BREE Research Paper

Ann Marie Matheny

July 31, 2019

Contents

1 Introduction 2
1.1 Objectives 2

2 Methods 2
2.1 Data 2
2.2 Bias Correction 4
2.3 Analyses 5

3 Results 5

4 Discussion 9

5 Conclusion 9

6 Appendix 10

7 Acknowledgements 18

1 Introduction

The Lake Champlain watershed, a region in northwestern Vermont, upstate New York, and part of Quebec, CA relies upon the hydrological system for ecological, economical, agricultural, and social needs. However, the integrity of the region's economy and socio-economic system is at risk to the impacts of extreme climate events resulting from global climate change. Recently, the area has experienced extreme flooding events, algae blooms and increases in summer temperatures, which are all linked to global climate change [4]. For instance, heavy precipitation can increase runoff to lakes and rivers, contributing to toxic algae growth and damage to the clean water supply, while severe heat waves are especially detrimental to vulnerable populations [7]. To plan for and mitigate changes in patterns of extreme events, accurate future climate projections of the Lake Champlain Basin are necessary.

Regional climate models (RCMs) are essential for generating future climate projections over regions such as the Lake Champlain basin. [5]. Despite recent advances in modeling, RCM climate projections, and especially those for extreme events, are characterized by nontrivial degrees of bias [5]. To correct the inherent bias in RCM climate projections, various bias correction methods are typically applied as a post-processing step [3]. Quantile mapping is a common bias correction technique that involves adjusting model distributions of climate variables such as maximum temperature to match those of observed data [3]. This technique has been used successfully in many other studies including research applications to which quantile mapping has been shown to better correct the mean and variance of model projections than simpler methods [3].

In this study, we compared extreme maximum temperature distributions from station data to those of RCMs and applied three implementations of bias correction (empirical quantile mapping) to RCM simulations based on station data. Our study is a first step in a larger research project investigating the changes in patterns of extreme events in the Lake Champlain Basin.

1.1 Objectives

Our primary objective was to compare patterns of extreme distributions of RCM daily maximum temperature (TMAX) projections and climate station data. We defined an extreme temperature event as the top 90th percentile for TMAX, similar to B. R. Bonsal et al. (2001) [2]. In particular, we investigated how extreme distributions varied temporally over a historical time period (1980-2014) and spatially over the Lake Champlain Basin. Our second objective was to compare the performance of three implementations of quantile mapping in correcting RCM extreme temperature distributions: annual correction, monthly correction, and bias correction based on a historic calibration period.

2 Methods

2.1 Data

The Lake Champlain Basin study area comprises the majority of Vermont and New Hampshire, as well as parts of upstate New York and southern Quebec (Figure 1). Climate projections for TMAX were generated by the Weather Research and Forecasting (WRF) Model, a regional climate model designed for atmospheric research and operational forecasting applications [6]. WRF output is a gridded climate product, and our simulations were generated at a 4 km resolution (Figure 2). We obtained observational data for TMAX from the Global Historical Climate Network (GHCN), a network of global climate station data checked for measurement and calibration error [1]. We used daily WRF simulations and GHCN station data from 1980-2014. Prior to analyses, we extracted only stations that had at least 70% complete records from 1980-2014 (73 stations). We then obtained the nearest WRF grid cell to each of the 73 stations to obtain 73 station-WRF "pairs".

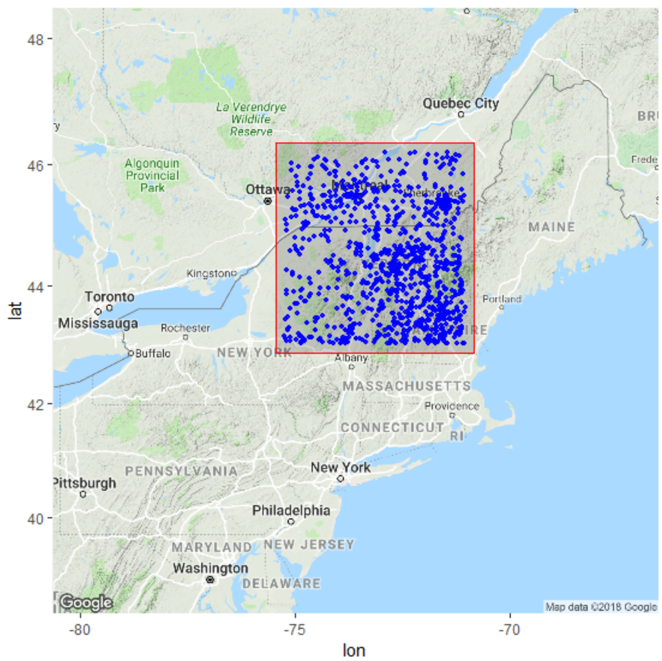

Figure 1: Global Historical Climate Network (HCN) weather stations in Lake Champlain Basin

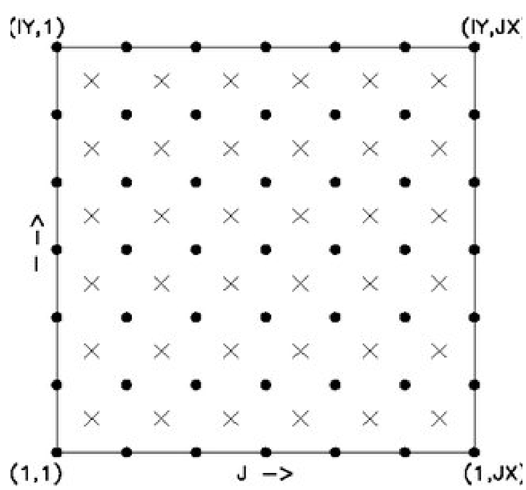

Figure 2: Illustration of WRF 4x4 km resolution grid system

2.2 Bias Correction

To quantify the distributional similarity between WRF projections and station data, we used the Perkins PDF skill score [8]. The Perkins Skill score quantifies the overlap between probability density functions of the model (m), and observation data (o) for a climate variable. To calculate the skill score, a discrete density histogram with *i* bins is made for both model and station data. The score is the sum of the minimum of the normalized density of model and station data over all *i* bins (eg 1). [8]. This equation can be expressed below:

$$S_{score} = \sum_{i=1}^{n} min(Z_{m_i}, Z_{o_i} \tag{1}$$

The Perkins skill score ranges between 0 and 1, where 1 indicates a perfect distributional match, and 0 denotes no match. We used three implementations of the empirical quantile mapping bias correction method to adjust distributions of each of the 73 WRF grid points based on the nearest station. In the first approach, annual correction (BCA) , we adjusted WRF extreme distributions for TMAX based on those of the entire time series of station data (eq 2)

$$P_{corr,t} = ecdf^{-1}_{obs,a}(ecdf_{raw,a}(P_{raw,t})), \tag{2}$$

where $P_{corr,t}$ was the corrected daily TMAX, $ecdf^{-1}_{obs,a}$ was the inverse empirical CDF of the entire distribution of TMAX station data over the historical time period,$ecdf_{raw,a}$ was the empirical CDF of the entire distribution of WRF TMAX projections, and $P_{raw,t}$ was the raw WRF daily TMAX value.

Next, we applied quantile mapping on a monthly basis(BCM). Instead of using CDFs of the entire annual time series to correct WRF simulations, we corrected daily TMAX values using the monthly distribution corresponding to the month of the value to be corrected (eq 3)

$$P_{corr,t} = ecdf^{-1}_{obs,m}(ecdf_{raw,m}(P_{raw,t})), \tag{3}$$

where $P_{corr,t}$ was the corrected daily TMAX, $ecdf^{-1}_{obs,m}$ was the inverse empirical CDF of the monthly distribution of TMAX station data over the historical time period, $ecdf_{raw,m}$ was the empirical CDF of the monthly distribution of WRF TMAX projections, and $P_{raw,t}$ was the raw WRF daily TMAX value.

In the third approach, we partitioned the historical data into calibration periods (1980-1995) and testing periods (1996-2014). We obtained the quantile mapping transfer function from the historical data calibration period, and then we applied the function to the WRF simulations only for the testing period (eq 4). Similar to BCM, we the correction was done on a monthly basis. We will refer to this method as BCT. The equation for BCT is as follows:

$$P_{corr:future,t} = ecdf^{-1}_{obs:past,m}(ecdf_{raw:future,m}(P_{raw:future,t})) \tag{4}$$

In (eq 4), $P_{corr:future,t}$ was the corrected daily TMAX value in the testing period, $ecdf^{-1}_{obs:past,m}$ was the inverse empirical CDF of the distribution of monthly TMAX station data over the training period from 1980-1994, $ecdf_{raw:future,m}$ was the monthly distribution of WRF TMAX in the test period 1995-2014, and $P_{raw:future,t}$ was the raw WRF daily TMAX value in the testing period.

2.3 Analyses

To determine how skill scores changed over time, we calculated the mean skill score over all WRF-station pairs for consecutive 5-year increments (e.g. 1980-1984, 1985-1989, 1990-1994, etc). To inspect changes in spatial patterning of skill scores, we plotted the skill scores for all WRF-station pairs on a map of the study area for each 5-year increment.

To determine the performance of the three bias correction implementations, we inspected the overall mean of the Perkins skill scores over time for all three bias correction implementations. In addition, we examined histograms of station data, raw WRF simulations, and and bias corrected WRF simulations for all WRF-station pairs over all 5-year time increments.

3 Results

We found that the similarity between extreme (90th quantile) TMAX distributions, based on Perkins skill scores, among the 73 WRF-station pairs changed only slightly over time. However, we did observe considerable spatial patterns in Perkins skill scores, and these patterns were consistent over the 35-year time period. BCM and to a lesser degree, BCA, resulted in substantial increases in Perkins skill scores, while the BCT implementation resulted in only marginal improvements in skill scores.

Overall, the average skill score for 1980-2014 ranged between 0.63 and 0.7 with a mean of 0.68. The cluster of stations in the southeast quadrant of the study area increased slightly over time (Figs 3 and 4).

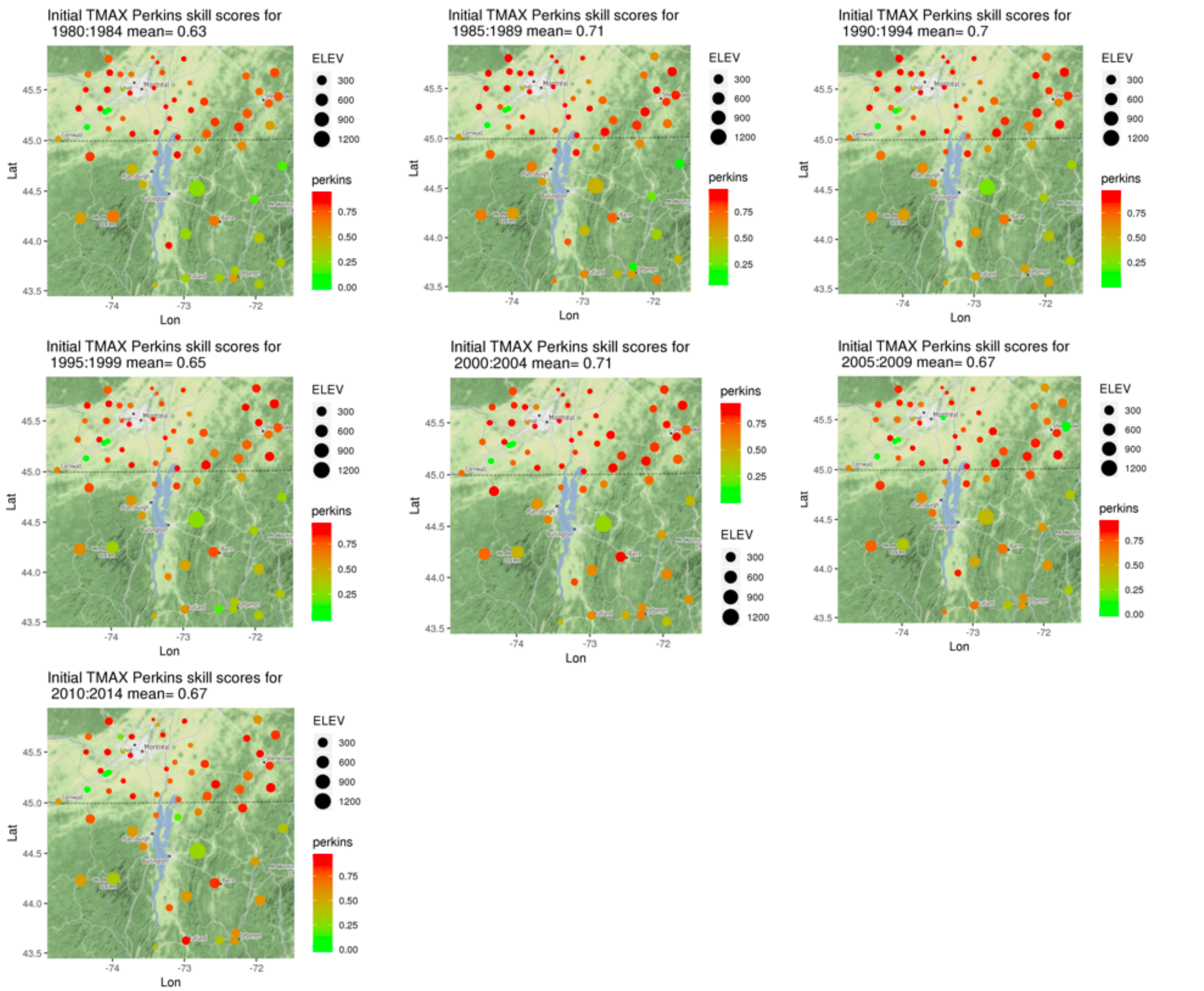

Figure 3: Raw Perkins score before bias correction implementation over the study area in 5-year increments

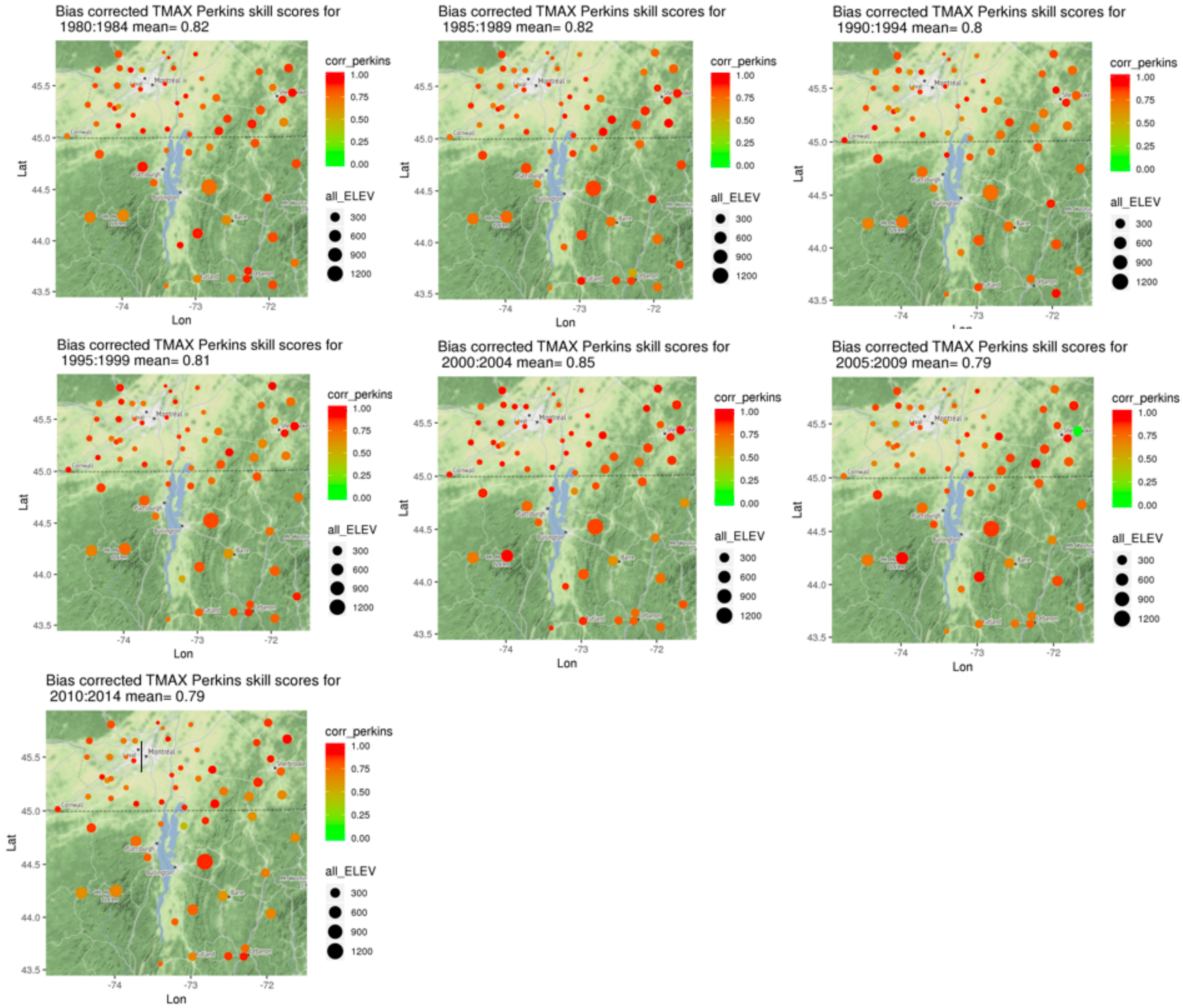

Figure 4: BCA corrected scores over the study area in 5-year increments

Generally, Perkins skill scores for both raw and bias corrected data were consistently greater in the northern portion of the study area (southern Quebec) compared to southern portion of the study area (Figures 3 and 4). There was variation of skill scores in the southeast quadrant of the study area across time for the raw data, but this variability decreased after bias correction was implemented. The skill score for the highest elevation station and nearest WRF grid point pair was always lower than most other locations over the entire historical time period. The skill score for the WRF-station pair east of St. Alban's Bay, in contrast, was very high (around 0.88), until the most recent time period (2010-2014), when it dropped to 0.32.

Average skill scores increased from 0.68 to 0.82 and 0.81, for BCM and BCA, respectively. After the application of BCM to WRF simulations, skill scores ranged from 0.81 to 0.85, with a mean of 0.82. Following the application of BCA, skill scores also increased for the entire time period, with the average skill score ranging between 0.79 to 0.85, with a mean of 0.81. BCT scores ranged from 0.78 to 0.85, with an average score of 0.78. The greatest skill score for a single station-WRF pair occurred during 1980-1984 when the average score increased from 0.63 to 0.82. Please see Table 3 for full results of all time increments and bias correction implementations.

In general, the means of raw WRF 90^{th} percentile distributions of TMAX exhibited a slight cold bias compared to those of station data (Fig. 6). After applying bias correction, the means of corrected WRF data were closer to those of station data. BCM and BCA both performed well in correcting the means, but means resulting from BCT were slightly lower. (See appendix Figures 14 and 13 for additional histograms)

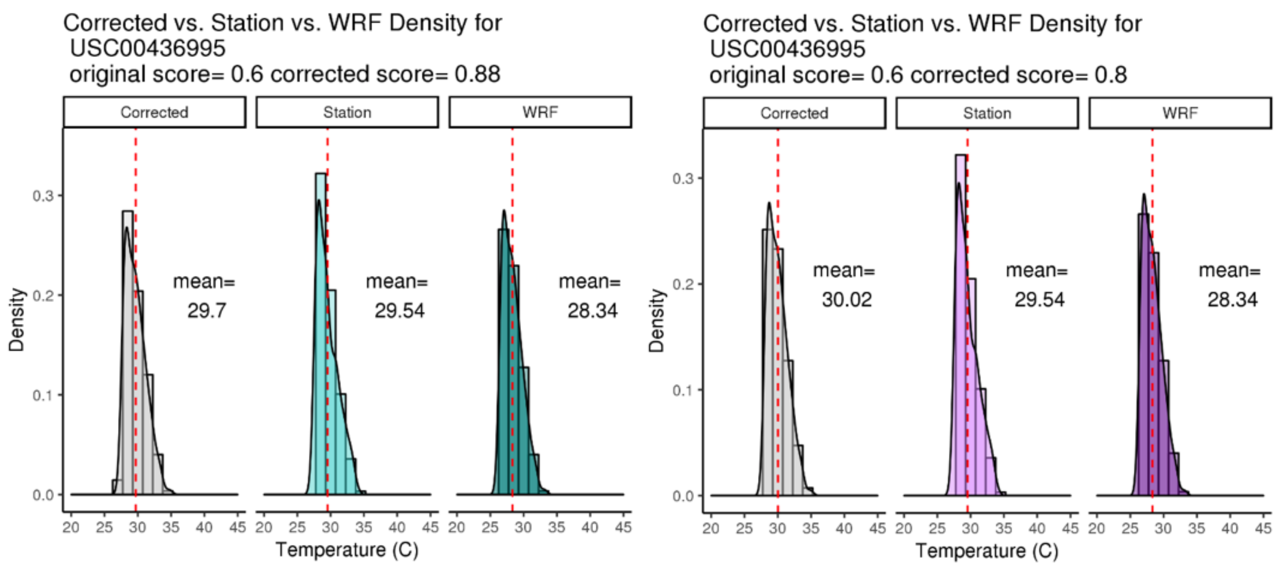

Figure 5: Left: Annually Corrected Data vs. Station vs. WRF 90th Percentile Density Plot for two station-WRF pairs over 2000-2004.

In a few rare cases, the skill score after bias correction for BCT was lower than before bias correction (Figs 13 and 14). However, some station-WRF pairs exhibited enormous improvements in skill scores; the average skill score between station USW00439984 and its nearest WRF grid point increased from 0.32 to 0.86, following BCT. The maximum Perkins score overall was 0.9, after BCT and BCM.

Year Chunk	Avg Initial	Avg BCA	Avg BCM	Avg BCT
1980-1984	0.635	0.820	0.832	0.811
1985-1989	0.708	0.819	0.815	0.790
1990-1994	0.701	0.798	0.813	0.779
1995-1999	0.666	0.811	0.828	0.793
2000-2004	0.707	0.847	0.850	0.846
2005-2009	0.680	0.788	0.820	0.824
2010-2014	0.686	0.791	0.813	0.809

Table 1: Average Perkins scores across year increments for each bias correction approach (*Initial* refers to the raw data before applying any bias correction implementation). BCM produced highest scores

4 Discussion

In this study, we compared the extreme (90th percentile) distributions of TMAX between WRF and station data spatially and temporally using Perkins skill scores. We also compared the effects of three implementations of quantile mapping on Perkins skill scores. Although we did not find considerable changes in Perkins skill scores over time for either raw or bias corrected WRF data, we did find that generally, skill scores were larger in the northern portion of the study area compared to the southern part, where there were more variable station scores over the historical time period. All bias correction implementations resulted in increases in Perkins skill scores, with BCM and BCA resulting in the greatest improvements in skill scores and BCT resulting in modest improvements in skill sores.

The southeast portion of the study area is characterized by complex topography, which, in addition to natural variability in TMAX over 1980-2014 could have played a role in the variability of Perkins skill scores in this area prior to bias correction. However, after bias correction, the variability decreased drastically. This is not altogether surprising; quantile mapping, by definition, corrects model distributions based on station distributions. Interestingly, Perkins skill scores were generally higher in southern Quebec, and this trend was consistent before and after bias correction and over 1980-2014, regardless of the implementation of bias correction. The pattern of higher skill scores in this region could be due to the complex geography and topography of the area. The portion of Canada in our study area was comparatively less topographically complex than the southern portion. There is perhaps less variability in extreme temperature distributions in less topographically complex regions because air temperature varies across regions with known physical influences [9]. WRF is likely better able to capture extreme events in less topographically complex regions. We did investigate Perkins scores across elevations to examine how topography might have effected skill scores, but we did not find any significant trends (Appendix Figure 11).

While all implementations of bias correction resulted in increases in skill scores, BCM produced the greatest increase, followed by BCA and BCT, with BCT resulting in the least amount of improvement. This is not very surprising. BCM/BCA is 'overfitting', as these implementations use the entire time series of station data in adjusting WRF data. BCT, in contrast, corrects WRF data based only on a subset of station data and is a more realistic representation of how quantile mapping would improve future WRF simulations.

5 Conclusion

The Lake Champlain basin is likely to experience significant changes in the patterns of extreme temperature events. The long-term consequences of such changes in extreme maximum temperatures could result in devastating ecological and agricultural impacts. Furthermore, these changes could impact economic industries such as recreation, farming, and maple syrup production. Regional climate models, such as WRF, may produce reasonably accurate projections of the future weather in this region, but they remain biased. We found that quantile mapping improved the similarity between WRF and station distributions for extreme maximum daily temperatures distributions. This study serves as a good first step in helping climate scientists understand, predict, and plan for future extreme temperature events in the Lake Champlain Basin.

6 Appendix

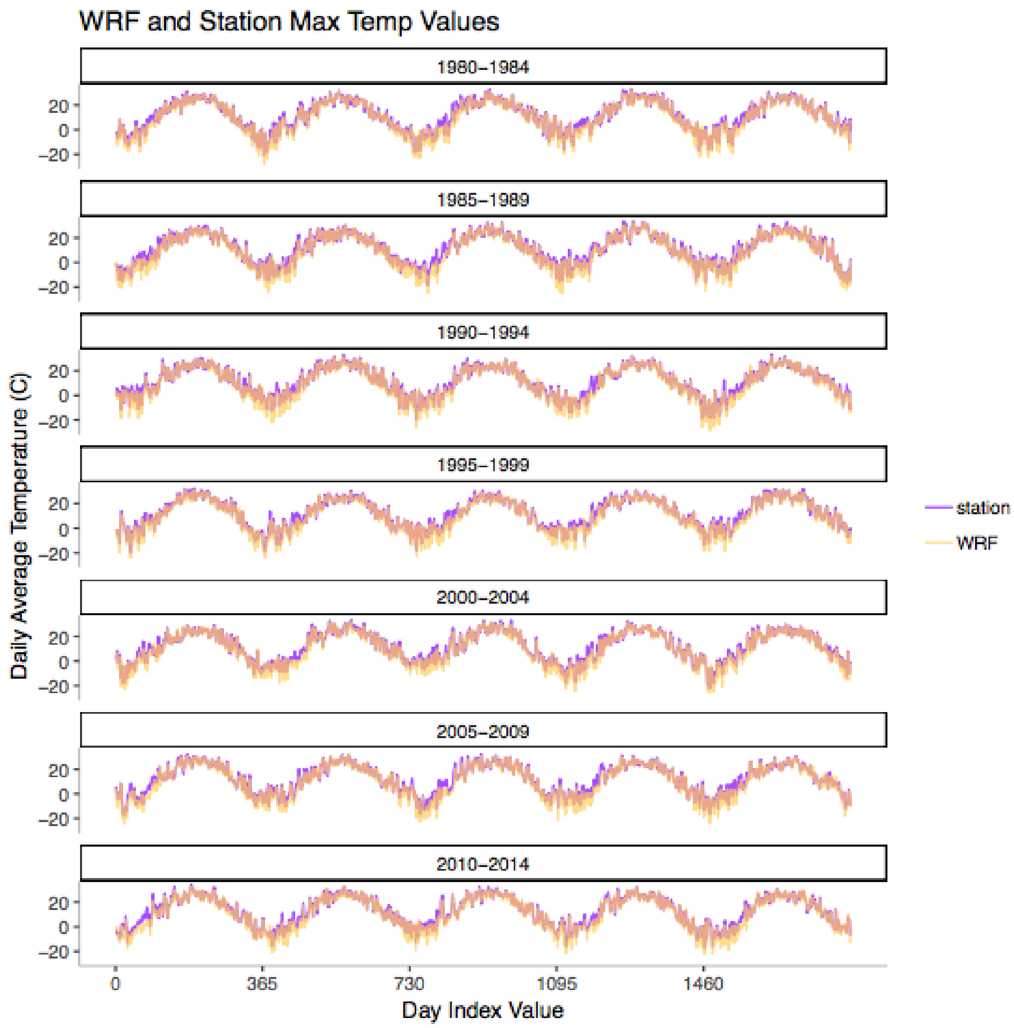

Figure 6: Time Series for daily Station and WRF Max Temperature data from 1980 to 2014 over 5 year periods. One can see that WRF overestimates winter maximum temperatures. As temperatures rise in the spring months, WRF tends to then underestimate daily maximum temperatures.

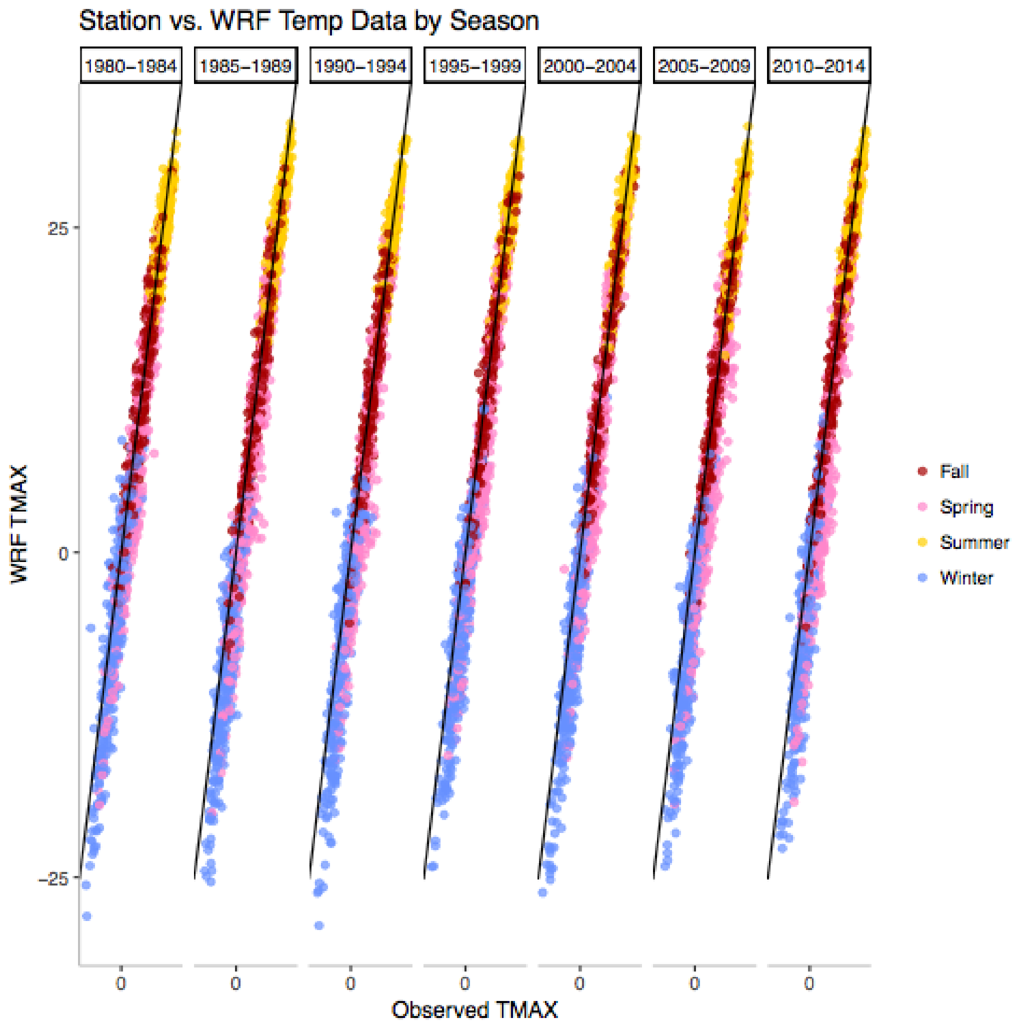

Figure 7: Scatter plots of observed station data versus WRF data in consideration of season. Again, winter is the most variable season across the years for daily maximum temperature.

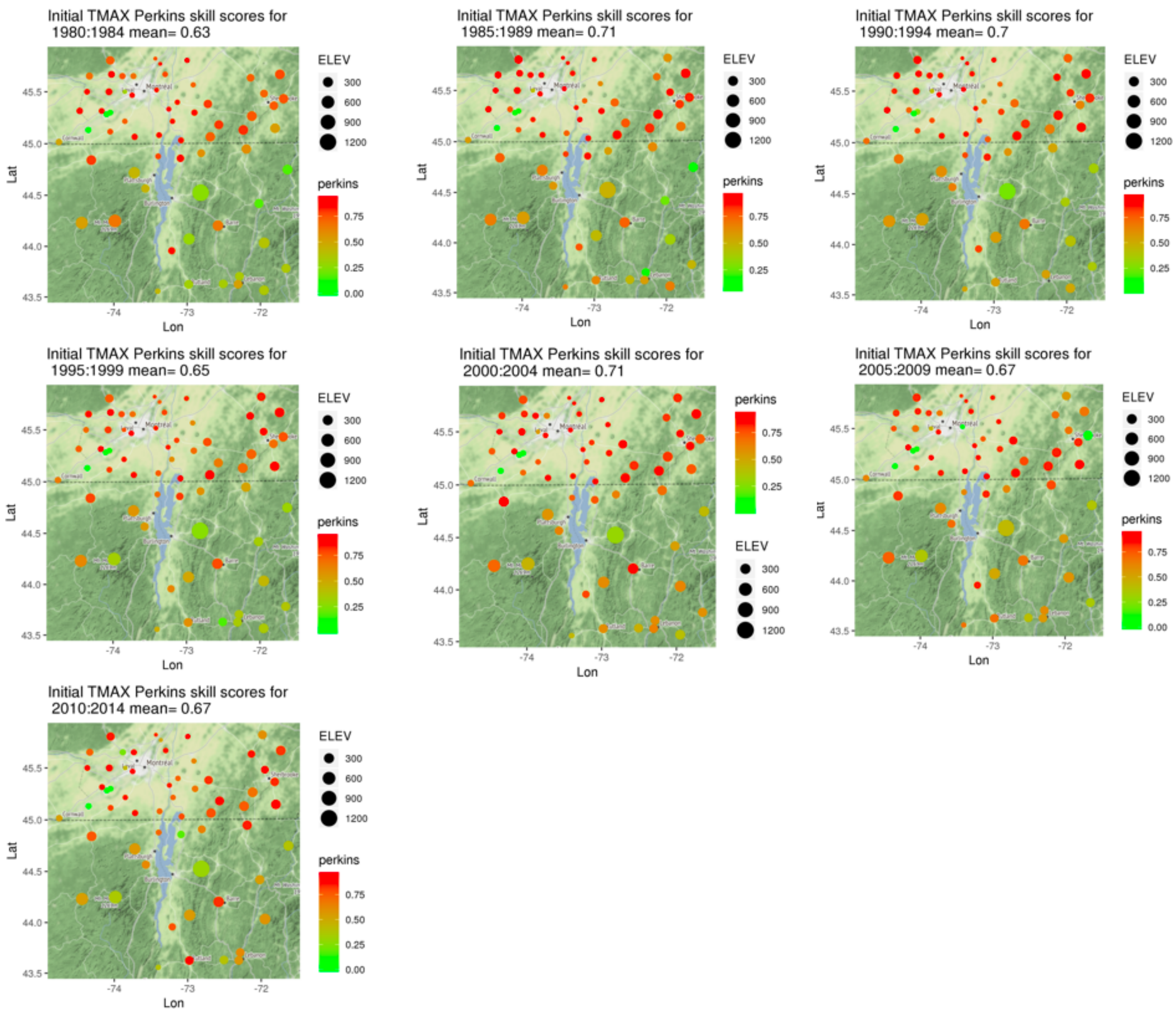

Figure 8: The 7 plots above each correspond to the seven 5-year time chunks from 1980-2014. The maps of the Lake Champlain watershed region pinpoint station locations indicated by different size and colors. Red points imply higher Perkins score in contrast to green points. Bigger points imply the station is located at a higher elevation in contrast to smaller points.

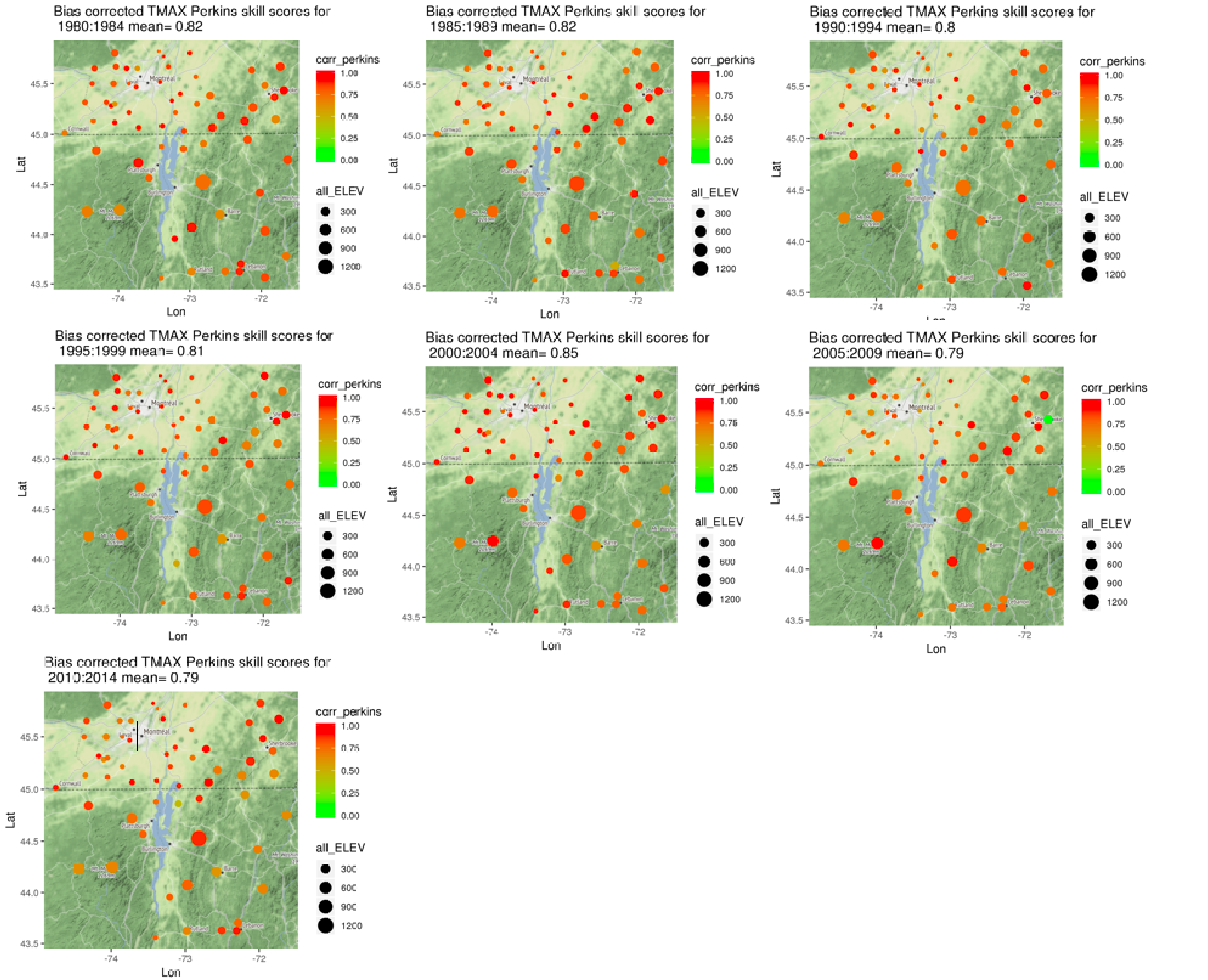

Figure 9: The 7 plots above each correspond to the seven 5-year time chunks from 1980-2014 BCM. The maps of the Lake Champlain watershed region pinpoint station locations indicated by different size and colors. Red points imply higher Perkins score in contrast to green points. Bigger points imply the station is located at a higher elevation in contrast to smaller points. The scores in contrast to before have been bias-corrected. Improvement is evident.

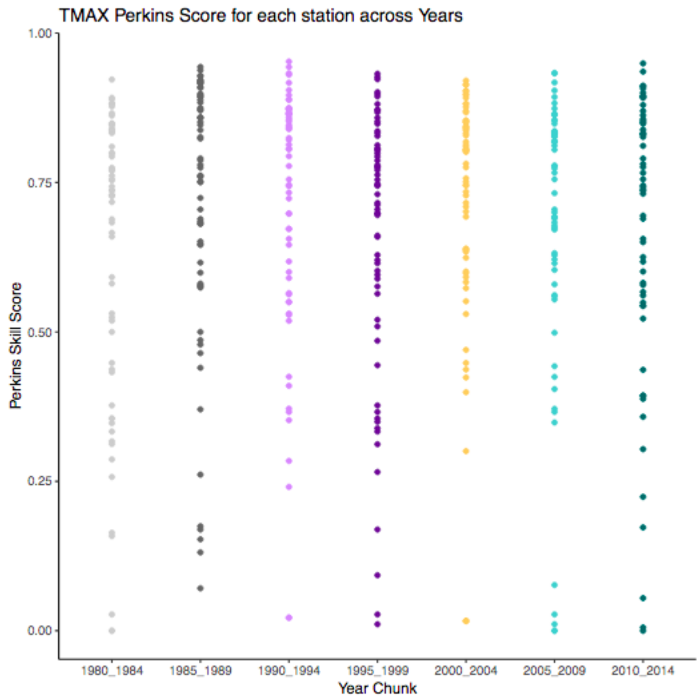

Figure 10: Perkins Skill Score for every station across 35 years

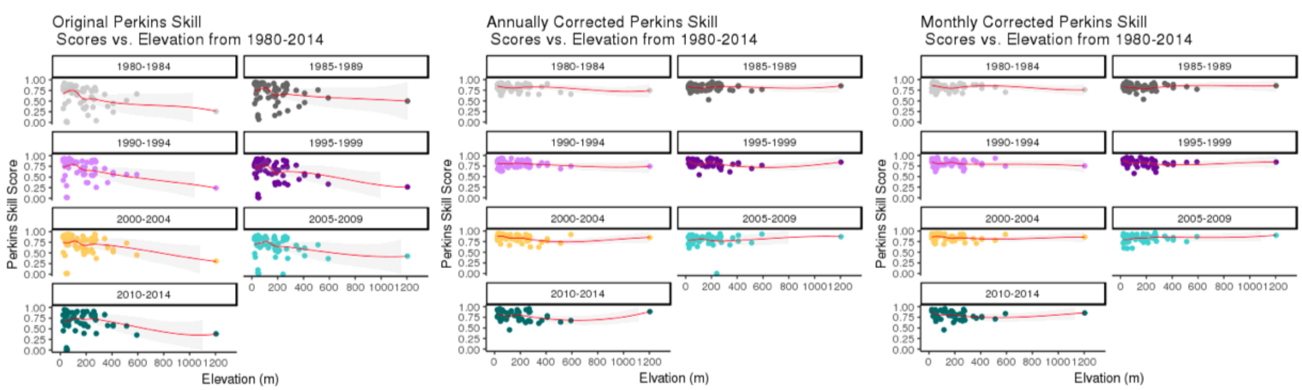

Figure 11: These are scatter plots comparing Perkins scores to station elevation from 1980 to 2014. This setup allows us to see if there exists any association or relationship between these two variables. To the left, the graph displays original data. To the right, the graph displays the bias corrected Perkins Skills Score across time and elevation. In this situation we corrected by year.

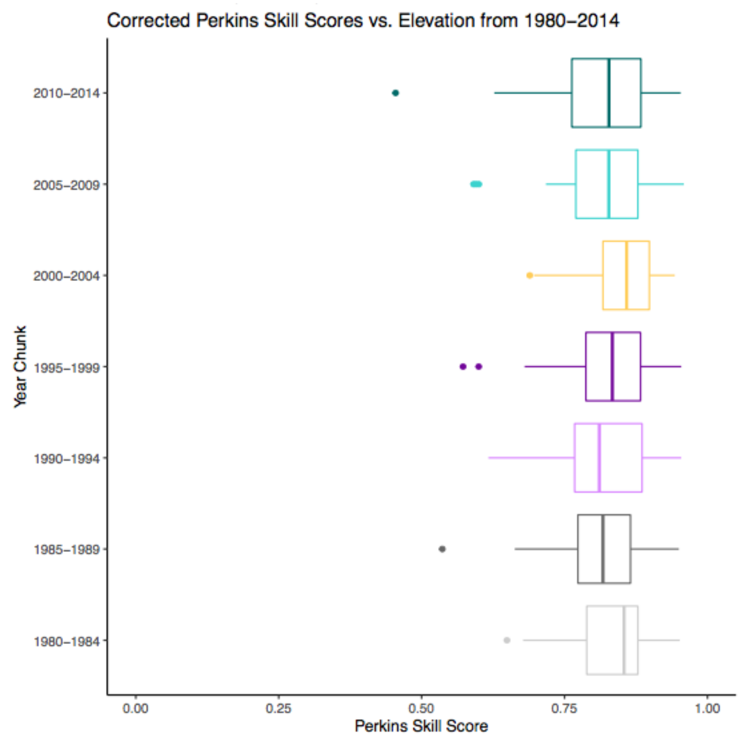

Figure 12: The box and whisker plots displayed show the distribution of Perkins scores for each year chunk. The widest distribution variation occurs from 2010-2014, where as earlier in 2000-2004, the scores where more highly concentrated in the top 70 percent.

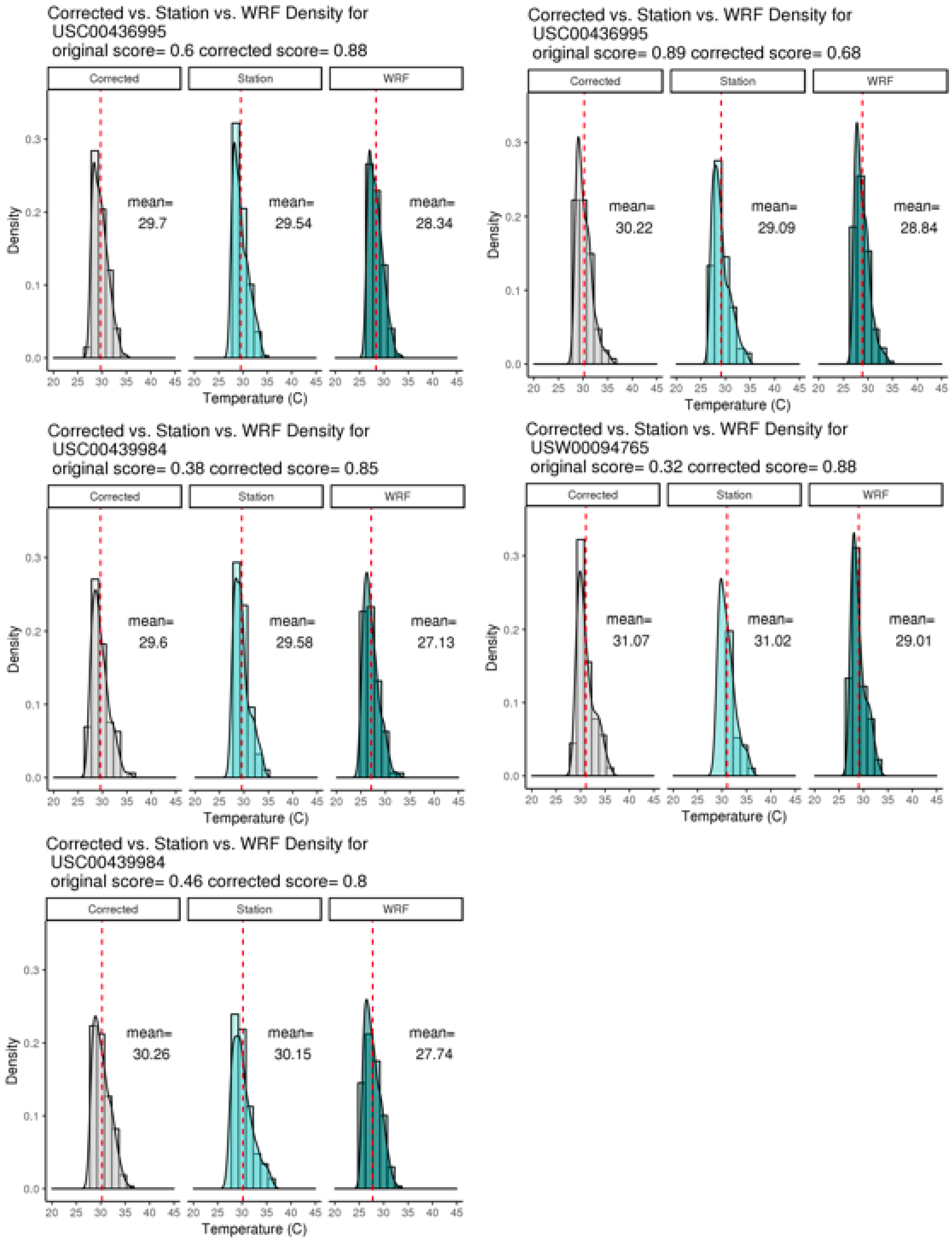

Figure 13: The graphs above are a comparison of the density distributions between the maximum daily temperature for station data, raw WRF data and the bias corrected WRF data as indicated in the different colored plots. The average (mean) Perkins Skills Score is printed on the right of each distribution and indicated by the red, dashed lines.

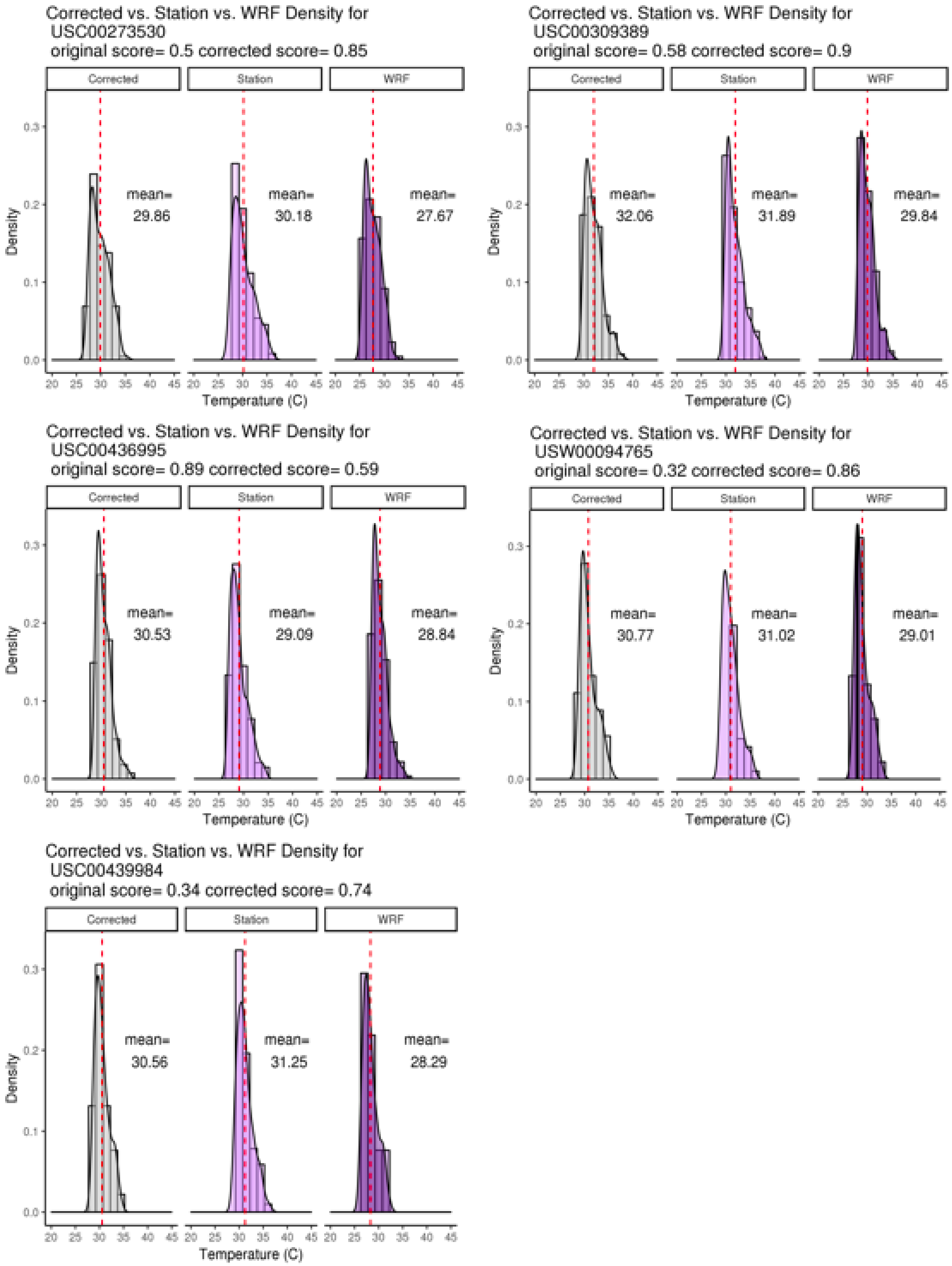

Figure 14: The methods follow the similar process. This time, data was only corrected from 1996-2014 based solely on data from 1980-1995. As before, The average Perkins Skills Score is printed on the right of each distribution and indicated by the red, dashed lines.

7 Acknowledgements

I would personally like to thank Maike Holthuijzen for her mentorship, advice, and guidance, as well as her readily available technical and professional consultation. Many thanks to the members of the BREE Climate team for technical support and access. Thank you to the Center for Workforce Development and Diversity for allowing me this opportunity to grow in my field and further challenge and develop my technical, analytical, and professional skills to bring to the scientific research community.

References

[1] Global historical climatology network (ghcn).

[2] B. R. Bonsal, X. Zhang, L. A. Vincent, and W. D. Hogg. Characteristics of daily and extreme temperatures over canada. *Journal of Climate*, 14(9):1959–1976, May 2001.

[3] Alex J. Cannon, Stephen R. Sobie, and Trevor Q. Murdock. Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, 28(17):6938–6959, Sep 2015.

[4] E. P. Maurer M. D. Dettinger M. Tyree Cayan, D. R. and K. Hayhoe. Climate change scenarios for the california region. *Clim. Change*, 87(42), 2008.

[5] Jie Chen, François P. Brissette, Diane Chaumont, and Marco Braun. Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over north america. *Water Resources Research*, 49(7):4187–4205, Jul 2013.

[6] Xin-Zhong Liang, Min Xu, Xing Yuan, Tiejun Ling, Hyun I. Choi, Feng Zhang, Ligang Chen, Shuyan Liu, Shenjian Su, Fengxue Qiao, and et al. Regional climate–weather research and forecasting model. *Bulletin of the American Meteorological Society*, 93(9):1363–1387, Sep 2012.

[7] Francesco Nordio, Antonella Zanobetti, Elena Colicino, Itai Kloog, and Joel Schwartz. Changing patterns of the temperature–mortality association by time and location in the us, and implications for climate change. *Environment International*, 81:80–86, May 2015.

[8] S. E. Perkins, A. J. Pitman, N. J. Holbrook, and J. Mcaneney. Evaluation of the ar4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over australia using probability density functions. *Journal of Climate*, 20(17):4356–4376, Jan 2007.

[9] K. Stahl, R.d. Moore, J.a. Floyer, M.g. Asplin, and I.g. Mckendry. Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. *Agricultural and Forest Meteorology*, 139(3-4):224–236, 2006.