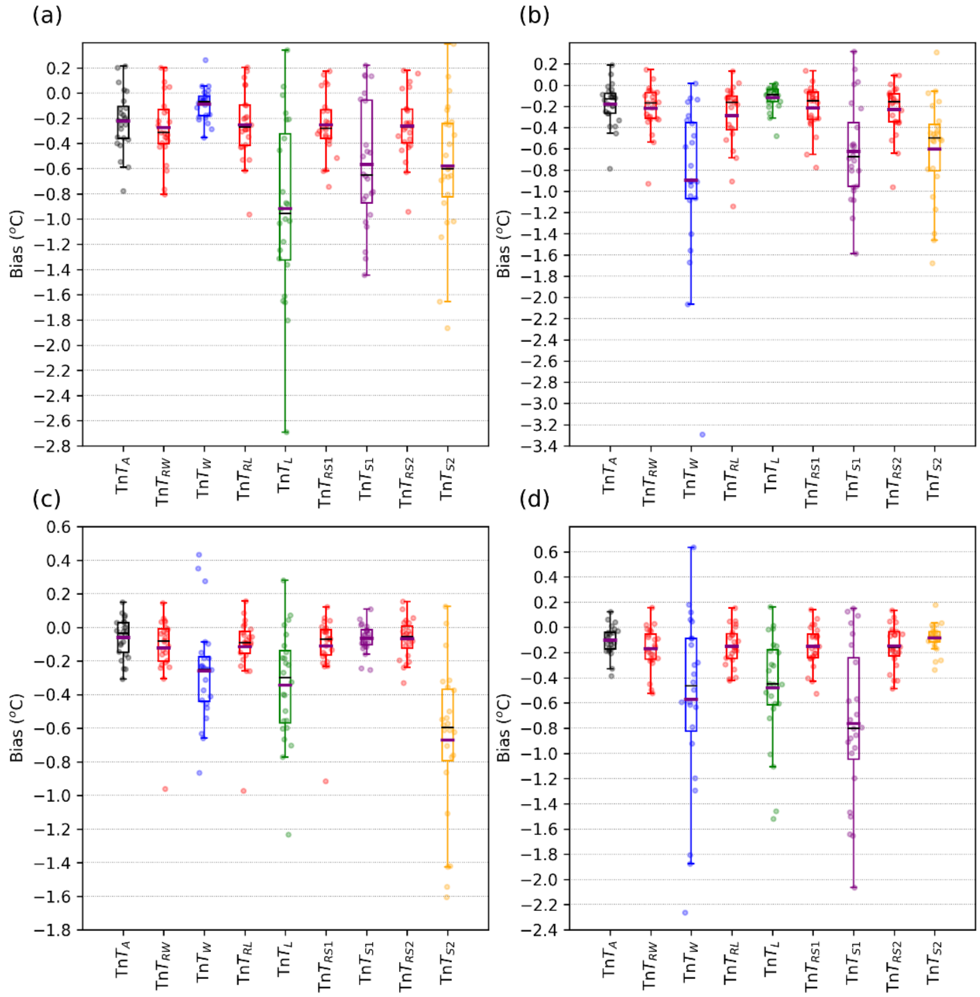

Figure 3.6 Bias results using the C8 predictor combination for the 24 coastal locations with various training periods. Results for this figure are limited to days where the mean wind speed for the relevant time was greater than 4 m/s. The wind speed threshold limited the number of days available for evaluation. Results are shown for the V_W evaluation period (a), V_L evaluation period (b), V_{S1} evaluation period (c), and V_{S2} evaluation period (d). The training periods are described on the x-axes. The T_A and T_R training periods are shown in black and red, respectively, while the T_W, T_L, T_{S1}, and T_{S2} training periods are shown in blue, green, purple, and orange, respectively. Median values are given by the thin black lines and mean values by the thick purple lines.

Tn Predictor Combination

RMSE results with the Tn predictor combination (and no wind speed threshold) are shown in Figure 3.7. The sole variable used for predictor generation was the reanalysis minimum temperature with the Tn combination. Mean RMSE values were less dependent on the wind direction of the training period when the Tn combination was used, with the difference in mean RMSE values between matched training/evaluation wind directions and opposite training/evaluation wind directions being less than that seen with the C8 combination (c.f. Fig. 3.7 with Fig. 3.3). Mean RMSE performance with the Tn combination was generally worse than that seen with the C8 combination, though there were some exceptions (c.f. Fig. 3.7 with Fig. 3.3). The gap was smallest and even negative in some instances when the wind direction of the training period was opposite that of the evaluation period. The most extreme example of this occurrence was when the training was done with T_W and the evaluation with V_L (c.f. Fig. 3.7b with Fig. 3.3b). This meant that information included in the C8 combination that was not the reanalysis minimum temperature was reducing the model's RMSE performance in these situations. Further investigation revealed that removing the wind variables from the C8 combination had the effect of lessening the loss in performance seen when the direction of the training wind was opposite that of the evaluation wind (not shown).

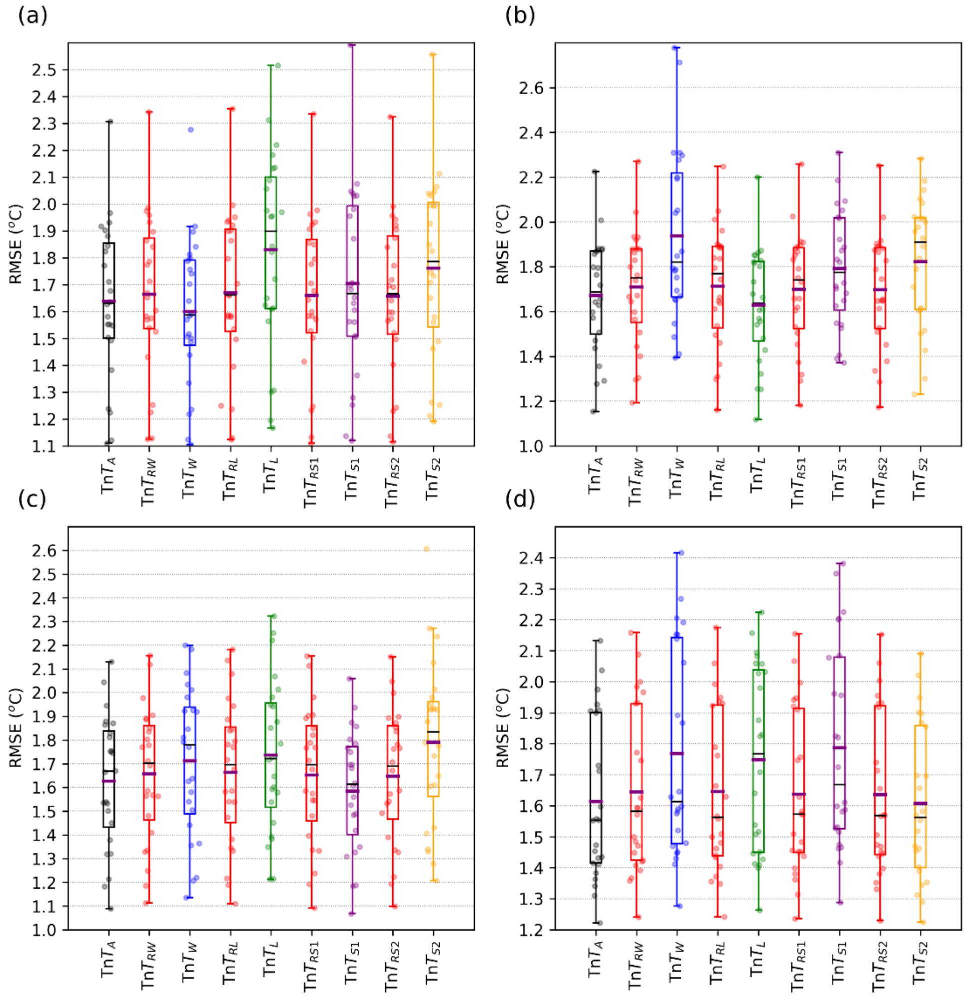

Figure 3.7 RMSE results using the Tn predictor combination for the 24 coastal locations with various training periods. Results are shown for the V_W evaluation period (a), V_L evaluation period (b), V_{S1} evaluation period (c), and V_{S2} evaluation period (d). The training periods are described on the x-axes. The T_A and T_R training periods are shown in black and red, respectively. The T_W, T_L, T_{S1}, and T_{S2} training periods are shown in blue, green, purple, and orange, respectively. Median values are given by the thin black lines and mean values by the thick purple lines.

Differences Between Coasts

RMSE results with the C8 combination for only those stations on the Pacific Coast are shown in Figure 3.8. RMSE values for the stations on the Pacific Coast tended to be less influenced by differences in the direction of wind for the training period and evaluation period compared to the other coastal locations (c.f. Fig. 3.8 with Fig. 3.3). One potential explanation for this smaller influence of wind direction was the lesser exposure to continental air masses on the Pacific Coast. If the air masses affecting the region were more uniform regardless of wind direction, the potential effects of wind direction could be reduced.

Domain Selection

The effects of changing the domain used for predictor generation are illustrated in Figure 3.9. Only matched training/evaluation wind directions were used for the weather classification training periods in Figure 3.9. Changing the domain used for predictor generation did not have much effect on RMSE performance, though slightly worse performance was seen with the water domain in all circumstances (W_D). Lower quality of data over water and lesser variability in meteorological data over water could potentially be explanations for the water domain's worse performance. For the inland locations, the results for changing the domain of predictors were small and are not detailed in a separate section.

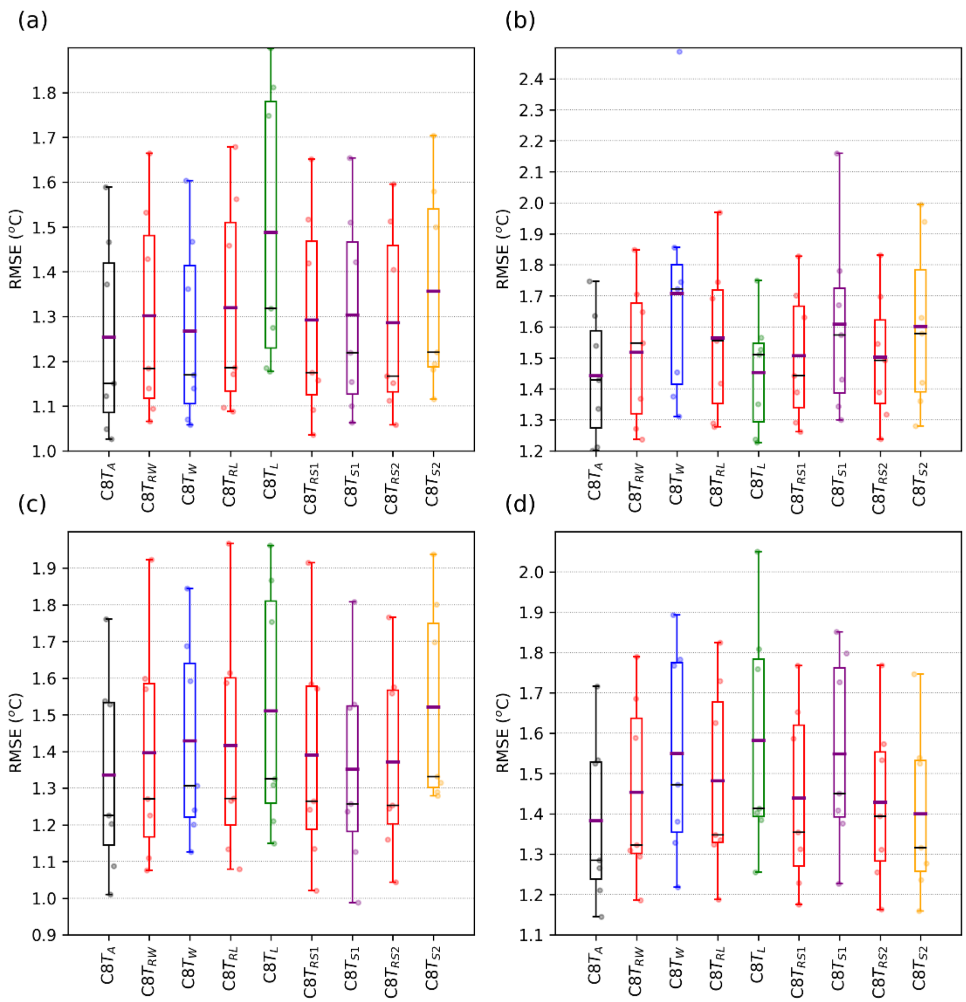

Figure 3.8 RMSE results using the C8 predictor combination for only the 7 locations on the Pacific Coast with various training periods. Results are shown for the V_W evaluation period (a), V_L evaluation period (b), V_{S1} evaluation period (c), and V_{S2} evaluation period (d). The training periods are described on the x-axes. The T_A and T_R training periods are shown in black and red, respectively. The T_W, T_L, T_{S1}, and T_{S2} are shown in blue, green, purple, and orange, respectively. Median values are given by the thin black lines and mean values by the thick purple lines.

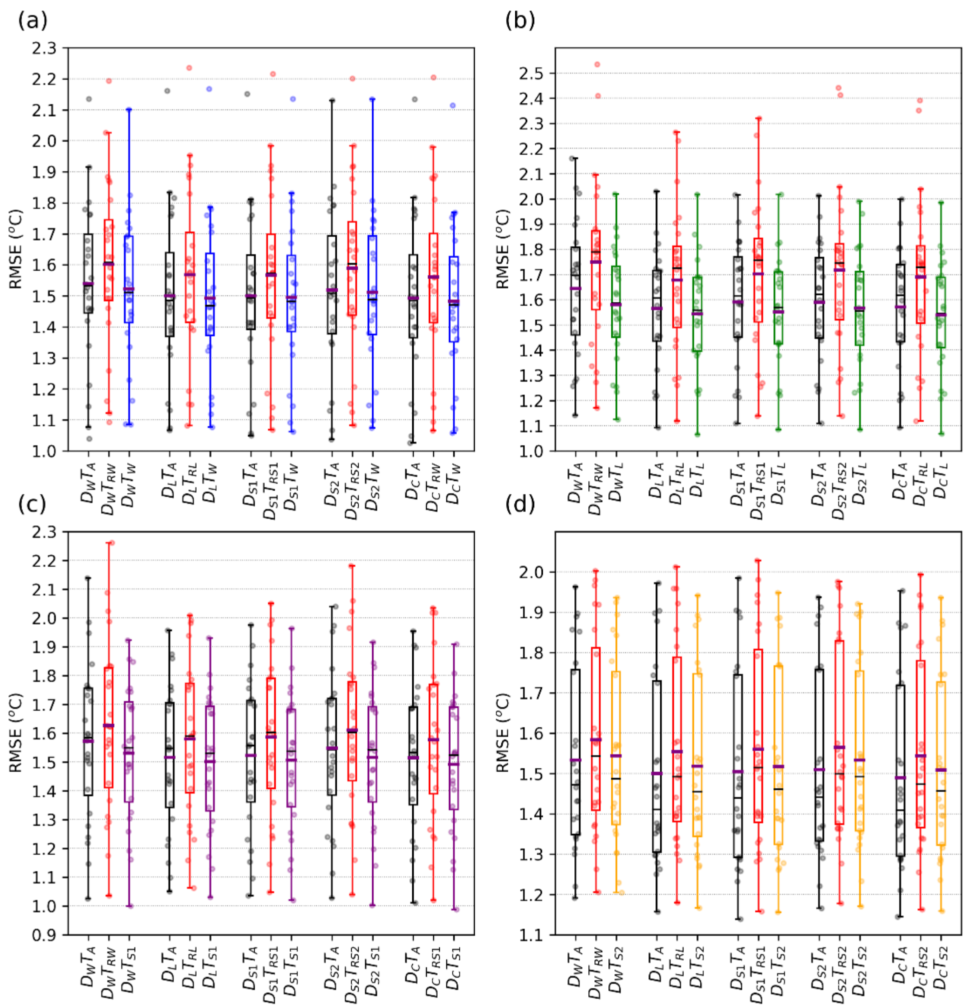

Figure 3.9 RMSE results using the C8 predictor at all 24 coastal locations using various predictor domain and training period combinations. Results are shown for the V_W evaluation period (a), V_L evaluation period (b), V_{S1} evaluation period (c), and V_{S2} evaluation period (d). The domain and training period combinations are described on the x-axes, with the domain coming first. The T_A and T_R training periods are shown in black and red, respectively. The T_W, T_L, T_{S1}, and T_{S2} are shown in blue, green, purple, and orange, respectively. Median values are given by the thin black lines and mean values by the thick purple lines.

3.3.2 Inland Locations

Results for downscaling performance of daily minimum temperature at the inland locations are discussed in this section. All 12 inland locations were included in the results. The different training periods used for the results in this section consisted of the T_A, T_R, T_N, T_S, T_{We}, and T_E periods. The T_A training periods consisted of all days available for training in the downscaling model, while the T_R training periods were made up of a random selection of days with size equivalent to one of the weather classification training periods. The four weather classification training periods (T_N, T_S, T_{We}, and T_E) consisted of the days where the mean wind direction came from the specified direction. The standard evaluation periods used for the results covered the same days that the four weather classification training periods covered (V_N, V_S, V_{We}, and V_E). As was the case with the coastal locations, additional evaluation periods were examined where the days were further limited based on wind speed.

Standard Evaluation Conditions

RMSE results for the standard training and evaluation periods are shown in Figure 3.10. Some of the same patterns seen with the coastal locations were present with the inland locations. Opposite training/evaluation periods (T_S, T_N, T_E, T_{We} in Fig. 3.10a, 3.10b, 3.10c, 3.10d respectively) again performed much worse with respect to mean RMSE than their matched counterparts' periods (T_N, T_S, T_{We}, T_E in Fig. 3.10a, 3.10b, 3.10c, 3.10d respectively). Matched training/evaluation periods performed better with respect to mean RMSE than the T_R training periods and were close in mean RMSE values to the T_A

training periods, as was the case with the coastal locations. Intriguingly, with respect to mean RMSE values, the T_N training period performed better with the V_{We} evaluation period than the T_E and T_S training periods did (c.f. T_N with T_E, T_S in Fig. 3.10c). The reverse was also the case (c.f. T_{We} with T_E, T_S in Fig. 3.10a). A similar but weaker pattern was seen with the T_E and T_S training periods. Given that the western and northern wind directions for the inland locations are more associated with continental air masses, while the southern and eastern wind directions are less so, these results are consistent with the hypothesis that the type of air mass associated with the wind direction was a factor in the predictor/predictand statistical relationships established by the downscaling model.

Bias results for the standard training and evaluation periods are depicted in Figure 3.11. As was the case with coastal locations, the matched training/evaluation periods exhibited very small absolute mean bias values as well as the smallest variance in bias values.

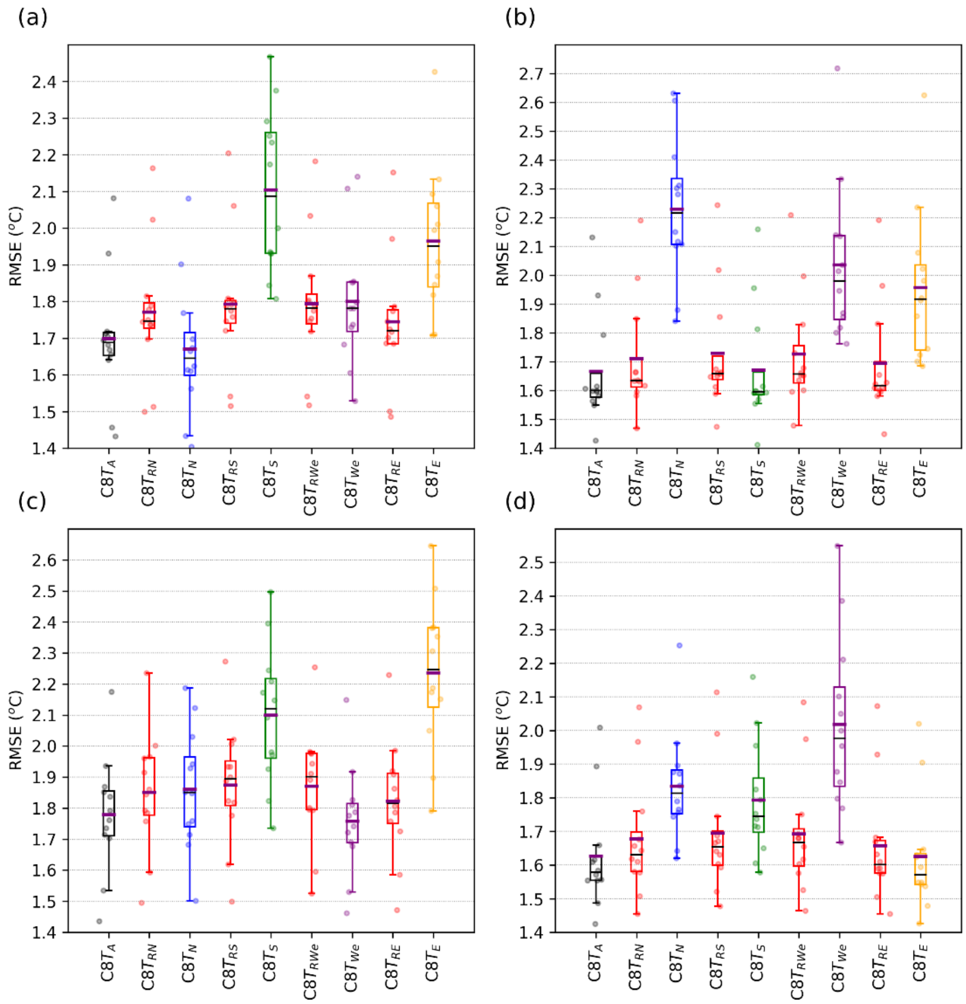

Figure 3.10 RMSE results using the C8 predictor combination for the 12 inland location with various training periods. Results are shown for the V_N evaluation period (a), V_S evaluation period (b), V_{We} evaluation period (c), and V_E evaluation period (d). The training periods are described on the x-axes. The T_A and T_R training periods are shown in black and red, respectively. The T_N, T_S, T_{We}, and T_E are shown in blue, green, purple, and orange, respectively. Median values are given by the thin black lines and mean values by the thick

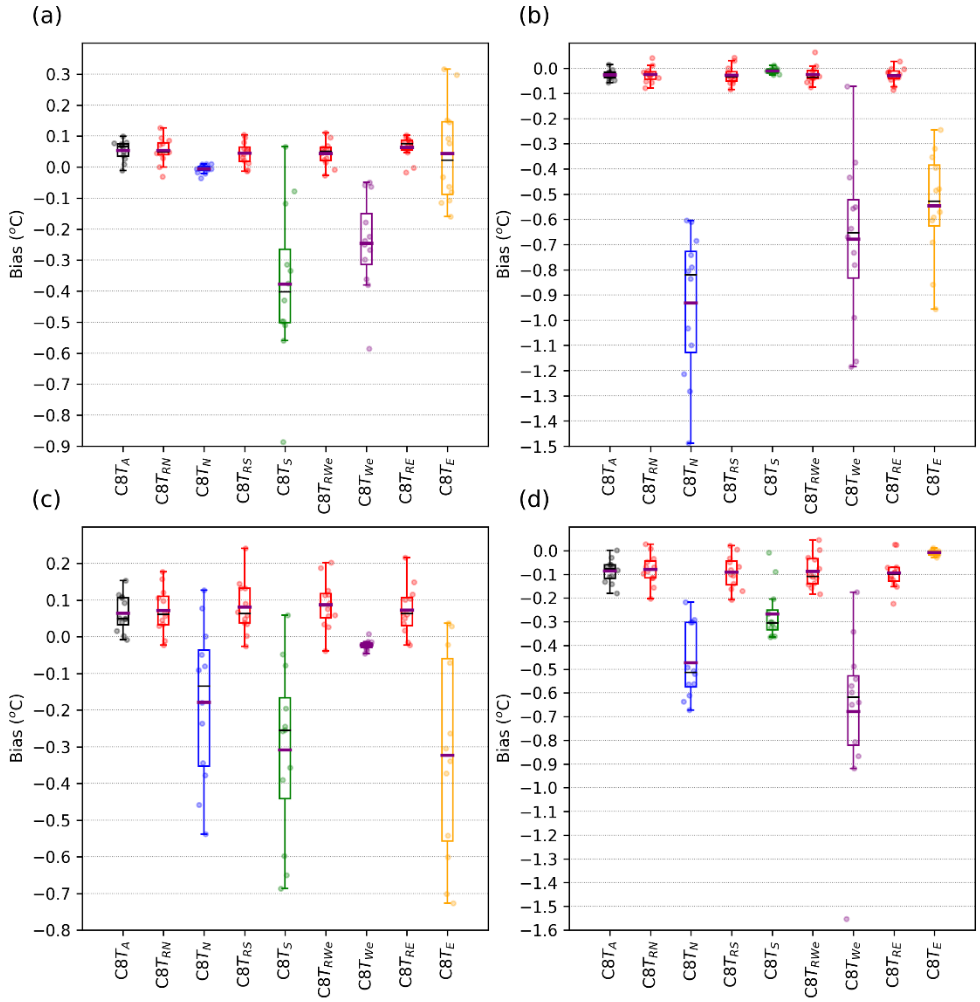

Figure 3.11 Bias results using the C8 predictor combination for the 12 inland location with various training periods. Results are shown for the V_N evaluation period (a), V_S evaluation period (b), V_{We} evaluation period (c), and V_E evaluation period (d). The training periods are described on the x-axes. The T_A and T_R training periods are shown in black and red, respectively. The T_N, T_S, T_{We}, and T_E are shown in blue, green, purple, and orange, respectively. Median values are given by the thin black lines and mean values by the thick purple lines.

Wind Speed Threshold

RMSE and bias results when the 4 m/s wind speed threshold was used are shown in Figures 3.12 and 3.13. Some of the patterns seen with RMSE without the wind speed threshold were amplified with the wind speed threshold, particularly the relative performances of the opposite and matched training/evaluation periods (c.f. Fig. 3.12 with Fig. 3.10). Bias results with the 4 m/s wind speed threshold were less clear, though the matched training/evaluation period did perform superior to the other tested training periods for the V_S and V_E evaluation periods.

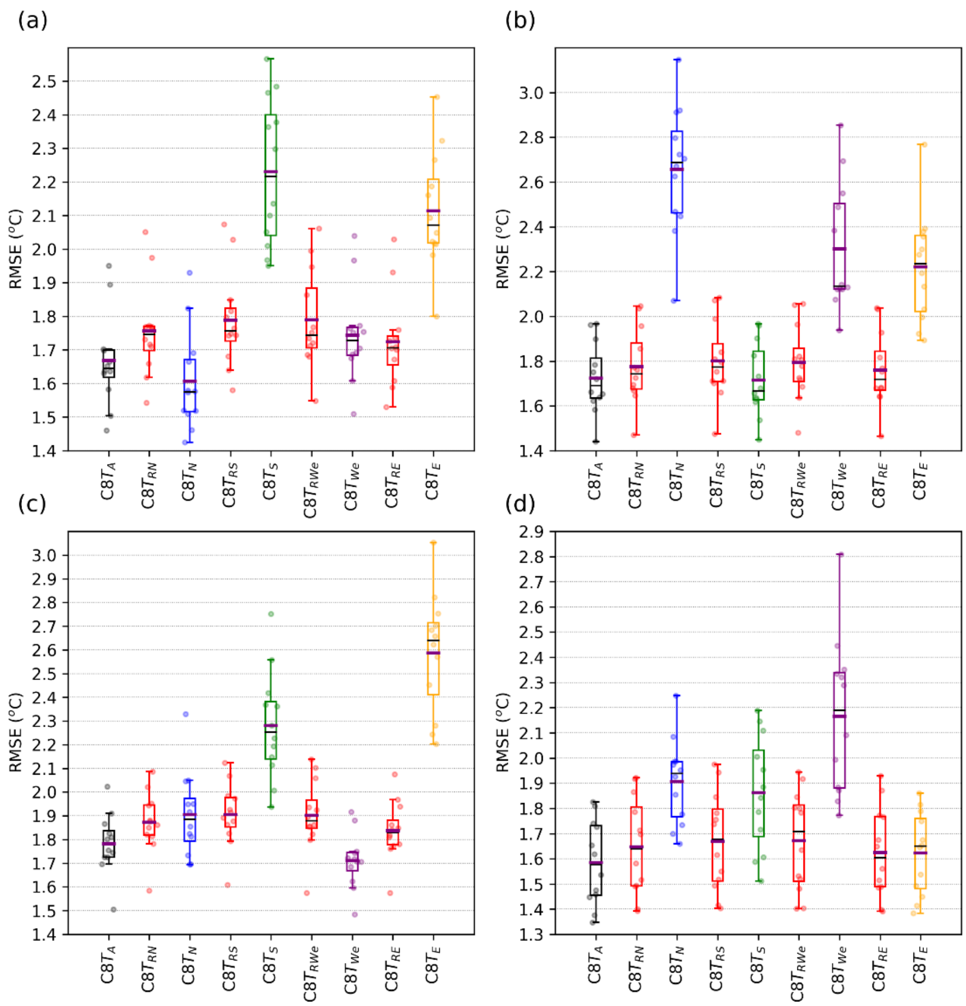

Figure 3.12 RMSE results using the C8 predictor combination for the 12 inland location with various training periods. Results for this figure are limited to days where the mean wind speed for the relevant time was greater than 4 m/s. The wind speed threshold limited the number of days available for evaluation. Results are shown for the V_N evaluation period (a), V_S evaluation period (b), V_{We} evaluation period (c), and V_E evaluation period (d). The training periods are described on the x-axes. The T_A and T_R training periods are shown in black and red, respectively. The T_N, T_S, T_{We}, and T_E are shown in blue, green, purple, and orange, respectively. Median values are given by the thin black lines and mean values by the thick purple lines.

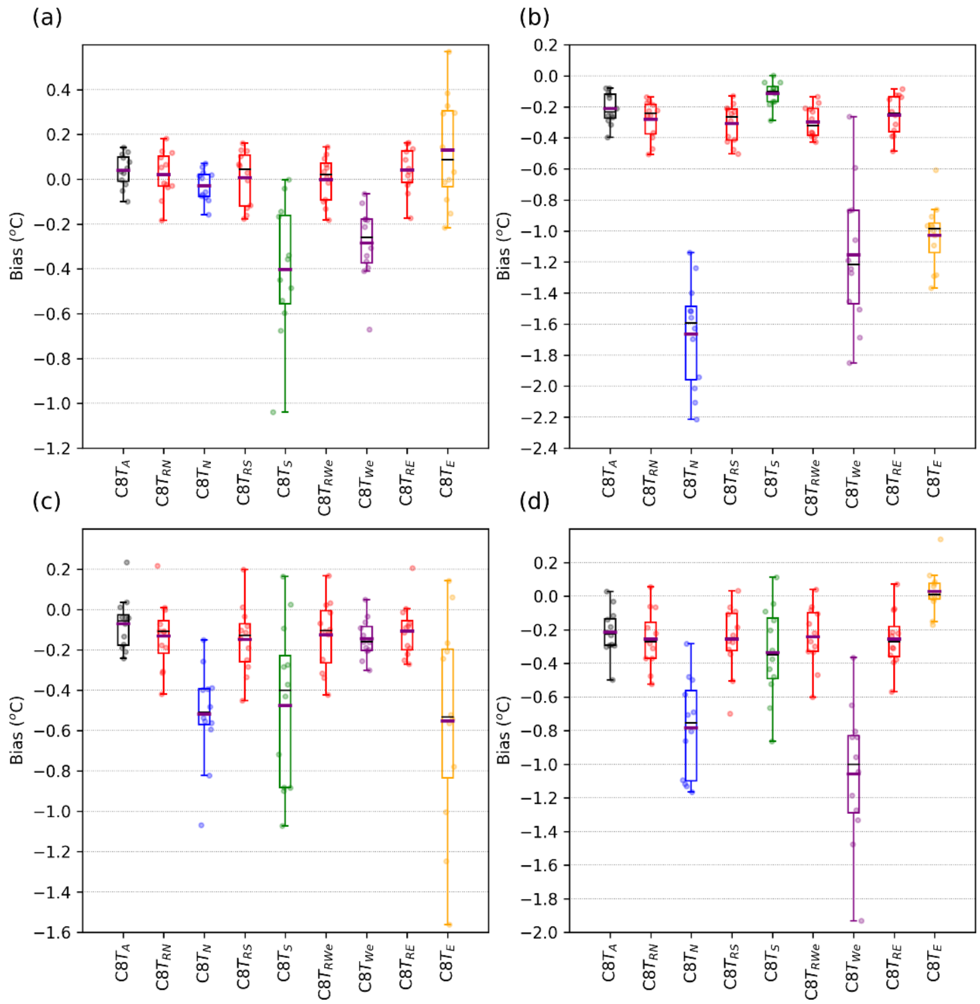

Figure 3.13 Bias results using the C8 predictor combination for the 12 inland location with various training periods. Results for this figure are limited to days where the mean wind speed for the relevant time was greater than 4 m/s. The wind speed threshold limited the number of days available for evaluation. Results are shown for the V_N evaluation period (a), V_S evaluation period (b), V_{We} evaluation period (c), and V_E evaluation period (d). The training periods are described on the x-axes. The T_A and T_R training periods are shown in black and red, respectively. The T_N, T_S, T_{We}, and T_E are shown in blue, green, purple, and orange, respectively. Median values are given by the thin black lines and mean values by the thick purple lines.

Tn Predictor Combination

RMSE results with the Tn predictor combination are illustrated in Figure 3.14. As was the case with coastal locations, mean RMSE values for opposite training/evaluation periods were closer to the matched training/evaluation periods than was seen with the C8 combination (c.f. Fig. 3.14 with Fig. 3.10). However, the largest difference in mean RMSE performance for opposite versus matched training/validation periods (T_N and T_S in Fig. 3.14b) was larger than any seen with the coastal locations (c.f. Fig 3.14 with Fig. 3.7). Overall RMSE performance again tended to be worse with the Tn combination than with the C8 combination, with some exceptions where the weather classification training and evaluation periods were not matched (c.f. Fig. 3.14 with Fig. 3.10). Wind variables in the C8 combination were again found to be reducing performance for the inland locations when the training and evaluation periods were of opposite wind directions (not shown).

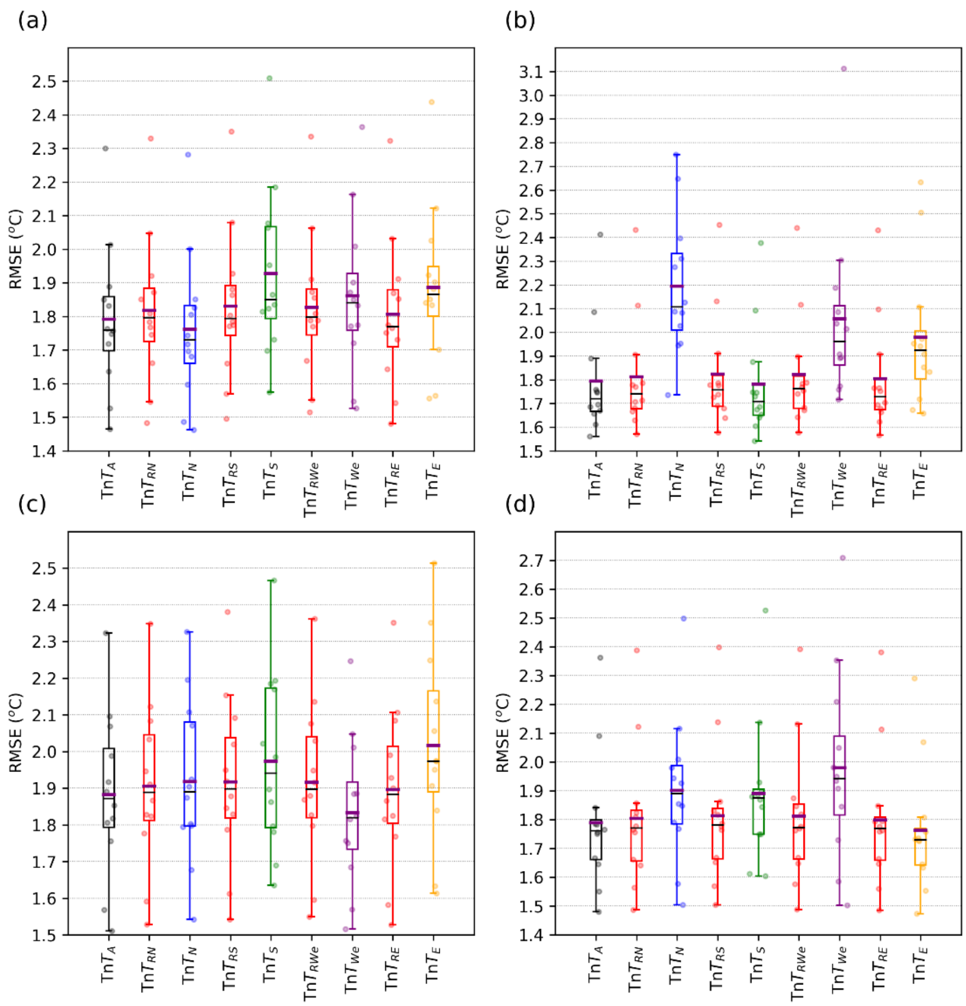

Figure 3.14 RMSE results using the Tn predictor combination for the 12 inland location with various training periods. Results are shown for the V_N evaluation period (a), V_S evaluation period (b), V_{We} evaluation period (c), and V_E evaluation period (d). The training periods are described on the x-axes. The T_A and T_R training periods are shown in black and red, respectively. The T_N, T_S, T_{We}, and T_E are shown in blue, green, purple, and orange, respectively. Median values are given by the thin black lines and mean values by the thick purple lines.

3.4 Conclusions

The primary objectives of the study were to evaluate potential effects of wind direction on statistical relationships between predictors and predictands in the downscaling of minimum daily temperature, and to determine whether a weather classification scheme could mitigate any problems related to these effects. To accomplish these objectives a downscaling model was developed that could evaluate the effects of wind direction on downscaling performance at 24 coastal locations and 12 inland locations in the United States.

3.4.1 Effects of Wind Direction

The study found that wind direction of the training period had the potential to play a major role in downscaling performance. If the wind direction of the training period was not matched to the evaluation period, downscaling RMSE performance was greatly reduced, with the greatest reductions seen when the wind direction of the training period was opposite that of the evaluation period. Further testing indicated that some of this reduction was linked to wind variables in the downscaling model. These results were unsurprising given the extrapolation required when evaluation and training periods were of opposite wind directions.

3.4.2 Weather Classification Scheme Performance

When the wind direction of the training period was matched to the evaluation period using the weather classification scheme, RMSE performance surpassed that seen with a training period of equivalent length but random selection of days (T_R period), while largely matching and under some conditions exceeding the mean RMSE performance seen with the full training period (T_A period). Given that the matched training/evaluation periods were shorter in length than the T_A period and thus more limited on variables, a longer overall training period could improve the relative performance of the matched periods compared to the T_A period. Bias results for the matched training/evaluation periods showed less variance than that seen with the T_R and T_A training periods and were similar or lower in mean absolute magnitude. These RMSE and bias results indicate that a weather classification scheme based on wind direction does have the potential to improve MLR-based downscaling performance in some situations.

3.4.3 Coastal Locations Versus Inland Locations

Both Inland and coastal locations saw the greatest reductions in mean RMSE performance when the wind direction of the training period was opposite that of the evaluation period. One outlier in this regard which saw the least difference in performance was the set of locations on the Pacific Coast. One potential explanation for this was the relative homogeneity of air masses affecting the Pacific Coast compared to the other locations, with wind direction having less of an effect on the type of air masses observed. Support for the hypothesis that air mass type influenced predictor/predictand relationships in the downscaling model was found in downscaling results for the inland locations. For the inland locations, the T_{We} and T_N training periods had better mean RMSE performance with each other's evaluation periods (T_{We} with V_N and T_N with V_{WE}) than with either the T_S or T_E evaluation periods. This was important because both the westerly and northerly wind directions for the inland locations are typically associated with continental air masses, while the southerly and easterly wind directions are not. Under this premise, wind direction of the training period was serving as a stand-in for the prevailing air mass regime of the training period.

3.4.4 Further Research

Improved predictor variable optimization, a longer period of data for training and evaluation, and additional station locations are areas for potential improvement and reduction of uncertainty for the study. One advantage of a longer period of study would be an increase in the number of days available for evaluation with the wind speed threshold.

Seasonal Effects

The frequency of wind direction as well as the type of air masses associated with each wind direction can be dependent on the season. Though examining potential consequences of these factors was not within the scope of this study, it is an area that could be important for ensuring optimal downscaling performance with the weather typing scheme.

Cardinal Versus Coastal Direction

For this study, wind directions were defined either relative to the axis of the coast for coastal locations or by the four cardinal directions for the inland locations. A further complication not considered in this study is that cardinal direction of the wind can influence the type of air mass seen in coastal locations. This is particularly relevant for the locations on the Great Lakes, which were grouped with the coastal locations in this study. Examining the role played by cardinal direction of wind combined with coastal direction of wind is an area that could provide additional understanding of how wind direction influences downscaling performance for certain locations.

Non-linearities in Statistical Relationships

More complex downscaling models that are not limited to linear relationships between predictors and predictands may be less affected by the problems encountered by MLR with changing wind directions. Additional improvements in downscaling performance may be realized through further research in this area.

Conclusions

The objective of the dissertation was to examine the effects of weather classification on regression-based downscaling of daily temperature extrema. Three different studies were conducted, with each study using a different form of weather classification. Results from each study indicated that downscaling performance was influenced by weather conditions seen in the training periods, and that weather classification had the potential to improve downscaling performance, depending on the circumstance.

The first study used weather classification based off time of day of daily temperature extrema. Results from the first study demonstrated that information about the time of day that the daily temperature extrema occurred was necessary for optimal downscaling performance on days where the diurnal temperature cycle was dominated by advective forcing.

Weather classification was based on temperature for the second study. Results from the second study indicated that downscaling performance could be greatly affected if the temperature conditions of the evaluation period were far outside of training period, with potential consequence for downscaling future climate model output.

The final study used weather classification based off wind direction. Results from the final study indicated that wind direction of the training periods had substantial effects on downscaling performance. As was the case with the second study, the effects were largest when the differences in training and evaluation conditions were greatest.

With respect to future research, there were several possible areas of focus. For the first study, exploring consequences of changes in the diurnal temperature cycles as the climate changes was identified as an area for further investigation. For the second and third studies, an area for future research was examining potential non-linearities in statistical relationships between daily temperature extrema and other meteorological variables. Exploring the influence of the seasons on statistical relationships used for downscaling was another potential subject for further research that was identified in the second and third studies.

Citations

Andreas, W., and B. Â. r. Gerd, 1998: Linear versus nonlinear techniques in downscaling. *Climate Research*, **10,** 83-93.

Barsugli, J. J., and Coauthors, 2013: The Practitioner's Dilemma: How to Assess the Credibility of Downscaled Climate Projections. *Eos, Transactions American Geophysical Union*, **94,** 424-425.

Bettolli, M. L., 2021: Analog Models for Empirical-Statistical Downscaling. Oxford University Press.

Bettolli, M. L., and O. C. Penalba, 2018: Statistical downscaling of daily precipitation and temperatures in southern La Plata Basin. *International Journal of Climatology*, **38,** 3705-3722.

Brands, S., S. Herrera, D. San-Martín, and J. M. Gutiérrez, 2011a: Validation of the ENSEMBLES global climate -models over southwestern Europe using probability density functions, from a downscaling perspective. *Climate Research*, **48,** 145-161.

Brands, S., J. J. Taboada, A. S. Cofiño, T. Sauter, and C. Schneider, 2011b: Statistical downscaling of daily temperatures in the NW Iberian Peninsula from global climate models: validation and future scenarios. *Climate Research*, **48,** 163-176.

Cammarano, D., M. Rivington, K. B. Matthews, D. G. Miller, and G. Bellocchi, 2017: Implications of climate model biases and downscaling on crop model simulated climate change impacts. *European Journal of Agronomy*, **88,** 63-75.

Camus, P., A. Rueda, F. J. Méndez, and I. J. Losada, 2016: An atmospheric-to-marine synoptic classification for statistical downscaling marine climate. *Ocean Dynamics*, **66,** 1589-1601.

Casanueva, A., S. Herrera, J. Fernández, M. D. Frías, and J. M. Gutiérrez, 2013: Evaluation and projection of daily temperature percentiles from statistical and dynamical downscaling methods. *Natural Hazards and Earth System Sciences*, **13,** 2089-2099.

Cortesi, N., 2014: *Analogues and Weather Typing Downscaling methods.*

Coulibaly, P., Y. B. Dibike, and F. Anctil, 2005: Downscaling Precipitation and Temperature with Temporal Neural Networks. *Journal of Hydrometeorology*, **6,** 483-496.

Dabhi, H., M. W. Rotach, M. Dubrovský, and M. Oberguggenberger, 2021: Evaluation of a stochastic weather generator in simulating univariate and multivariate climate extremes in different climate zones across Europe. *Meteorologische Zeitschrift*, **30,** 127-151.

Dirksen, M., W. H. Knap, G.-J. Steeneveld, A. A. M. Holtslag, and A. M. G. K. Tank, 2020: Downscaling daily air-temperature measurements in the Netherlands. *Theoretical and Applied Climatology*, **142,** 751-767.

Dixon, K. W., and Coauthors, 2016: Evaluating the stationarity assumption in statistically downscaled climate projections: is past performance an indicator of future results? *Climatic Change*, **135**.

Duhan, D., and A. Pandey, 2015: Statistical downscaling of temperature using three techniques in the Tons River basin in Central India. *Theoretical and Applied Climatology*, **121,** 605-622.

Erlandsen, H. B., K. M. Parding, R. Benestad, A. Mezghani, and M. Pontoppidan, 2020: A Hybrid Downscaling Approach for Future Temperature and Precipitation Change. *Journal of Applied Meteorology and Climatology*, **59,** 1793-1807.

Fan, X., L. Jiang, and J. Gou, 2021: Statistical downscaling and projection of future temperatures across the Loess Plateau, China. *Weather and Climate Extremes*, **32,** 100328.

Flint, L., and A. Flint, 2012: Downscaling future climate scenarios to fine scales for hydrologic and ecological modeling and analysis. *Ecological Processes*, **1,** 1-15.

Gaitan, C. F., W. W. Hsieh, A. J. Cannon, and P. Gachon, 2014: Evaluation of Linear and Non-Linear Downscaling Methods in Terms of Daily Variability and Climate Indices: Surface Temperature in Southern Ontario and Quebec, Canada. *Atmosphere-Ocean*, **52,** 211-221.

Gough, W. A., A. Žaknić-Ćatović, and A. Zajch, 2020: Sampling frequency of climate data for the determination of daily temperature and daily temperature extrema. *International Journal of Climatology*, **40,** 5451-5463.

Goyal, M. K., and C. S. P. Ojha, 2011: Downscaling of surface temperature for lake catchment in an arid region in India using linear multiple regression and neural networks. *International Journal of Climatology*, **32,** 552-566.

Grouillet, B., D. Ruelland, P. Vaittinada Ayar, and M. Vrac, 2016: Sensitivity analysis of runoff modeling to statistical downscaling models in the western Mediterranean. *Hydrol. Earth Syst. Sci.*, **20,** 1031-1047.

Gutiérrez, J. M., D. San-Martín, S. Brands, R. Manzanas, and S. Herrera, 2013: Reassessing statistical downscaling techniques for their robust application under climate change conditions. *Journal of Climate*, **26,** 171-188.

Hernanz, A., J. A. García-Valero, M. Domínguez, P. Ramos-Calzado, M. A. Pastor-Saavedra, and E. Rodríguez-Camino, 2021: Evaluation of statistical downscaling methods for climate change projections over Spain: Present conditions with perfect predictors. *International Journal of Climatology*, **n/a**.

Hernanz, A., J. A. García-Valero, M. Domínguez, P. Ramos-Calzado, M. A. Pastor-Saavedra, and E. Rodríguez-Camino, 2022: Evaluation of statistical downscaling methods for climate change projections over Spain: Present conditions with perfect predictors. *International Journal of Climatology*, **42,** 762-776.

Huth, R., 2002: Statistical downscaling of daily temperature in central Europe. *Journal of Climate*, **15,** 1731-1742.

Huth, R., 2004: Sensitivity of Local Daily Temperature Change Estimates to the Selection of Downscaling Models and Predictors. *Journal of Climate*, **17,** 640-652.

Huth, R., S. Kliegrová, and L. Metelka, 2007: Non-linearity in statistical downscaling: does it bring an improvement for daily temperature in Europe? *International Journal of Climatology*, **28,** 465-477.

Jang, S., and M. L. Kavvas, 2015: Downscaling Global Climate Simulations to Regional Scales: Statistical Downscaling versus Dynamical Downscaling. *Journal of Hydrologic Engineering*, **20,** A4014006.

Khan, M. S., P. Coulibaly, and Y. Dibike, 2006: Uncertainty analysis of statistical downscaling methods. *Journal of Hydrology*, **319,** 357-382.

Kim, M., S. T. Kim, and Y. Jeong, 2020: Weather Generator–Based Downscaling of EAWM Strength Prediction to the Climate of a Korean Basin. *Journal of Applied Meteorology and Climatology*, **59,** 1581-1605.

Kostopoulou, E., C. Giannakopoulos, C. Anagnostopoulou, K. Tolika, P. Maheras, M. Vafiadis, and D. Founda, 2007: Simulating maximum and minimum temperature over Greece: a comparison of three downscaling techniques. *Theoretical and Applied Climatology*, **90,** 65-82.

Lanzante, J. R., K. W. Dixon, M. J. Nath, C. E. Whitlock, and D. Adams-Smith, 2018: Some Pitfalls in Statistical Downscaling of Future Climate. *Bulletin of the American Meteorological Society*, **99,** 791-803.

Lin, G.-F., M.-J. Chang, and J.-T. Wu, 2017: A Hybrid Statistical Downscaling Method Based on the Classification of Rainfall Patterns. *Water Resources Management*, **31,** 377-401.

Manzanas, R., and Coauthors, 2018: Dynamical and statistical downscaling of seasonal temperature forecasts in Europe: Added value for user applications. *Climate Services*, **9,** 44-56.

Maraun, D., 2013: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate*, **26,** 2137-2143.
Merkenschlager, C., and E. Hertig, 2019: A circulation-based approach considering non-stationarities within the scope of statistical downscaling: An example of seasonal daily precipitation extremes in the Mediterranean area. *International Journal of Climatology*, **39,** 1912-1926.
Merkenschlager, C., S. Koller, C. Beck, and E. Hertig, 2021: Assessing local daily temperatures by means of novel analog approaches: a case study based on the city of Augsburg, Germany. *Theoretical and Applied Climatology*, **145,** 31-46.
Mesinger, F., and Coauthors, 2006: North American regional reanalysis. *Bulletin of the American Meteorological Society*, **87,** 343-360.
Miksovsky, J., and A. Raidl, 2005: Testing the performance of three nonlinear methods of time seriesanalysis for prediction and downscaling of European daily temperatures. *Nonlin. Processes Geophys.*, **12,** 979-991.
Nojarov, P., 2015: Statistical downscaling of regional climate models in Bulgarian mountains and some projections. *Theoretical and Applied Climatology*, **119,** 83-98.
Peterson, T. C., and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology*, **18,** 1493-1517.
Pichuka, S., and R. Maity, 2018: Development of a time-varying downscaling model considering non-stationarity using a Bayesian approach. *International Journal of Climatology*, **38,** 3157-3176.
Pichuka, S., S. S. Maity, and R. Maity, 2022: Benefit of time-varying downscaling model for the assessment of urban temperature rise. *Modeling Earth Systems and Environment*, **8,** 2431-2447.
Preisendorfer, R. W., and C. D. Mobley, 1988: *Principal Component Analysis in Meteorology and Oceanography.* Elsevier.
Radan, H., 1999: Statistical downscaling in central Europe: evaluation of methods and potential predictors. *Climate Research*, **13,** 91-101.
Ribalaygua, J., L. Torres, J. Pórtoles, R. Monjo, E. Gaitán, and M. R. Pino, 2013: Description and validation of a two-step analogue/regression downscaling method. *Theoretical and Applied Climatology*, **114,** 253-269.
Robertson, A. W., A. V. M. Ines, and J. W. Hansen, 2007: Downscaling of Seasonal Precipitation for Crop Simulation. *Journal of Applied Meteorology and Climatology*, **46,** 677-693.

Sachindra, D., and B. Perera, 2016: Statistical downscaling of general circulation model outputs to precipitation accounting for non-stationarities in predictor-predictand relationships. *PloS one*, **11,** e0168701.

Salvi, K., S. Ghosh, and A. R. Ganguly, 2016: Credibility of statistical downscaling under nonstationary climate. *Climate Dynamics*, **46,** 1991-2023.

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*, **21,** 773-790.

Semenov, M., and E. Barrow, 1997: USE OF A STOCHASTIC WEATHER GENERATOR IN THE DEVELOPMENT OF CLIMATE CHANGE SCENARIOS. *Climatic Change*, **35,** 397-414.

Solman, S. A., and M. N. Nuñez, 1999: Local estimates of global climate change: a statistical downscaling approach. *International Journal of Climatology*, **19,** 835-861.

Tabari, H., S. M. Paz, D. Buekenhout, and P. Willems, 2021: Comparison of statistical downscaling methods for climate change impact analysis on precipitation-driven drought. *Hydrol. Earth Syst. Sci.*, **25,** 3493-3517.

Takayabu, I., H. Kanamaru, K. Dairaku, R. Benestad, H. von Storch, and J. H. Christensen, 2015: Reconsidering the quality and utility of downscaling. *J. Meteor. Soc. Japan*, **94A,** 31-45.

Timbal, B., and B. J. McAvaney, 2001: An analogue-based method to downscale surface air temperature: application for Australia. *Climate Dynamics*, **17,** 947-963.

Timbal, B., A. Dufour, and B. McAvaney, 2003: An estimate of future climate change for western France using a statistical downscaling technique. *Climate Dynamics*, **20,** 807-823.

Van Uytven, E., J. De Niel, and P. Willems, 2020: Uncovering the shortcomings of a weather typing method. *Hydrol. Earth Syst. Sci.*, **24,** 2671-2686.

Vesely, F. M., L. Paleari, E. Movedi, G. Bellocchi, and R. Confalonieri, 2019: Quantifying Uncertainty Due to Stochastic Weather Generators in Climate Change Impact Studies. *Scientific Reports*, **9**.

Vincent, L. A., X. Zhang, B. R. Bonsal, and W. D. Hogg, 2002: Homogenization of Daily Temperatures over Canada. *Journal of Climate*, **15,** 1322-1334.

Vincent, L. A., E. J. Milewska, R. Hopkinson, and L. Malone, 2009: Bias in Minimum Temperature Introduced by a Redefinition of the Climatological Day at the Canadian Synoptic Stations. *Journal of Applied Meteorology and Climatology*, **48,** 2160-2168.

Vrac, M., M. Stein, K. Hayhoe, and X. Z. Liang, 2007: A general method for validating statistical downscaling methods under future climate change. *Geophysicial Research Letters*, **34**.
Wang, Y., G. Sivandran, and J. M. Bielicki, 2018: The stationarity of two statistical downscaling methods for precipitation under different choices of cross-validation periods. *International Journal of Climatology*, **38,** e330-e348.
Wigley, T. M. L., P. D. Jones, K. R. Briffa, and G. Smith, 1990: Obtaining sub-grid-scale information from coarse-resolution general circulation model output. *Journal of Geophysical Research: Atmospheres*, **95,** 1943-1953.
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,** -.
Žaknić-Ćatović, A., and W. A. Gough, 2021: Identification of radiative and advective populations in Canadian temperature time series using the Linear Pattern Discrimination algorithm. *International Journal of Climatology*, **41,** 5100-5124.
——, 2022: Seasonal Aspects of Radiative and Advective Air Temperature Populations: A Canadian Perspective. *Atmosphere*, **13,** 1017.
Zorita, E., and H. von Storch, 1999: The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods. *Journal of Climate*, **12,** 2474-2489.

VITA

The author, Quinn Pallardy, is a native of Columbia, Missouri. Quinn is a long-time fan of Missouri sports, and enjoys bike rides, reading, programming, and video games.

Quinn has been interested in the weather since an early age, which is why he decided to study atmospheric science. Many aspects of atmospheric science have interested Quinn, including but not limited to climate science, hurricanes, tornados, radar, and numerical weather modeling.