IMPROVING THE ACCURACY OF BIAS CORRECTED, GRIDDED PRECIPITATION FOR DROUGHT MONITORING

A Thesis

by

ALEX RICHARD SMITH

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	John W. Nielsen-Gammon
Committee Members,	Christopher J. Nowotarski
	Zhe Zhang
Head of Department,	Ramalingam Saravanan

August 2022

Major Subject: Atmospheric Sciences

ABSTRACT

Effective drought monitoring relies upon accurate long-term estimates of precipitation. Stage IV precipitation estimates provide a high spatial and temporal resolution that is effective for short-term applications, but in long-term periods useful for drought monitoring, compounding biases can reduce its accuracy. Errors in the dataset include beam blockage, mean field and range dependent, and two-dimensional biases. This study improves upon a three-step bias adjustment methodology that corrects for the Stage IV biases east of the Rocky Mountains, adds a fourth step to remove discontinuities caused by independent analyses, determines an optimal interpolation method for adjusting the data, and assesses the performance of each step within different regions and accumulation periods. Beam blockage is identified using image filtering and ridge detection software, then adjusted using unblocked data. Mean field and range dependent biases are adjusted using radar estimates, normal precipitation, and rain gauges. Discontinuities are adjusted with an inverse distance weighting (IDW) method that blends the data. Two-dimensional biases are adjusted using gauge-radar biases that are interpolated to the entire precipitation field. Extensive testing of the performance is done using a combination of data withholding and comparison of radar estimates to gauges.

The Stage IV bias adjustments generally result in lower root mean square error (RMSE), median absolute error (MAE) and median bias (MB) compared to gauges. The largest improvements are seen in gauge-based adjustments, which include the mean field and range dependent and two-dimensional steps, and in regions where there are fewer

gauges and greater variations in seasons. The smallest improvements and cases of increased error arise in radar-based adjustments, which include the beam blockage and discontinuity adjustment steps. When testing interpolation methods, IDW is optimal versus inverse distance weighting squared (IDW2) and ordinary kriging (OK) for its lower error and computational expense.

ACKNOWLEDGEMENTS

There were many influential people involved in my graduate school experience who have supported and believed in me in my time here. Firstly, I would like to express my deepest gratitude to Dr. Nielsen-Gammon for giving me the opportunity to work with him since I was an undergraduate student. He helped me believe that I could succeed as a graduate student when I was his REU student, he let me pursue internship opportunities that helped me explore my interests, and he helped guide me into being a better scientist. Additionally, I would like to thank my committee members, Dr. Nowotarski for helping me with edits and new ideas throughout the thesis process, and Dr. Zhang for agreeing to be on my committee last minute and for helping with the final push to the defense.

My research and friend groups have helped strongly support me in my time here. Thank you to my research group for making me feel like I belong here and for always being available to help with anything. Thank you to my friends for making my time here enjoyable. From intramural sports, to trivia nights, frequent hangouts, and concerts, I could not have had the amazing experience I had here without you.

Finally, I would like to thank my family for their unwavering love and support. Thank you to my parents for always being there for me and for frequently visiting and checking in on me. Our many phone calls and texts made me feel like I was still at home. Also, I'd like to thank my sisters for their support by visiting and sending tik toks and memes. Finally, I would like to thank my grandparents. I thoroughly enjoyed all of our dinners in the BCS area and the visits to your house. It made Texas feel like home.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professor John Nielsen-Gammon and Professor Christopher Nowotarski of the Department of Atmospheric Sciences, and Professor Zhe Zhang of the Department of Geography.

Funding Sources

This work was made possible by funding from NOAA Grant Number NA17OAR4310143, NOAA funding from the Southern Regional Climate Center, and the Office of the State Climatologist of Texas.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

Effective drought monitoring is of critical importance to the public because it can allow decision makers to respond quickly and reduce drought impacts (NIDIS 2006). A crucial aspect of improving monitoring is precipitation estimation, but it can be inaccurate due to low and uneven rain gauge spatial resolution and radar estimation biases and errors. Combining gauge observations with radar estimates can improve quantitative precipitation estimates (QPE) used in drought monitoring and overcome the low spatial resolution of gauges. Additionally, satellites can aid in filling in data gaps. However, wind, tipping errors, and under catchment in gauges, precipitation type errors and temporal resolution in satellites, and beam blockage, mean field, range dependent, and two-dimensional biases reduce the accuracy of precipitation estimations. This study improves upon a three-step methodology by McRoberts and Nielsen-Gammon (McRoberts and Nielsen-Gammon 2017, McRoberts 2014) to adjust for radar biases. Additionally, extensively quality controlled rain gauges and an optimal interpolation method are used to combine the radar and gauge estimates. The result is an improved QPE for use in drought monitoring.

This study uses Stage IV precipitation estimates east of the Rocky Mountains as the QPE data to improve upon, as well as quality controlled (QC) rain gauges to compare and aid in bias correction adjustments. Then, the improved bias correction methodology is tested at each step in order to quantify the improvements and optimize the input parameters. The steps for the bias corrections include a beam blockage, mean

field and range dependent, radar domain discontinuity, and two-dimensional step. For the radar-based adjustments, which are the beam blockage and radar domain discontinuity steps, the radar estimates are compared with corresponding rain gauges to assess performance. For the gauge-based steps, the mean field and range dependent and two-dimensional steps, data withholding is used to assess performance. Additionally, in the two-dimensional step, a series of data withholding tests are run to determine an optimal interpolation method and parameters to use in its adjustment between ordinary kriging (OK), inverse distance weighting (IDW), and inverse distance weighting squared (IDW2). With the optimized steps of the correction methodology, the Stage IV data is adjusted. Using the output, seasonal, one-year and three-year accumulation periods are tested in regions of high and low topographic relief, different climate characteristics, and high and low gauge observation density in order to quantify the performance of the correction methodology in different conditions. The results of this study demonstrate that the bias correction methodology improves QPE data, and the performance varies with respect to accumulation period, region, topographical features, and gauge observation density. The benefits of these improvements and tests can help improve QPE data for drought monitoring.

2. BACKGROUND AND HYPOTHESES

Drought is a recurring natural disaster that can devastate civilizations and ecosystems with a deficit of rainfall. The American Meteorological Society describes drought as "A period of abnormally dry weather sufficiently long enough to cause a serious hydrological imbalance" (AMS 2019). Its occurrences can bring water shortages, agricultural damages, ecological disasters, and costly economic losses. Five common categories of drought that exist are meteorological, hydrological, agricultural, socioeconomic, and ecosystem drought. Meteorological drought is a result of a reduction in precipitation over a prolonged period, hydrological drought is when surface or subsurface water supplies are reduced over a prolonged precipitation deficit, agricultural drought is a result of a dryness at a critical time in the growing season, socioeconomic drought is a result of supply and demand changes related to drought, and ecological drought is when ecosystems are harmed due to drought (Hao et al. 2017, Heim 2002). Different drought indices exist for each category in order to provide information on drought conditions. Index definitions can vary for different drought types, but precipitation deficits are a prominent feature of each. As the climate warms, droughts will become more frequent and prolonged, which increases the need for more reliable and timely warnings in order for decision makers to adequately respond and aid in recovery (Mukherjee et al. 2018, Hayes et al. 2012). In order to improve the reliability and timeliness of drought monitoring, the accuracy of precipitation measurements must be improved. The work in this project will help improve drought monitoring by

correcting Stage IV estimated precipitation datasets through a series of radar bias corrections.

Numerous drought monitoring indices are already available for quantifying severity and impact, but they are subject to different definitions of drought depending on the index. Due to the different types and broad definition of drought, it is difficult to have a universal drought index (Heim 2002). For example, the Palmer drought severity index (PDSI) and the Z index can be derived from the same data, but the PDSI may be better suited for hydrological drought while the Z index performs better for meteorological and agricultural drought (Quiring 2009). In addition to indices, assessment and evolution of drought can be monitored using land surface models (LSM). Soil moisture, evapotranspiration, precipitation, and streamflow are examples of drought related parameters that can be simulated with LSMs. In order to effectively detect and quantify drought, these parameters must be accurately depicted.

Satellites and their instruments are an invaluable observation source for aiding in drought monitoring because they can view drought impacts and measure moisture-related variables. They are able to provide near real-time monitoring, consistent records, and high spatial resolution across the globe (AghaKouchak 2015). Satellites measure a wide variety of variables including evapotranspiration, soil moisture, precipitation, snow cover, vegetation, surface temperature, evaporation, water storage, and land cover (AghaKouchak 2015). These measurements can be used as drought indicators and assimilated into models. Commonly used satellites include the Soil Moisture Active Passive (SMAP), Gravity Recovery and Climate Experiment (GRACE), and the Soil

Moisture and Ocean Salinity (SMOS). SMAP retrievals are able to provide top layer soil moisture estimates and soil-moisture-based drought indices (Mladenova et al, 2019, Liu et al. 2017). However, SMAP performance may vary with respect to different climate regimes and vegetation, which can lead to inconsistent drought monitoring (Mishra et al. 2017). GRACE is effective at monitoring terrestrial water storage and detecting drought occurrence and severity, but its spatial and temporal resolution makes it difficult to find flash-drought and local variations (Li et al. 2019, Thomas et al. 2014, Houburg 2012). Finally, SMOS can determine changes in soil moisture from heavy precipitation well, but it is prone to errors in measuring light precipitation and drying too quickly after rainfall events (Blankenship et al. 2016, Shellito et al. 2016). These examples are just a few of the many pros and cons to satellite measurements for drought monitoring. As technology progresses, there will be an abundance of new and more accurate ways to classify drought (AghaKouchak 2015).

In addition to satellite measurements, rain gauges and radar precipitation estimates are essential for drought monitoring. Rain gauges give in-situ observations while radars fill in gaps of information between the gauges. These methods are subject to numerous errors that can lead to inaccurate measurements. Rain gauges are prone to under catchment caused by wind, tipping losses, splashing, and mechanical errors (Legates 2000, Hunter 1996). Rain gauges are assumed to be “ground truth” because they directly measure rainfall, but since they are point estimates, they are not representative of spatially continuous data (Yilmaz et al. 2005). Errors in radar precipitation estimates are caused by beam blockage, range dependent and mean field

biases, improper Z-R relationships, hardware calibration, ground clutter, and bright banding (Jayakrishnan et al. 2004, Steiner et al. 1999, Ciach and Krajewski 1999). Radar estimated precipitation has high spatial resolution, but its values are derived from reflectivity-rain rate (Z-R) relationships, so estimates can vary based on the different relationships. An advantageous way to incorporate the "ground truth" of gauge observations and high spatial resolution of radars, as well as minimize each of their respective errors is through merging. Interpolation methods like kriging with external drift (Rabiei and Haberlandt 2015), kriging for uncertain data (Cecinati et al. 2017), as well as IDW (Foehn et al. 2018) can be used to combine gauges with radar data in order to increase precipitation estimate accuracy. Since drought can vary on local scales, successful drought monitoring depends on high spatial resolution measurements.

The measurement data for drought variables are important for use in data assimilation systems, LSMs, and drought indices. For example, the North American Land Data Assimilation Phase II (NLDAS II) product is used as an input forcing dataset for LSMs in CONUS (Xia et al. 2012a, Xia et al. 2012b). Its forcing data is extensive, providing over 40 years of information at hourly intervals. The LSMs associated with it are Noah, Variable Infiltration Capacity, Sacramento Soil Moisture Accounting, and Mosaic. Most of the forcing data is derived from the North American Regional Reanalysis (NARR), which includes temperature, specific humidity, wind speed, surface pressure, incoming solar and longwave radiation, and precipitation (Xia et al. 2012a, Xia et al. 2012b). Limitations with NLDAS II include its coarse resolution with grid spacing of 15km, possible quality control issues with input rain gauges, improper blending with

NARR, and real-time latency. Reliability of LSMs and drought indices rely on meteorological data, so it is crucial that the forcing data is accurate.

Modeling relies on high quality input data. One of the weaknesses of this data is the depiction of high-resolution rainfall accumulation. Commonly, Stage IV quantitative precipitation estimates (QPE) made by National Centers for Environmental Prediction (NCEP) are used in modeling for its high spatial and temporal resolution. From 2002 to the present, Stage IV QPEs produce hourly and 6-hourly mosaicked maps from the 12 River Forecast Centers (RFC) at a spatial resolution of 4km, and they are made up of merged data from 140 radars and approximately 5500 rain gauges (Lin and Mitchell 2005). The original Stage IV product was designed for assimilation into weather forecast models to improve precipitation forcing, but it is useful for other applications such as intercomparison to satellite QPE, flash flood forecasting, ecological applications, and drought monitoring (Eldardiry 2017, Nelson et al. 2016). Limitations of Stage IV data include tendencies to underestimate heavy and overestimate light rainfall (Westcott et al. 2008), have missing data and grid coverage errors leading to erroneously low values (Case et al. 2013), and contain the major radar biases including beam blockage, range dependent and mean field, and two-dimensional biases. Because Stage IV is mostly used for real-time and short-range flood analysis, its long-term biases and errors are often not strongly considered during the RFC QC phase. The biases are difficult to identify in the short term but can have severe impacts in long-term applications. The use of Stage IV in drought monitoring is challenging with these errors due to the nature of drought occurring on longer timescales.

Beam blockage is one of the sources of error in Stage IV QPEs. This error is caused by an obstruction to the radar beam, which leads to a systematic underestimation of precipitation beyond the blockage along the beam. When there is a blockage in a radar domain and a beam hits it, the beam is scattered and absorbed at the obstacle, then the returned power from targets beyond the obstacle is diminished or nonexistent, which leads to the underestimated rainfall values (Krajewski et al. 2006). Beam blockages are most often caused by complex terrain, tall buildings, trees, water towers, and cell towers, and they generally occur at the lowest elevation angles (Overeem et al. 2009, Bech et al. 2003). The lowest elevation angles of the radar beam provide the best estimates of precipitation, which makes it crucial to correct for these biases (Baeck and Smith 1998). In order to produce accurate Stage IV QPEs, beam blockages must be identified and corrected.

Mean field and range dependent biases are another common source of error for Stage IV QPEs. Mean field biases are caused by errors in radar calibration and inaccurate Z-R relationships. Radar calibration errors are a result of changes over time to the radar constant C related to components, such as the transmitter, receiver, and antenna system (Villarini and Krajewski 2010b). Errors caused by Z-R relationships are associated with mischaracterized drop size distributions within storms (Anagnostou et al. 1998). Range dependent biases can be attributed to the vertical profile of reflectivity (VPR) effect, which is related to changes in precipitation intensity with height (Krajewski et al. 2011). As the beam height increases at longer ranges from the radar, hydrometeors with different properties can be sampled, which leads to different

estimates in surface precipitation. Range dependent errors can also be caused by attenuation, partial beam filling, and overshooting of low clouds at longer ranges (Villarini and Krajewski 2010a). Corrections to both mean field and range dependent errors can be done through comparison of the radar field and rain gauges.

Discontinuities at the boundaries of each radar domain can be an unintended consequence of the beam blockage and mean field and range dependent bias adjustments. These discontinuities occur because each radar domain is independently analyzed leading up to this point. For example, if one radar domain is adjusted downwards and the neighboring one is adjusted upwards away from each other, the grid cells that border each other will have a larger difference in precipitation estimates after the adjustments are performed. Visually, this creates a discontinuity that clearly shows a line between the two radar domains. In order to correct for this, the magnitude of the discontinuity is calculated, then the grid cells are either increased or decreased at an inverse distance depending on which side is higher or lower than the discontinuity. This correction is performed in this study.

Two-dimensional biases are another source of error within Stage IV QPEs. These errors occur because precipitation microphysics vary in space, where the Z-R relationship may perform better in certain situations than others. The errors are small in shorter accumulation periods but can be pronounced in longer ones. Additionally, independent analyses for each RFC and radar domain can further exacerbate two-dimensional biases. In order to correct for the biases, the entire precipitation field is

considered rather than individual radar domains, and reliable gauge data is incorporated through the use of interpolation.

In order to correct for biases in Stage IV QPEs, a three-step correction methodology was developed by McRoberts and Nielsen-Gammon (McRoberts and Nielsen-Gammon 2017, McRoberts 2014). This methodology first detects and corrects for beam blockage that leads to an underestimation in precipitation amount. The detection is done by separating each radar domain into multiple annuli, fitting a low-pass Fourier series to long term QPEs in each annulus, and flagging cells that are largely different from their neighboring cells. If there is a dense area of flagged cells beyond a distance in a radar radial, the beam is considered to be blocked. This is corrected using radar estimated accumulations of unblocked nearby cells. The next step corrects for mean field and range dependent biases. This is done by calculating the observed bias between radar and rain gauges, then modeling these biases as a function of range for each radar. Cells are grouped according to distance from the radar, and a composite bias is calculated. A single model is made from the resulting calculations and takes the form of a simple linear fit. Then, the grid cells are adjusted accordingly. Finally, the methodology adjusts for residual two-dimensional biases that were not already accounted for in the other steps. The biases computed in the previous step are interpolated to the entire precipitation field using Ordinary Kriging, then the data are adjusted based on the interpolation. The results of this correction methodology give a new QPE in near real time that is more accurate and representative of the precipitation field. This can help improve drought monitoring because the datasets are high resolution

both spatially and temporally, and areas with sparse in-situ measurements are better estimated.

For this study, adjustments are made to the three-step correction methodology in order to fix beam blockage identification and adjustment errors, remove radar domain discontinuities, and increase the efficiency of the mean field and range dependent and two-dimensional bias scripts. In the beam blockage identification algorithm, there are issues with the grid point flagging. Figure 2.1 shows two different Fourier series fits used for the flagging procedure within the KABR radar domain for the 36-month accumulation period ending December 2012. Annuli 3 and 13 represent the HRAP grid points and their percent of normal (PoN) values for all azimuth angles in the 30-39km and 130-139km annuli respectively. The Fourier fit for annulus 3 is a better approximation than at annulus 13. One reason for the difference in fits is that annulus 13

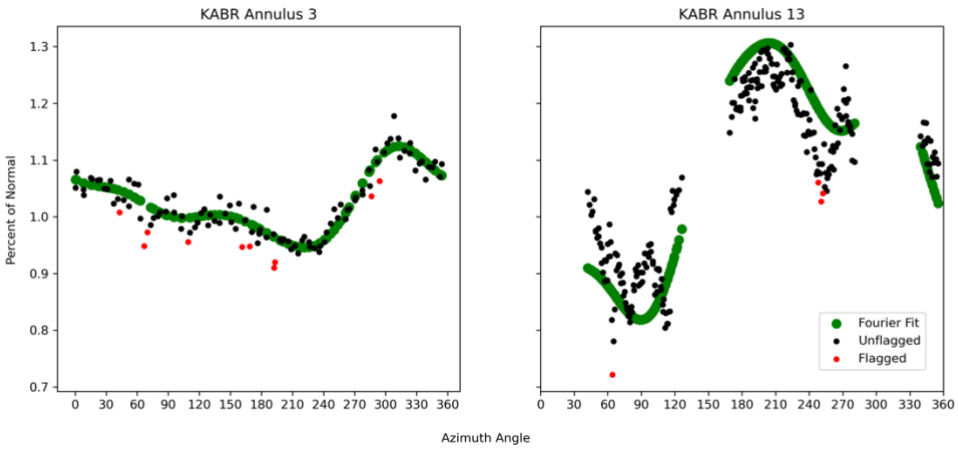

Figure 2.1: Comparison between short-range annulus (left) and long-range annulus (right) and their Fourier fits and flagging. Improper flagging can occur due to less availability of grid points at longer ranges.

has gaps of available grid points in certain azimuth angles, such as the 0-30-, 130-170-, and 280–330-degree gaps. Certain azimuth angles may not contain grid points at further annuli because the radar domains are not circular. At longer ranges, this can cause certain azimuth angles to not have grid points within them. The wide gaps in data create a discontinuity that the Fourier series may improperly estimate. In a poor fit like in annulus 13, grid cells near the data gaps may be flagged for appearing anomalously low compared to the fit. This can lead to unblocked beams being flagged as blocked, and vice versa. Figure 2.2 shows the effects of improper flagging. The red circle in the left PoN image is the location of a beam blockage in the KABR domain. However, the yellow points in the right image are flagged grid points. If incorrect beams are flagged, the resulting corrections can increase the estimates at unblocked beams by a difference approximately equal to the distance from the Fourier fit to the erroneously flagged grid points. The adjustment to this step in the correction methodology addresses this by using

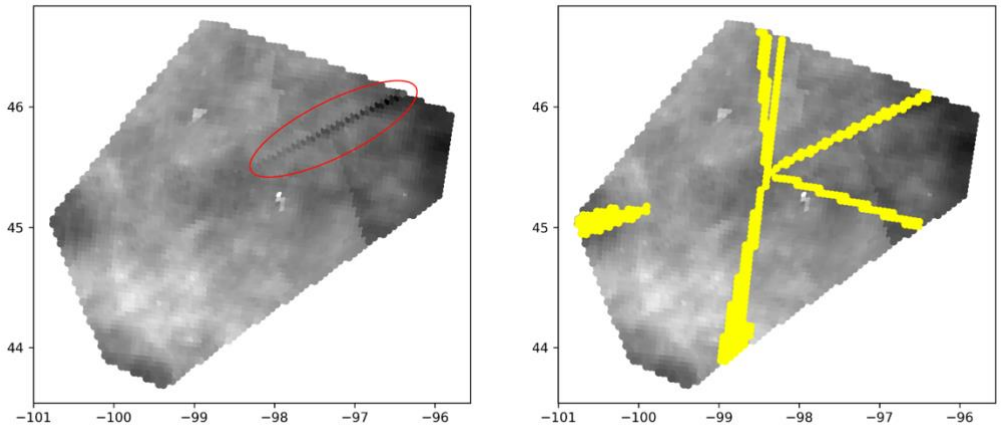

Figure 2.2: Percent of Normal for KABR 36-month accumulation ending December 2012. The circled area in red is a blocked beam, but the flagged beams are in yellow.

image filtering to find radar beam-like features, then applying an adjustment. In addition to increasing the efficiency of the mean field and range dependent step and the two-dimensional step, a radar domain discontinuity adjustment is executed in between the two steps. The discontinuity corrections increase or decrease PoN values along each radar domain boundary in order to remove discontinuities that may exist. Finally, the two-dimensional step is tested using inverse distance weighting (IDW) and inverse distance weighting squared (IDW2) along with Ordinary Kriging (OK) to determine the best performing interpolation method to use in its adjustment. The final radar estimated precipitation analysis will ideally agree closely with the gauge estimates. This is an appropriate goal because gauge networks are made up of independent observations and thus, should be free of spatially dependent biases, which can help reduce the biases associated with Stage IV QPEs (McRoberts 2014). All of the steps and their changes will be discussed further in the methods section.

Most prior studies have focused on improving QPEs for the purpose of short-term events, like flooding and weather forecasting. Little work has been done using the high spatial and temporal resolution benefits of Stage IV QPE for improving drought monitoring. The work in this study will evaluate how well each step in the three-step correction methodology by McRoberts and Nielsen-Gammon performs relative to raw Stage IV data. Additionally, evaluation of multiple interpolation methods will be performed for different QPE accumulation periods using the corrected data to determine the optimal interpolation method for the two-dimensional adjustment. The benefits of this analysis will be to better understand the performance of the bias correction methods,

and to quantitatively evaluate its performance on different time scales and interpolation methods. The applications of this will be useful for assimilation into drought models and monitors so that the input precipitation information is more accurate. The existing literature on Stage IV QPEs and drought monitoring motivates the testing of the following hypotheses:

1. QPE data will be improved with extensively tested interpolation methods.
2. Overall, Ordinary Kriging will perform better than Inverse Distance Weighting and Inverse Distance Weighting Squared.
3. Regions with fewer gauges and greater topographic relief will benefit more from the corrections.
4. Bias adjustments using gauge-based adjustments will perform better than those only using radar.

3. DATA AND METHODS

3.1. Radar Data

The NWS Stage IV QPE data are analyzed for the period of 2005 through 2020 over the region east of -103°W in CONUS. This time period is selected because it is the full length of temporally complete data in which all steps of the bias correction can be

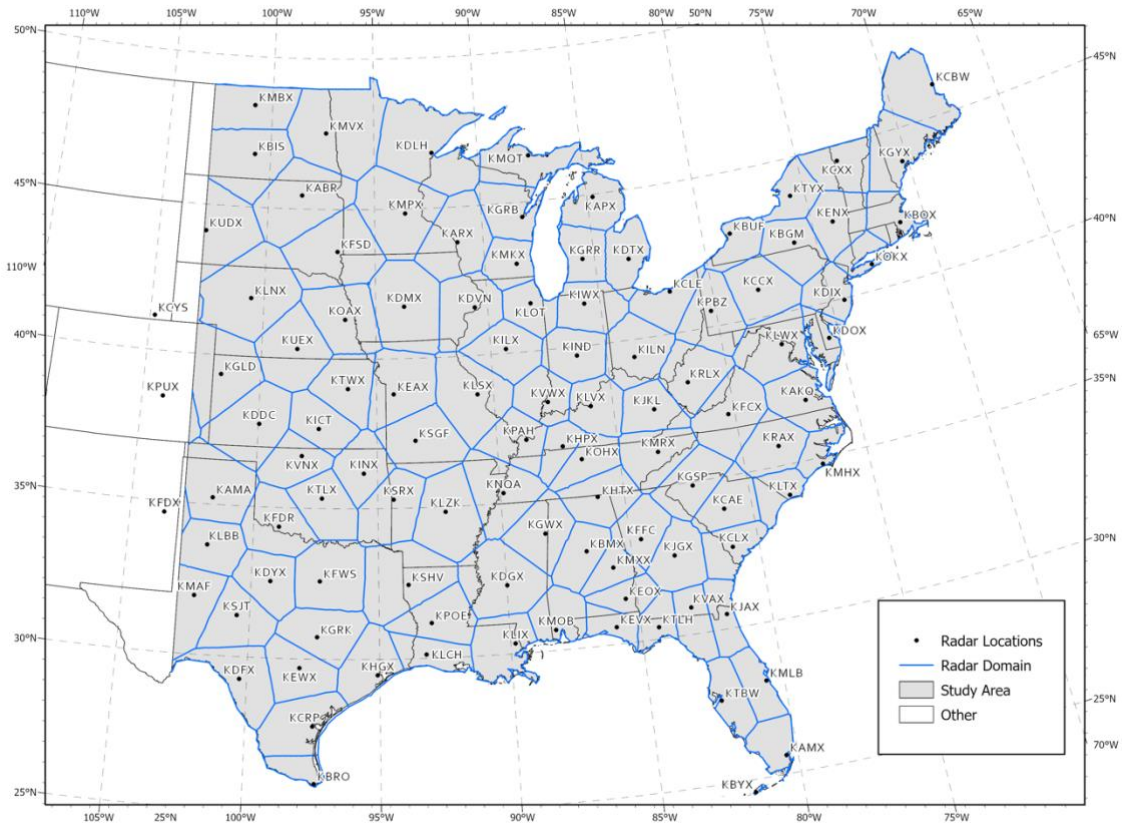

Figure 3.1: The study area (gray) and radar domain boundaries (blue) for this study.

applied. The western extent of the study region avoids the Rocky Mountains because the complex topography causes numerous cases of beam blockage in the low elevation

angles, and many stratiform rain or snow events could go undetected (Maddox et al. 2002). Figure 3.1 shows the location and domains of all radars used in this study. All radar domains (blue lines) within the gray shaded region are adjusted and tested in this study. Because the KCYS, KPUX, and KFDX domains do not completely fit within the study area, the bias adjustment calculations are performed for their entire domains, and only the grid points that are east of -103°W are used. Additionally, the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) from 1981 to 2010 climate normals are aggregated to 1 to 36-month totals as a comparison to the Stage IV QPE. Both the Stage IV and PRISM data are projected on the Hydrologic Rainfall Analysis Project (HRAP) coordinate system, which is a polar stereographic projection with an approximately 4-km resolution (Fulton et al. 1998). The radar estimated PoN is from the Stage IV QPE divided by the PRISM normal value at the corresponding HRAP grid point.

3.2. Rain Gauge Data

Rain gauges used in this study are selected based on reliability and data availability. Figure 3.2 shows the locations of every gauge that is used in this study. The gauges that are used are from the Weather-Bureau-Army-Navy (WBAN), Cooperative Observer Network (COOP), Federal Aviation Administration (FAA), World Meteorological Organization (WMO), International Civil Aviation Organization (ICAO), Global Historical Climatology Network (GHCN), and National Weather Service (NWS) networks. These networks are assumed to be reliable because they are frequently monitored and are quality controlled. The daily rainfall values for the gauges

are aggregated to create the same accumulation periods as the radar estimates of 1 to 36-months. In order to compare the gauge accumulations to the radar estimates, each gauge is assigned to the HRAP grid ID for which they are located. Using this information, radar-gauge bias and gauge PoN are calculated for use in the bias adjustment methodology. To ensure that the bias adjustments can be compared accurately, extensive

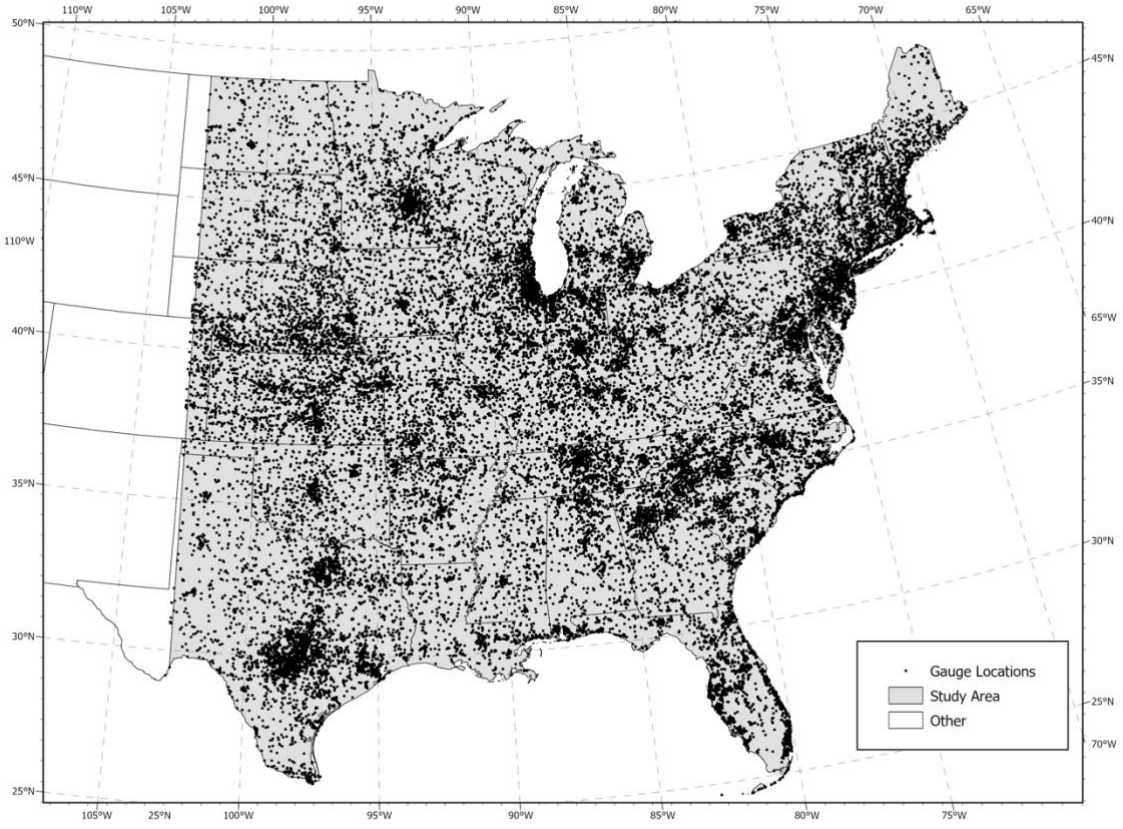

Figure 3.2: The locations of every gauge used in this study.

quality control is performed in order to remove spurious values. The gauge quality control method uses information from both the gauges and Stage IV precipitation. Two equations are developed to determine the minimum and maximum allowable gauge PoN.

For a given gauge, the PoN is calculated and compared to the radar PoN at its respective grid point. Equation 1 calculates the minimum and equation 2 calculates the maximum threshold PoN for the gauge. The values q, a, r, and b are constants that are determined

$$\min PoN = q * (radar\ PoN - a) \qquad (1)$$

$$\max PoN = r * (radar\ PoN + b) \qquad (2)$$

from testing. If the gauge PoN falls outside of this range, it is considered to be an outlier and dropped. These equations were developed by considering how the gauge and radar PoN should be approximately related. Since the gauges are the "ground truth" in this study, the equations account for variations above and below the radar PoN. However, the radar PoN values are used because it is reasonable to assume that the expected gauge value should be close to the radar value. The results and values of the constants are presented in section 3.4.

3.3. Bias Adjustment Procedure

The bias adjustment procedure is adjusted from the original methods presented by McRoberts and Nielsen-Gammon (McRoberts and Nielsen-Gammon 2017, McRoberts 2014). The necessary changes include an improved beam blockage detection algorithm, more efficient mean field and range dependent and two-dimensional adjustments, and a newly added radar domain discontinuity adjustment. The procedure starts with use of Python scripts to download the radar, rain gauge, and HRAP grid metadata. The data are then aggregated to monthly values because drought monitoring requires longer term data. Additionally, the gauge data is assigned the HRAP grid ID that it is located within. With the grid information, the gauge and radar values are paired

so that they can easily be compared. Once the sorting is complete, the beam blockage correction script is run. The algorithm detects, then corrects for beam blockage for all accumulation periods within every radar domain in the study area. Then, the beam blockage corrected data is run through the mean field and range dependent bias script. The mean field and range dependent biases are calculated on a seasonal basis, then adjusted for all accumulation periods. With this output, the radar domain discontinuity correction script is run. The goal of this step is to either increase or decrease the PoN values along radar domain boundaries in order to remove discontinuities that may exist at those locations. Finally, the beam blockage, mean field and range dependent, and radar domain discontinuity corrected QPEs are run through the two-dimensional bias script to account for any remaining biases in the data. This adjustment relies on optimally interpolated QC gauge data. Additionally, the two-dimensional step considers the precipitation field as a whole rather than individual radar domains. The results of the bias adjustment procedure create Stage IV datasets for 1-to-36-month accumulation periods. The output datasets for each step and time period are evaluated in this study.

Step 1, the beam blockage correction methodology, first inputs the raw Stage IV 36-month radar estimated accumulations at a given radar domain. The 36-month period is selected in order to minimize Type I and Type II errors (McRoberts 2014). Type I errors occur when low values of precipitation due to random variability in shorter time periods get mistaken with beam blockage. Type II errors occur when spatial variations versus precipitation amount is too large. A period of 36-months is optimal because it is long enough to overcome random variability in short time scales, but short enough to

account for possible changes with time. With this data, the grid points are sorted by ascending range from the radar and the azimuth angle is calculated for each point. From there, the grid point PoN is calculated, then converted to an image using scikit-image processing in Python. The goal of the image processing is to use ridge detection to find beam blockages. Ridge detection is able to select curves that are local maxima relative to neighboring pixels in an image (Sato et al. 1998). Multiple filters exist to detect different types of ridges, and for this study, Sato filtering is selected. The Sato filter has the ability to detect continuous, curvilinear ridges then calculate the fraction of the whole image that contains such features. Additionally, different smoothing factors ranging from 0.1 to 10 can be applied to the ridge features. Through testing, the smoothing factor for this study is 0.5. The results of the testing are shown in section 4.1.

In order to attempt to differentiate between potential beam blockage and noise in the filtered image, a series of tests are run. First, the algorithm searches in 0.1° wide radials around the entire domain. The radial width of 0.1° is selected in order to easily detect subtle changes in linearity within the pixels of the filtered image. Then, the image pixel values are checked. If there are non-zero values within a radial, this means a ridge is detected. If a ridge is detected, two additional tests are run in order to determine if the ridge is due to beam blockage, and if it is, where the blockage begins. The first test determines if the ridge is due to beam blockage by calculating the correlation between the range of the grid cells and their filtered value. If a detected ridge is due to beam blockage, the filtered value will increase with range. If the correlation is greater than 0.50, the entire beam is flagged for beam blockage. The minimum correlation of 0.50

can easily determine most beam blockage occurrences; however, it can also pick up noise if the radial has multiple grid cells with increasing values at increasing ranges. If this test determines that beam blockage exists in the radial, then the second test is conducted. This test determines the location where the blockage begins. This is accomplished with step detection using convolution. In a radial with beam blockage, there will be a sharp increase in the Sato filtered pixel values as a function of range where the blockage starts. The convolution method can detect these sharp increases, which are called peaks. If a peak is detected at a range greater than 10km from the radar, the range of the peak is the starting point for the blockage, and all cells beyond that point are flagged. If no peak is detected, the entire radial is flagged. Because the values of the pixels of the detected ridges are calculated as a fraction of the whole image, radar domains with several beam blockages and additional noise can make it difficult to

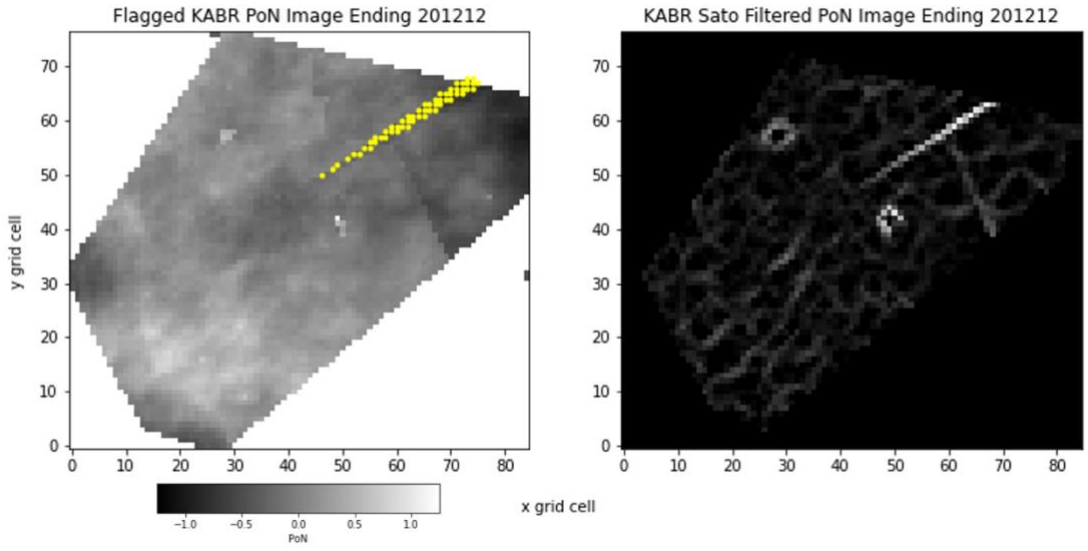

Figure 3.3: PoN image ending December 2012 with beam blockage flagging detected by Sato image filtering (left) and Sato ridge detection filtering of the PoN image (right).

identify a peak because of the pervasiveness of the non-zero pixels and their small values.

Figure 3.3 shows an example of the beam blockage identification. The selected domain is KABR, and the accumulation period is from the 36-months ending in December 2012. The right image is the 36-month PoN Sato filtering output, and the left image is the resulting beam blockage flagging. The bright, white linear feature towards the northeast is a blocked beam. The other white colors in the image are noise. The resulting adjustments from the flagging in this domain only adjust the definite beam blockage. Compared to the original beam blockage adjustment methodology, this method adjusts the beams differently, which can result in less error if improper beams are flagged. Instead of determining the magnitude of adjustment based on a Fourier fit, the adjustment in this methodology averages neighboring grid points. For each flagged grid cell, the adjusted value is equal to the mean of the 6 closest unflagged grid cells. It is fair to assume that a blocked grid cell will have similar estimated precipitation to its neighbors because of the long accumulation period. The original algorithm's erroneous adjustments lead to beams becoming much higher relative to their neighbors, while this method smooths the data in cases where too many grid cells are flagged. Fortunately, radials that are adjusted when beam blockage does not exist and vice versa can still be corrected in the other steps of the bias adjustment algorithm. Finally, once the 36-month adjustment for all of the flagged grid cells is determined, the percent increase between the unadjusted and adjusted beam blockage values are calculated, then applied to all accumulation periods leading up to the 36-months. It is assumed that the magnitude of

the beam blockage is approximately the same in the preceding 1-to-35-month accumulation periods because the obstruction is likely to be the same.

Step 2 is the mean field and range dependent adjustment algorithm. The mean field and range dependent biases are calculated using the output from the beam blockage adjustment. This methodology uses information from both the radar and rain gauge accumulations. For a given radar domain, four different adjustment factors are calculated for every HRAP grid point based on seasons where December, January, and February are winter, March, April, and May are spring, June, July, and August are summer, and September, October, and November are fall. Seasonal calculations are performed because the mean field and range dependent biases can vary by seasons depending on the location due to climatological features. For example, in the high plains, deep, convective cells are more likely in the spring and summertime, but shallow, stratiform rainfall or snow is more likely in the fall and winter. In deep, convective cells, large biases may exist due to extreme rain rates and hail contamination (Villarini and Krajewski, 2009). Wintertime biases are often caused by bright banding and overshooting of the beam. Additionally, differences in drop size distributions between the seasons are challenging to account for in Z-R relationships (Villarini and Krajewski 2010b). In order to calculate the adjustment factors at each grid point, equation 3 is developed using information from both the radar estimated and gauge observed precipitation, and it is calculated separately for each season. Equation 3 is as follows:

$$p_{new}(r) = p(r) * \left(\frac{psn(r)}{ps(r)}\right) * \ pg \quad (3)$$

The value, $p(r)$, is the average radar estimated precipitation as a function of range. This is calculated by averaging every radar estimate for all occurrences of a given season in the study period at each possible grid point range from the radar within a domain. The ratio in the middle is $psn(r)$, the smoothed average PRISM normal precipitation as a function of range, and $ps(r)$, the smoothed average radar precipitation as a function of range. The value, $psn(r)$, is a smoothed curve of $pn(r)$, the average PRISM normal precipitation as a function of range. The calculation of $pn(r)$ is similar to $p(r)$, except the normals are used instead of the estimates. The value, $ps(r)$, is a smoothed curve of $p(r)$. The term, pg, is the average gauge PoN, which is calculated by taking the average of all gauge PoN observed for the entire study period in the given radar domain. The smoothing is performed using a Savitzky-Golay filter, where the window length is 10% of the number of grid points in a given radar domain, and the order of the polynomial is three. The result, $p_{new}(r)$, is the new average radar estimated precipitation as a function of range for the entire study period. From $p_{new}(r)$, the adjustment factor for each range can be calculated as shown in equation 4. The result from this calculation gives factors at each range that can multiplied by the grid cell values that are located at the same ranges. For example, if equation 3 is calculated for winter, then the factors calculated in

$$adjustment\ factor = 1 - \frac{(p(r) - p_{new}(r))}{p(r)} \quad (4)$$

equation 4 are applied to all grid point radar estimated values with the corresponding range that fall within the winter months. Once all of the seasonal factors are calculated and applied, the seasonal accumulation periods can be summed in order to compute longer accumulation periods.

Step 3 is the radar domain discontinuity adjustment. The radar domain discontinuity step relies upon the beam blockage and mean field and range dependent

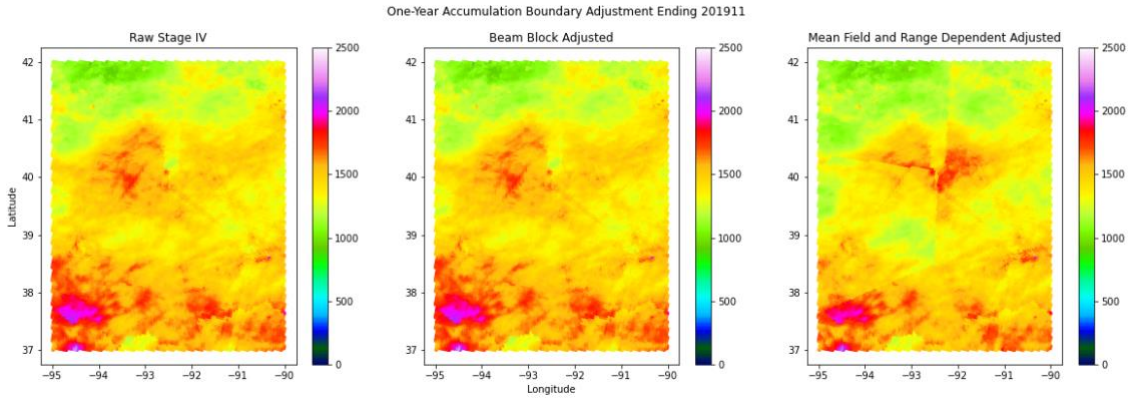

Figure 3.4: Raw (left), beam blockage adjusted (middle), and mean field and range dependent adjusted (right) Stage IV data for a central US location for the one-year accumulation period (mm) ending December 2019.

corrections. Figure 3.4 illustrates the radar domain discontinuity problem. The figure shows three different radar estimated precipitation images where the left is the raw Stage IV, the middle is the beam blockage adjusted, and the right is the mean field and range dependent corrected data for the one-year accumulation period ending December 2019. It can be seen in the figure that distinct lines appear at -92.5°W in the north-south and approximately 40°N in the west-east directions in the mean field and range dependent corrected data. These are an example of discontinuities at the edges of multiple radar domains. One cause of this is from anisotropies in the raw Stage IV data (McRoberts 2014). Each River Forecast Center (RFC) produces independent daily analyses of the Stage IV data. Differences that exist in the analyses between each RFC can lead to biases. In the short-term, the biases are minimal but can compound as the accumulation

periods increase. If an individual radar is biased low and a neighboring radar is biased high, the compounding biases will lead to a low estimate directly next to a high estimate, which creates a discontinuity. Another cause is from the adjustments leading up to the discontinuities. Both the beam blockage and mean field and range dependent adjustments are done independently for each radar domain in one-dimension. This leads to residual two-dimensional biases, and the discontinuities are the most prominent example of this type of bias. These discontinuities are undesirable because it implies sharp differences of precipitation at the edges of different radar domains, which is inaccurate at longer accumulation periods. It is desirable to remove these discontinuities because it can help increase the accuracy of the precipitation field. In order to smooth the discontinuities, average PoN along the boundaries are calculated and adjusted using an IDW method. Equations 5 and 6 show how the IDW method is calculated for the case where one side of a discontinuity is lower than the other, and one is higher than the other respectively. Both equations are calculated for individual grid points from 0km to 25km

$$P = \left(PoN - \left(\frac{Discont/_2}{D/_{D_0}} \right) \right) * norml \quad (5)$$

$$P = \left(PoN + \left(\frac{Discont/_2}{D/_{D_0}} \right) \right) * norml \quad (6)$$

away from the discontinuity. The variables are the same in each equation, where *P* is the new radar estimated precipitation value, *PoN* is the percent of normal of the grid point, *D* is the distance the grid point is from the discontinuity, D_0 is a reference distance, and *norml* is the normal precipitation value for the grid point. The purpose of the reference

distance is to make the distances dimensionless, and for this study, the value is 1km. For example, if the average PoN along boundary 1 is 1.10 and 0.90 along boundary 2, there is a discontinuity of 0.2. The equations increase the PoN on the lower side and decreases the higher side for grid points up to 25km from the discontinuity. The 25km radius from each boundary was arbitrarily selected as an appropriate distance because the IDW calculations cause infinitesimal changes beyond this range. By using the blending method, the discontinuities at radar domain boundaries can be removed and the IDW method prevents new discontinuities from forming. This method helps improve the precipitation estimates because it reduces the high and increases the low biases which minimizes each of their respective errors.

Step 4 adjusts for the remaining biases in the corrected data by using the two-dimensional adjustment methodology. The input for this step is the beam blockage, mean field and range dependent, and boundary discontinuity corrected data and QC gauges. The biases that remain could be from compounding errors and those as a result of considering each domain independently. Unlike the prior adjustments, the two-dimensional adjustment is done for the entire precipitation field rather than independently at each radar. This adjustment first calculates the bias between all gauges and their respective grid cell estimate. Then, this bias is interpolated to the entire study area using an optimal interpolation method. The method with the lowest error and least amount of bias for all accumulation periods is tested between OK, IDW, and IDW2. Additionally, the number of known points to consider in the interpolation calculations is tested for an optimal value. Finally, the interpolated field is added to the input radar

estimated field to get the adjusted values. This procedure is run for every tested accumulation period in this study.

3.4. Data Analysis Methods

Using the boundary corrected mean field and range dependent adjustment data, OK, IDW, IDW2 are tested in the two-dimensional adjustment to determine which method produces the lowest error. All of these methods calculate weights from surrounding known values in order to interpolate, but their approaches differ. OK calculates variograms using spatial autocorrelation at different distances from the point in question. This method is beneficial because it automatically accounts for clustering and can take directional bias into account, but its complexity and computational expense can be a drawback (Gentile et al. 2013). In this study, OK with a spherical, exponential, gaussian, power, hole-effect, and linear model are tested. For IDW and IDW2, weights are only based upon the distances from the point in question to the known points, where IDW2 places a higher importance on nearby points. These methods are simple and straightforward, but they are not smoothed (Gentile et al. 2013). One of the main differences between OK, and IDW and IDW2 is the way that the known values are handled. With IDW and IDW2, the analyses are required to exactly agree with the known points at their locations. For OK, it assumes that the known points have random error, so exact agreement at the locations of the known points is not required.

This study analyzes the accuracy of each interpolation method on different time scales because performance can vary temporally. Rainfall accumulation periods are tested using seasonal, one-year, and three-year accumulation periods where winter,

spring, summer, and fall are the seasons, December through November are the years, and the aggregates from December 2005 through November 2008 and subsequent 36-month groupings are the three-year periods. Long accumulation periods are the main focus of this study due to the goal of improving Stage IV estimates for drought monitoring. The interpolation performance and biases across different seasons vary due to climatological differences in each season such as precipitation type and intensity, and the structure of convection and stratiform events. Additionally, short-term errors can compound as the accumulation period increases to one and three-year periods. As part of the assessments, this study determines the reliability of each interpolation method in the different time periods, as well as select the top performers for each time period. It is desirable to use the same method for all cases for simplicity.

Extensive data withholding tests are run in order to help determine the performance of the interpolation methods and the accuracy of other tests in this study like gauge QC parameters and determination of the western extent of the study area. This method first works by randomly withholding a certain amount of known data points. In this study, the known data points are gauge observations and 20% are withheld. Then, statistics are calculated between the withheld points and base analysis points in question. For example, in testing the performance of the two-dimensional bias adjustment, the base analysis would be the boundary blended mean field and range dependent adjusted data. The root mean square error (RMSE), median absolute error (MAE), and median bias (MB) are the calculated statistics for this study. The median based statistics are useful for their resistance to outliers. Then, the non-withheld points are run through the

analysis that is being tested. Using the two-dimensional bias correction example, the bias between the gauge and radar estimates at the non-withheld points would be interpolated to the HRAP grid. From the output, the RMSE, MAE, and MB are calculated at the locations of the withheld points between the observed and newly estimated values. The performance is determined by the difference in the pre-analysis and post-analysis statistics. The goal is to find the analysis with the greatest reduction in RMSE and MAE, and the value closest to 0 for the MB for each test. For each test, the data withholding is run with two different randomly withheld points in order to determine if the results are consistent using different gauges. Finally, for non-gauge-based tests and adjustments, no gauges are withheld in testing the error statistics because gauges are not used to calculate the analysis.

The gauge QC parameters are tested using data withholding, where IDW with 30 neighboring points in each grid point calculation is the interpolation method. This method is selected because it has a small computational expense. The 30 neighboring points are used because it is enough points to sample in all directions from an unknown point, but not enough points to sample at distances too far away. Additionally, the input radar data is the beam blockage corrected data, which is used because it is the step before gauge data is necessary for adjustment. Because the gauges involved in the withholding are not QC'd, the data withholding process is modified. The modification tests the interpolation values with all of the withheld points, then again with the QC applied using the values being tested. This modification can help determine if the performance is consistent between different gauges. The procedure for testing the four

constants in equation 1 is as follows: *r* and *b* are held constant for the duration of the test using 1.2 and 0.4 respectively. These values are selected because the resulting gauges appeared the most reasonable to the eye. Then, *q* is tested using multiple values while *a* is held constant using a value of 0.1. Once this run is complete, the optimal value of q is selected, then *a* is tested while holding *q* constant. Finally, the optimal value of *a* is selected, then the test is rerun for both constants to check if the optimal values remain the same. The testing for equation 2 is performed the same way, except *q* and *a* are held constant for the duration using their optimal values and *r* and *b* are the constants in question.

The results for the gauge QC equation constant testing are shown in figure 3.5. The solid lines indicate that all withheld points are used (non-QC) in the RMSE, MAE, and MB calculations, and the dashed lines indicate that the QC is applied to the withheld points. The optimal values are the ones that minimize the RMSE and MAE and have the MB closest to zero. For the optimal *q* value, 0.7 is selected. The non-QC RMSE starts to increase, the non-QC MAE flattens then increases, and both MB lines increase above this value. There are small improvements in QC RMSE and MAE above 0.7, but they are not large enough to justify the increasing error in the other metrics. Using this optimal *q* value, the optimal *a* value is tested and determined to be 0.2. The non-QC RMSE remains approximately the same beyond this point. Additionally, both MAE lines increase and both MB become more negative beyond this point. Finally, for the equation 2 variables, *r* and *b*, there is very little change in error and bias throughout all of the tested values. This suggests that equation 1 in the QC plays a bigger role in reducing the

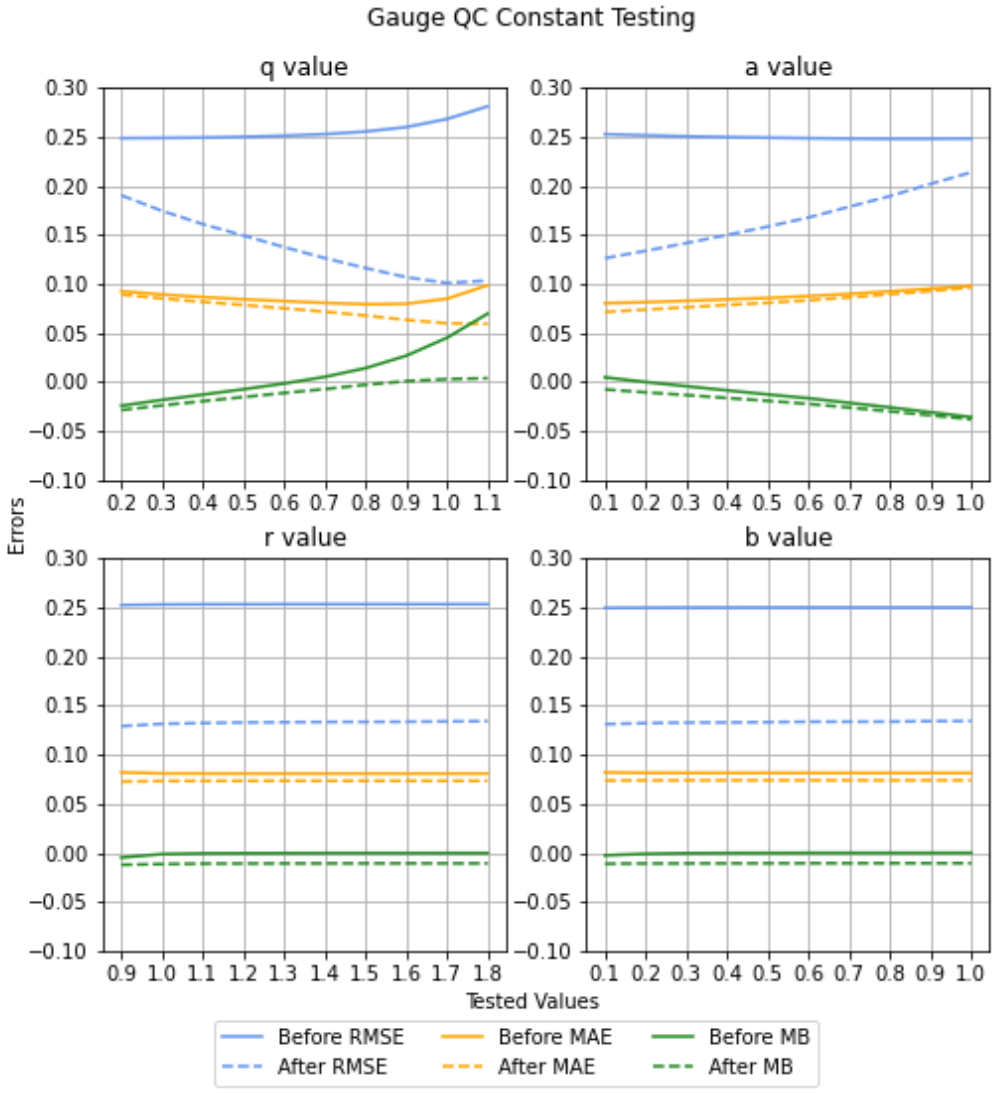

Figure 3.5: The RMSE (blue), MAE (orange), and MB (green) of constants *q* (upper left), *a* (upper right), *r* (lower left), and *b* (lower right). Solid lines are before, and dashed lines are after the QC is applied to the withheld gauges.

outliers in the gauges. This is an expected result because gauges are more likely to underestimate relative to the radar than overestimate. The values of *r* and *b* are selected to be 1.2 and 0.2 respectively. These are selected because of the small dip in MAE leading up to the values. Figure 3.6 shows an example of how the gauge QC works. The

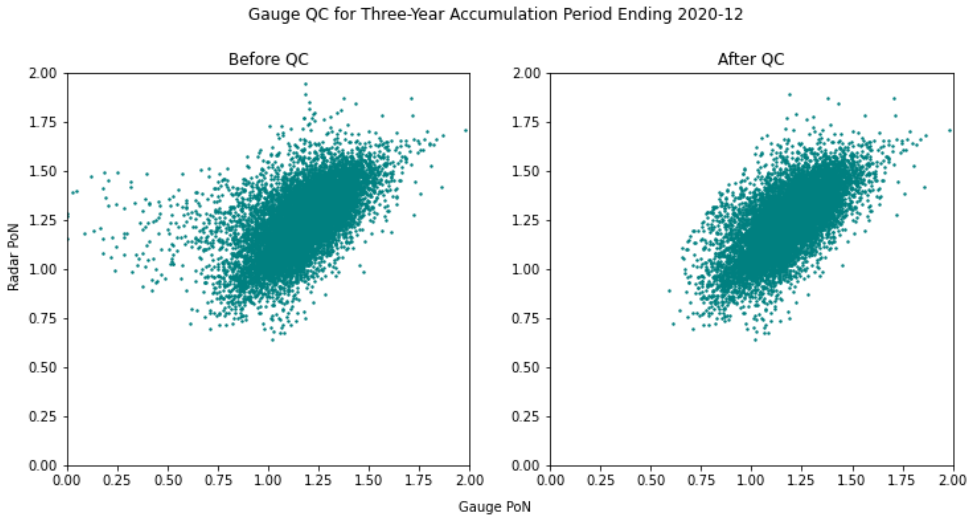

Figure 3.6: Gauge PoN versus radar PoN for the three-year accumulation period ending December 2020 before (left) and after (right) the gauge QC is applied.

scatterplots show the gauge PoN versus the radar PoN for the three-year accumulation period ending in December 2020, where the left image is before and the right is after the QC is applied. It can be seen that most of the removed points are those where the gauge PoN values are lower than the minimum acceptable PoN. Additionally, when applying the gauge QC, approximately 8% of the gauges are removed. The QC gauges are used in all steps that require gauge data and data withholding.

The study area extent testing is performed in order to determine how far west the study area can be without drastically reducing the radar estimate accuracy from the effects of the Rocky Mountains. The raw Stage IV data is the input analysis, and QC gauges are used for comparison with the analysis. Because gauges are not used in the input analysis, the raw Stage IV and gauges can be directly compared without withholding. The results from the comparison are shown in figure 3.7, where the average

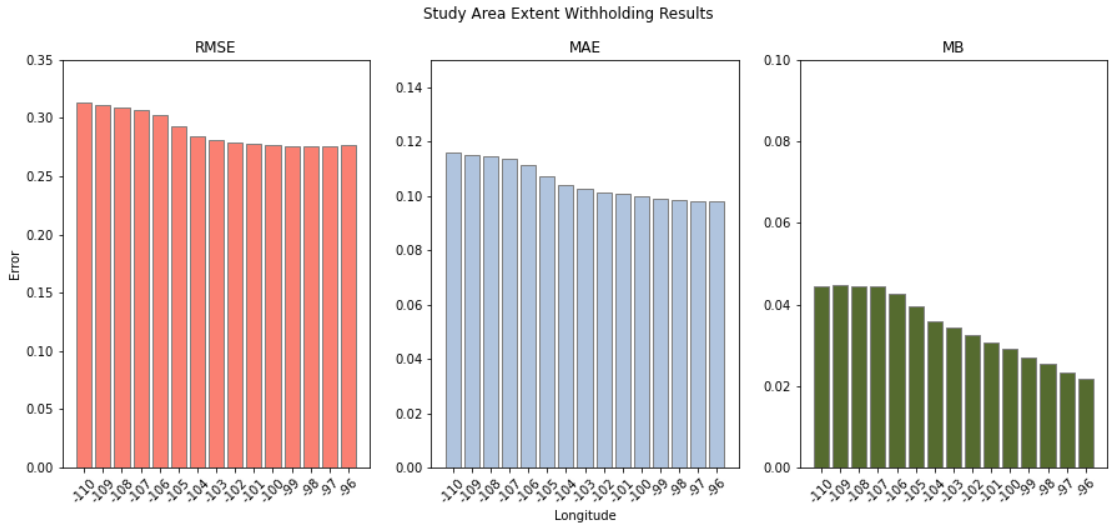

Figure 3.7: The average RMSE (left), MAE (middle), and MB (right) for each tested western longitude extent.

RMSE, MAE, and MB for each tested longitude are shown from left to right. For the RMSE and MAE, the errors decrease quickly between -110°W and -105°W, then flatten out. The MB remains fairly constant between -110°W and -107°W, then gradually decreases. The goal for this study is to use the largest possible study area while maintaining a low error. Because of this, -103°W is selected as the western extent. This extent does not have the lowest RMSE, MAE, or MB, but the tradeoff of selecting a smaller study area for a minimal improvement in error is not justifiable. Each step in the bias correction methodology is performed within the bounds of this study area.

The metrics for determining accuracy are applied across four different study regions. The overall study region is shown in Figure 3.1, which includes all of the CONUS east of 103°W. The western CONUS is not included because the topography causes more reliance on digital elevation models and rain gauges rather than radars. This

study also assesses the accuracy of the QPEs across different subregions. Figures 3.8, 3.9, and 3.10 show different subregions based on similar spatial characteristics. The subregions in figure 3.8 are selected for similarities in climatological characteristics. For example, errors in radar estimates and gauges may be higher in the northern plains (green) due to frozen precipitation and high wind. Figure 3.9 shows selected subregions based on the sum of the effective observation density per 10 square kilometers. The effective observation ratio is calculated by dividing the amount of gauge observations by

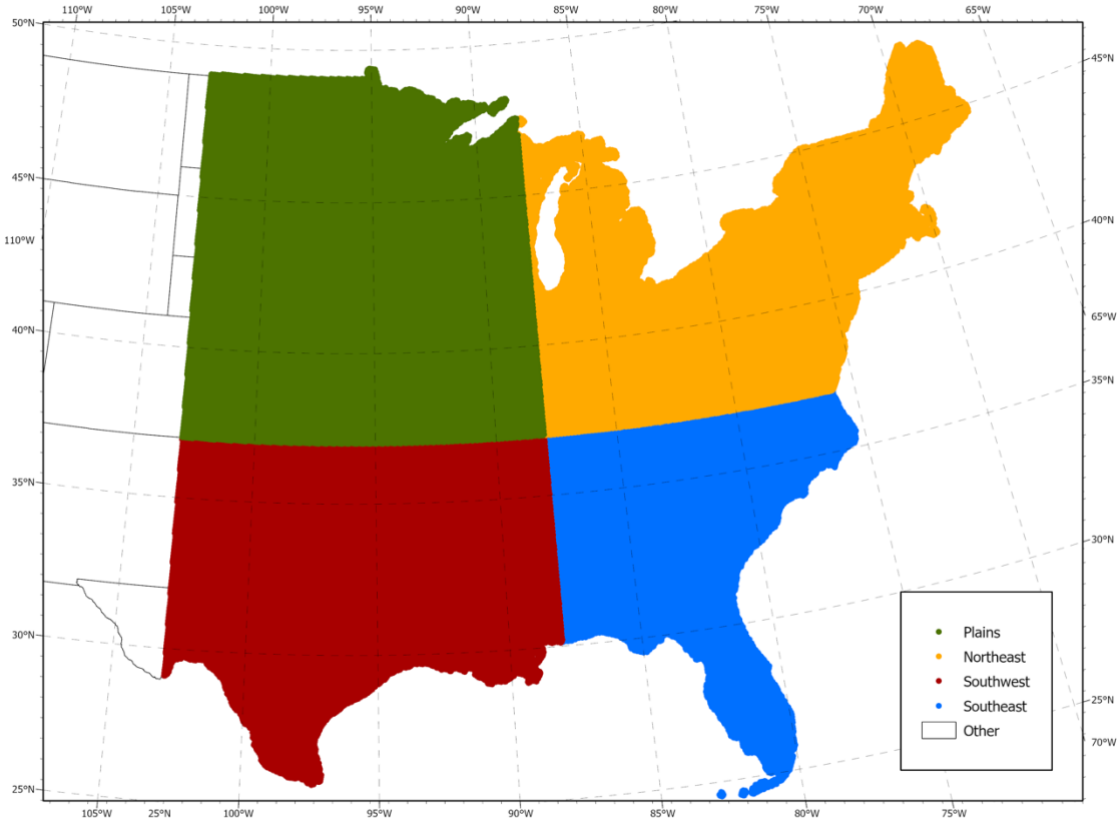

Figure 3.9: Subregions selected based on similar climatological characteristics. These regions are the Plains (green), Northeast (orange), Southwest (Red), and Southeast (blue).

the total possible number of observations in the study period for 36-month periods. The periods used are for the 36-month totals ending December 2008, 2011, 2014, 2017, and

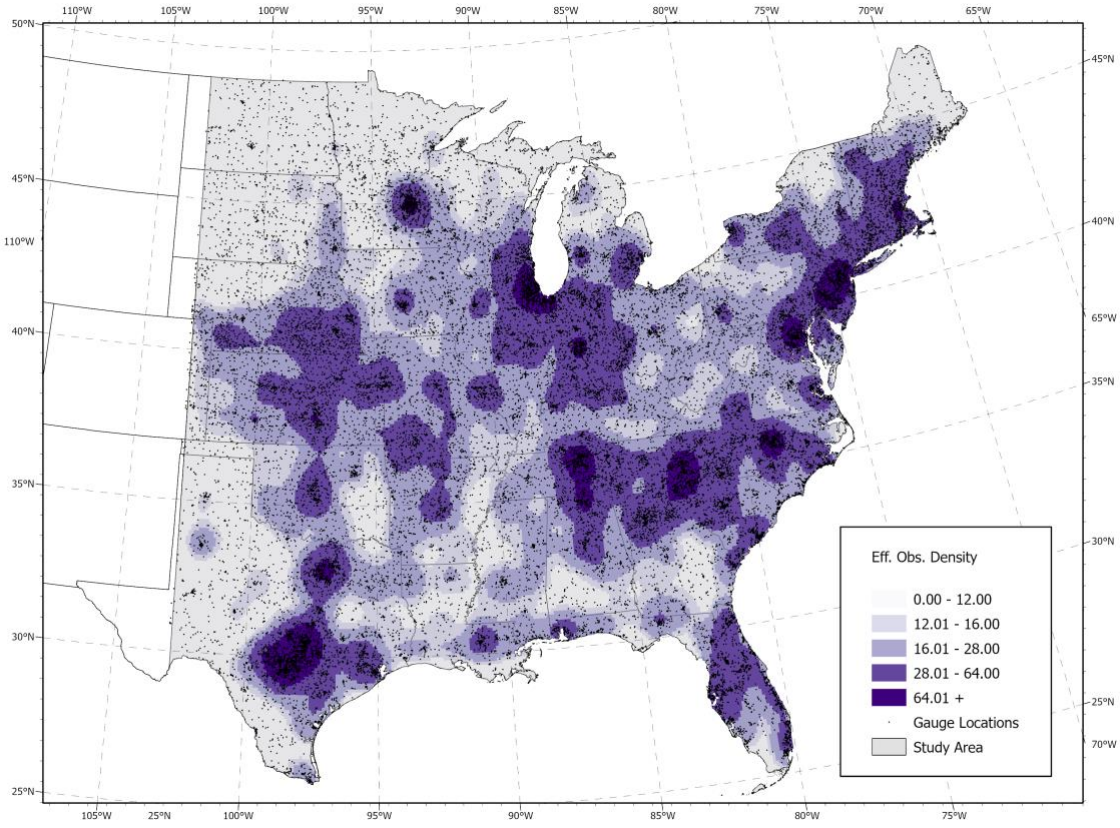

Figure 3.8: Study regions selected based on the effective observation ratio per 10 square kilometers. The regions are the five different shades of purple.

2020. The longest accumulation period in this study is used because gauges that are available at the longest accumulation period are also complete for all of the shorter periods. There are five different classes of effective observation density in Figure 3.9. These classes are selected based on the geometric interval. This method calculates the sum of squares, minimizes them, then selects classes in a way that makes the number of

Figure 3.9: Subregions selected based on terrain. The green areas are mountainous regions, and the gray regions are not.

observations in each class approximately equal (ESRI 2022). Since the radar estimates are compared with gauges, the amount of available gauge data can influence the performance of the bias correction and interpolation methods. Finally, figure 3.10 selects two regions based on terrain. The green polygon is a modified version of the Federal Aviation Agency (FAA) designated mountainous area (Durham and Haviland 2020). The modification occurs in the Mohawk Valley in central New York, where the FAA map has a gap in the polygon in it, but for this study, it is filled in. The valley is filled in because radar beams can still be blocked due to the neighboring mountains. The rest of

the map is considered non-mountainous. The mountainous areas can lead to less reliable results due to the higher likelihood of beam blockages. The previously discussed statistics and time periods are applied to the entire study area and subregions.

4. RESULTS

4.1. Beam Blockage Adjusted Results

The results of the performance of step 1, the beam blockage adjustments, are discussed in this section. Because the adjustments are radar based and do not consider gauge accumulations, there is no gauge withholding in the determination of performance. In order to calculate the performance between the unadjusted and adjusted data, the RMSE, MAE, and MB are taken at grid cells that are flagged and adjusted for beam blockage and their corresponding gauges. Additionally, the same metrics are calculated at the grid points that are not adjusted in order to show the magnitude of the beam blockage improvements relative to the unflagged grid cells. First, the results of the input parameter testing for the beam blockage adjustment algorithm will be shown. The tested parameters are the smoothing factor (σ), and the minimum correlation of the Sato filter pixel values with range. Then, this section will show the differences between the raw Stage IV and beam blockage corrected data for the entire study area for the three-year, one-year, and seasonal accumulations. Then, the results of the performance in the similar climatological characteristics and mountainous versus non-mountainous subregions are shown. Finally, the strengths and weaknesses of the algorithm are discussed.

The optimal input parameters for step 1 are determined through a series of tests. First, the algorithm is run three times using three different values of σ: 0.1, 0.5, and 1, while holding the minimum correlation constant at 0.5. Then, the three different

correlation values are tested: 0.4, 0.5, and 0.6 while holding sigma at 0.1. Only the three-year accumulations are tested because the adjustments are determined using these periods. Additionally, the results are generated for the entire study area. Table 1 shows the results of the σ value testing. The error metrics are the differences between the step 1

Step 1 vs Raw Stage IV Errors Using Different σ

σ Value	RMSE	MAE	MB
0.1	-0.0041	-0.0015	0.0359
0.5	-0.0219	-0.0104	0.0401
1.0	-0.0009	-0.0006	0.0365

Table 1: Step 1 minus the raw Stage IV PoN error metrics testing different smoothing values (σ) while holding the minimum correlation at 0.5.

and raw Stage IV data errors. A negative RMSE and MAE represents a decrease in error, and a positive MB represents an increase in bias due to the adjustments. It can be seen that the greatest decreases in error occur when σ is equal to 0.5. Also, this is the value with the greatest increase in bias. Because beam blocked cells are negatively biased, it is advantageous to select the parameter that increases the values. Using 0.5 as σ, the differences in error between the step 1 and raw Stage IV data are shown for the three

Step 1 vs Raw Stage IV Errors Using Different Correlations

Correlation Value	RMSE	MAE	MB
0.4	-0.0057	-0.0031	0.0312
0.5	-0.0219	-0.0104	0.0401
0.6	-0.0192	-0.0079	0.0433

Table 2: Step 1 minus the raw Stage IV error metrics testing different correlation values while holding σ at 0.5.

tested correlations in table 2. The greatest decreases in error occur when the correlation is 0.5, and the greatest increase in bias is when the correlation is 0.6. However, the difference in MB between the 0.5 and 0.6 correlations are small, so 0.5 is the selected value because it has the greatest improvement in error. With the optimal σ and correlation values, step 1 is run for all accumulation periods for the full study area.

The performance of step 1 is tested within the full study region at different accumulations. The results of the testing at the locations of the grid points that are adjusted are shown in table 3. The values in the table are representative of the average of

Step 1 vs Raw Stage IV Errors for the Full Study Area at Adjusted Locations

Accumulation Period	Raw RMSE	Step 1 RMSE	Raw MAE	Step 1 MAE	Raw MB	Step 1 MB
Three-Year	0.1835	**0.1413**	0.0958	**0.0843**	-0.0172	**-0.0009**
One-Year	0.1857	**0.1637**	0.1012	**0.0943**	-0.0127	**0.0079**
Winter	0.2903	0.2916	0.1455	**0.1433**	-0.0069	0.0135
Spring	0.2144	**0.1998**	0.1144	**0.1122**	0.0040	0.0226
Summer	0.2293	**0.2077**	0.1273	**0.1239**	-0.0221	**-0.0043**
Fall	0.2089	**0.1901**	0.1112	**0.1094**	-0.0262	**-0.0075**
Season Avg.	0.2353	**0.2219**	0.1243	**0.1219**	-0.0129	**0.0060**

Table 3: Step 1 versus raw Stage IV RMSE, MAE, and MB for all the adjusted locations within the full study area at all accumulation periods. Bolded numbers indicate an improvement in Step 1 versus the raw Stage IV data.

the given statistic for the entire study period for each accumulation. For example, the three-year winter RMSE is the average of all RMSE calculations for every 36-month period between 2005 and 2020. Additionally, if a number is in bold, it indicates that the adjustment led to an improvement over the raw data. For all of the results in table 3,

there are small differences in the original versus the adjusted data. Additionally, it was expected that the errors would decrease for all accumulations, however the results partially follow that hypothesis. The RMSE decreases for every accumulation except for the winter. The MAE, however, improves in all accumulations. When considering MB, the adjustments increase the biases for every accumulation. The adjusted MB is closer to zero in all of the periods except for winter and spring relative to the raw data, but the negatively biased beam blocked cells are increased for every period. To further diagnose

Step 1 Flagged vs Unflagged Grid Cell Errors for the Full Study Area

Accumulation Period	No Flag RMSE	RMSE Change	No Flag MAE	MAE Change	No Flag MB	MB Change
Three-Year	0.1375	0.9173	0.0775	0.6284	0.0167	0.4808
One-Year	0.1552	0.7213	0.0875	0.5036	0.0216	0.6005
Winter	0.2741	-0.0802	0.1348	0.2056	0.0214	0.7208
Spring	0.1947	0.7411	0.1058	0.2558	0.0329	0.6435
Summer	0.201	0.7632	0.1161	0.3035	0.0276	0.3581
Fall	0.182	0.6988	0.0982	0.1384	-0.0041	0.8461
Season Avg.	0.213	0.6008	0.1137	0.2264	0.0195	0.5833

Table 4: Unflagged grid cell RMSE, MAE, and MB versus the magnitude of the improvement after the step 1 adjustment relative to the unflagged grid cells.

the performance of the flagging of the beam blocked cells and how well the algorithm performs, the results in table 4 show the magnitude of improvement for each metric from the step 1 adjustments relative to the unflagged grid cell metrics. The closer the adjusted metrics are to the unflagged grid cell metrics, the better the performance. A value of 1.0 would mean that the adjustment error matches the unflagged grid cell error. Overall, the

magnitude of improvement in RMSE after step 1 is greater than 60%, except for the winter. The MAE improvements are smaller, but none are lower than 13%. Finally, the MB sees large improvements across all accumulations.

In order to identify how step 1 performs in different regions, the data are tested using the subregions selected based upon similar climatological characteristics for all three accumulations. It is expected that the results of the performance will vary based on the region and seasons because differences in precipitation type and intensity can exist. The three-year accumulation performance is shown in table 5. It can be seen that the RMSE decreases for every region. Additionally, the MAE decreases everywhere except

Three-Year Step 1 vs Raw Stage IV Errors in Climate Regions

Three-Year	Raw RMSE	Step 1 RMSE	Raw MAE	Step 1 MAE	Raw MB	Step 1 MB
Plains	0.1805	**0.1685**	0.1042	0.1102	0.0369	0.0635
Southwest	0.1546	**0.1322**	0.0880	**0.0750**	-0.0030	-0.0067
Northeast	0.2003	**0.1375**	0.1105	**0.0844**	-0.0396	**-0.0105**
Southeast	0.1467	**0.1289**	0.0885	**0.0774**	-0.0463	**-0.0301**

Table 5: Step 1 versus the raw Stage IV RMSE, MAE, and MB for the subregions selected based on similar climatological characteristics for the three-year accumulation period. Bolded numbers indicate an improvement in Step 1 versus the raw Stage IV data.

for the plains. For the MB, only the northeast and southeast biases move closer to zero, but the plains values increase, which follows the pattern of the adjustments increasing the bias for the full study area results in table 3. The three-year change in error metrics after step 1 is applied is shown in table 6 for the climatologically based subregions. For every RMSE, the magnitude of improvement is greater than 60%, which is similar to the

Three-Year Step 1 Change in Adjustment Relative to Unflagged Grid Cells

Three-Year	No Flag RMSE	RMSE Change	No Flag MAE	MAE Change	No Flag MB	MB Change
Plains	0.1635	0.7058	0.1006	-1.666	0.0760	0.6803
Southwest	0.1182	0.6153	0.0659	0.5882	0.0106	-0.2720
Northeast	0.1334	0.9387	0.0779	0.8006	0.0244	0.4546
Southeast	0.1190	0.6425	0.0686	0.5577	-0.0103	0.4500

Table 6: Unflagged grid cell RMSE, MAE, and MB versus the magnitude of the improvement after the step 1 adjustment relative to the unflagged grid cells in the climatological subregions for the three-year period.

results for the full study area. Additionally, most MAE values show improvement except for the plains. Finally, the plains, northeast, and southeast show large improvements in bias. The southwest MB, however, decreases, which goes against the pattern of step 1

One-Year Step 1 vs Raw Stage IV Errors in Climate Regions

One-Year	Raw RMSE	Step 1 RMSE	Raw MAE	Step 1 MAE	Raw MB	Step 1 MB
Plains	0.1989	**0.1913**	0.1188	**0.1179**	0.0520	0.0683
Southwest	0.1470	**0.1396**	0.0831	**0.0805**	0.0004	0.0089
Northeast	0.2041	**0.1602**	0.1071	**0.0962**	-0.0224	**0.0047**
Southeast	0.1651	**0.1471**	0.1038	**0.0940**	-0.0703	**-0.0363**

Table 7: Step 1 versus the raw Stage IV RMSE, MAE, and MB for the subregions selected based on similar climatological characteristics for the one-year accumulation period. Bolded numbers indicate an improvement in step 1 versus the raw Stage IV data.

increasing the MB. For the one-year accumulation in table 7, the results follow a similar pattern as the three-year except the MAE in every region is improved. The adjustments at the shorter accumulations are the same percent difference as those in the three-year, so

small differences in error could exist due to subtle changes in blockage strength. Table 8 shows the one-year change in error metrics after step 1 is applied for the climatological

One-Year Step 1 Change in Adjustment Relative to Unflagged Grid Cells

One-Year	No Flag RMSE	RMSE Change	No Flag MAE	MAE Change	No Flag MB	MB Change
Plains	0.1827	0.4691	0.1074	0.0789	0.0738	0.7477
Southwest	0.1393	0.9610	0.0792	0.6667	0.0095	0.9340
Northeast	0.1477	0.7783	0.0878	0.5647	0.0220	0.6103
Southeast	0.1336	0.5714	0.0781	0.3813	-0.0115	0.5782

Table 8: Unflagged grid cell RMSE, MAE, and MB versus the magnitude of the improvement after the step 1 adjustment relative to the unflagged grid cells in the climatological subregions for the one-year period.

subregions. The results differ from those in table 6, where the plains have a smaller RMSE improvement, but a larger MAE improvement. Additionally, the southwest rather than the northeast has the largest RMSE and MAE improvements. Finally, the errors within the seasonal accumulations are shown in tables 9 and 10, where table 9 tests the adjustment performance and table 10 quantifies the magnitude of the adjustments relative to the unflagged grid cells. The plains subregion consistently has the highest error out of the subregions for all accumulations, and the cold months have the highest error. Only the southern RMSE and MAE values show improvement within the winter. Additionally, the MB is increased for every region in the winter, but only the southern ones get closer to zero. The springtime errors decrease for all regions except for the northeast MAE, but the southeast MB is the only one that moves closer to zero. The biases, however, all increase following the adjustment. In the summer, the RMSE and

Seasonal Step 1 vs Raw Stage IV Errors in Climate Regions

Winter	Raw RMSE	Step 1 RMSE	Raw MAE	Step 1 MAE	Raw MB	Step 1 MB
Plains	0.4621	0.4715	0.3116	0.3418	0.2146	0.2650
Southwest	0.1828	**0.1782**	0.1063	**0.0995**	-0.0379	**-0.0315**
Northeast	0.2467	0.2476	0.1383	0.1506	0.0340	0.0574
Southeast	0.1875	**0.1770**	0.1219	**0.1070**	-0.0838	**-0.0598**
Spring	Raw RMSE	Step 1 RMSE	Raw MAE	Step 1 MAE	Raw MB	Step 1 MB
Plains	0.2369	**0.2320**	0.1361	**0.1359**	0.0787	0.0972
Southwest	0.2010	**0.1897**	0.1068	**0.1017**	0.0048	0.0082
Northeast	0.2157	**0.1862**	0.1153	0.1184	0.0036	0.0314
Southeast	0.1821	**0.1748**	0.1179	**0.1119**	-0.0623	**-0.0330**
Summer	Raw RMSE	Step 1 RMSE	Raw MAE	Step 1 MAE	Raw MB	Step 1 MB
Plains	0.2119	**0.2117**	0.1154	0.1156	0.0103	0.0227
Southwest	0.2136	**0.2086**	0.1385	**0.1291**	0.0213	0.0344
Northeast	0.2435	**0.2000**	0.1324	**0.1244**	-0.0592	**-0.0356**
Southeast	0.2187	**0.2053**	0.1391	**0.1273**	-0.0600	**-0.0285**
Fall	Raw RMSE	Step 1 RMSE	Raw MAE	Step 1 MAE	Raw MB	Step 1 MB
Plains	0.2124	**0.2056**	0.1196	0.1237	0.0311	0.0457
Southwest	0.1744	0.1748	0.1016	**0.1006**	-0.0127	**-0.0032**
Northeast	0.2149	**0.1720**	0.1171	**0.1063**	-0.0428	**-0.0176**
Southeast	0.2042	**0.1917**	0.1103	**0.1097**	-0.0691	**-0.0496**

Table 9: Step 1 versus the raw Stage IV RMSE, MAE, and MB for the subregions selected based on similar climatological characteristics for the seasonal accumulation periods. Bolded numbers indicate an improvement in step 1 versus the raw Stage IV data.

MAE are improved in every region like the spring, except for the plains MAE. In addition to the southeast, the northeast MB moves closer to zero. Finally, the fall has the

Seasonal Step 1 Change in Adjustment Relative to Unflagged Grid Cells

Winter	No Flag RMSE	RMSE Change	No Flag MAE	MAE Change	No Flag MB	MB Change
Plains	0.4026	-0.1579	0.2375	-0.4075	0.1654	-1.024
Southwest	0.1896	-0.6764	0.0990	0.9315	-0.0340	1.6410
Northeast	0.2315	-0.0592	0.1366	-7.2352	0.0520	1.3000
Southeast	0.1716	0.6603	0.0996	0.6681	-0.0453	0.6233
Spring	No Flag RMSE	RMSE Change	No Flag MAE	MAE Change	No Flag MB	MB Change
Plains	0.2222	0.3333	0.1284	0.0259	0.0838	3.6274
Southwest	0.1896	0.9912	0.0989	0.6455	0.0145	0.3505
Northeast	0.1755	0.7338	0.1024	-0.2403	0.0356	0.8687
Southeast	0.1704	0.6239	0.0978	0.2985	-0.0096	0.5559
Summer	No Flag RMSE	RMSE Change	No Flag MAE	MAE Change	No Flag MB	MB Change
Plains	0.2116	0.6667	0.1221	0.0298	0.0516	0.3002
Southwest	0.1985	0.3311	0.1171	0.4392	0.0464	0.5219
Northeast	0.1871	0.7712	0.1123	0.3980	-0.0044	0.4306
Southeast	0.1934	0.5296	0.1162	0.5152	0.0151	0.4194
Fall	No Flag RMSE	RMSE Change	No Flag MAE	MAE Change	No Flag MB	MB Change
Plains	0.1987	0.4963	0.1088	-0.3796	0.0411	1.4600
Southwest	0.1721	-0.1739	0.0922	0.1063	-0.0053	1.2837
Northeast	0.1625	0.8187	0.0949	0.4864	-0.0109	0.7899
Southeast	0.1775	0.4681	0.0999	0.0576	-0.0299	0.4974

Table 10: Unflagged grid cell RMSE, MAE, and MB versus the magnitude of the improvement after the step 1 adjustment relative to the unflagged grid cells in the climatological subregions for the seasonal periods.

most regions with improved biases, and it mostly has decreases in error. The biases for all seasons increase with the adjustment, but the unadjusted fall biases are the most

negative. Overall, the performance of the beam blockage algorithm within the different regions is seasonally dependent. In table 10, the causes of the increase in errors can be identified. In the winter, the southern regions showed improvement relative to the unadjusted data, however, every RMSE difference is further from the unflagged grid cell RMSE. This suggests that the increases in error can be attributed to biased data being incorporated into the adjustment calculations. Another potential issue with the algorithm can be seen when the change magnitudes are greater than 1. This suggests a possible error in identifying blocked beams, however, the adjustment calculations are performed on the 36-month period and may degrade as the accumulation periods decrease in length. Incorporation of rain gauge data in the next steps of the bias adjustment methodology can help adjust for these errors.

The final tested subregions are the mountainous versus non-mountainous regions.

Step 1 vs Raw Stage IV Errors in Mountainous Region

Accumulation Period	Raw RMSE	Step 1 RMSE	Raw MAE	Step 1 MAE	Raw MB	Step 1 MB
Three-Year	0.1588	**0.1423**	0.1033	0.1035	-0.0532	**-0.0217**
One-Year	0.1590	**0.1565**	0.1055	0.1092	-0.0374	**-0.0108**
Winter	0.2176	0.2323	0.1456	0.1576	-0.0103	**0.0012**
Spring	0.1831	0.1847	0.1111	0.1154	-0.0148	**0.0095**
Summer	0.2013	**0.1935**	0.1264	**0.1256**	-0.0444	**-0.0248**
Fall	0.1832	**0.1791**	0.1174	**0.1147**	-0.0628	**-0.0479**
Season Avg.	0.1952	0.1971	0.1249	0.1281	-0.0335	**-0.0160**

Table 11: Step 1 versus the raw Stage IV RMSE, MAE, and MB for the mountainous region at all accumulation periods. Bolded numbers indicate an improvement in step 1 versus the raw Stage IV data.

It is expected that the adjustments will more greatly improve the data in the non-mountainous over the mountainous region because there is less likely to be widespread beam blockages in a non-mountainous radar domain, so identification of the blockages is easier, and the adjustment is more likely to use less biased grid cells. It is also important to note that the entirety of the mountainous region is within the southeast and northeast subregions. Using the same testing methods as the climatological subregions, table 11 shows the results of step 1 versus the raw Stage IV data at the adjusted grid cells for all accumulations in the mountainous region. For all accumulations, the dataset becomes more positively biased with the adjustments, where the MB moves closer to zero in every accumulation. The RMSE improves in every accumulation except winter and spring, but the MAE only improves in the summer and fall. Like the previous beam blockage adjustment errors, the colder months do not see improvement with the

Step 1 Change in Adjustment Relative to Unflagged Grid Cells in Mountainous

Accumulation Period	No Flag RMSE	RMSE Change	No Flag MAE	MAE Change	No Flag MB	MB Change
Three-Year	0.1297	0.5670	0.0766	-0.0074	-0.0080	0.6969
One-Year	0.1451	0.1798	0.0864	-0.1937	-0.0107	0.9962
Winter	0.2169	-21.00	0.1309	-0.8163	-0.0012	1.2637
Spring	0.1728	-0.1553	0.1005	-0.4056	0.0047	1.2461
Summer	0.1848	0.4727	0.1118	0.0547	-0.0199	0.8000
Fall	0.1703	0.3178	0.1020	0.1753	-0.0389	0.6234
Season Avg.	0.1862	-0.2111	0.1113	-0.2352	-0.0137	0.8838

Table 12: Unflagged grid cell RMSE, MAE, and MB versus the magnitude of the improvement after the step 1 adjustment relative to the unflagged grid cells in the mountainous region for all periods.

adjustment over the raw data. When looking at the change in the step 1 adjusted data relative to the unflagged grid points in the mountainous region in table 12, the results show variable performance. The longer accumulations show large improvements in RMSE, but the other errors move further away from the unflagged grid cell errors. Additionally, the MB shows changes greater than 1 in the winter and spring. The changes in error and bias suggest that the flagging does not always perform well in the mountainous region as expected. Tables 13 and 14 are calculated the same way as tables 11 and 12, except they represent the non-mountainous results instead. The errors of the

Step 1 vs Raw Stage IV Errors in Non-Mountainous Region

Accumulation Period	Raw RMSE	Step 1 RMSE	Raw MAE	Step 1 MAE	Raw MB	Step 1 MB
Three-Year	0.1775	**0.1409**	0.0926	**0.0839**	-0.0139	**0.0026**
One-Year	0.1900	**0.1647**	0.1003	**0.0933**	-0.0080	0.0110
Winter	0.3024	**0.3012**	0.1445	**0.1404**	-0.0053	0.0178
Spring	0.2190	**0.2018**	0.1146	**0.1110**	0.0077	0.0288
Summer	0.2335	**0.2099**	0.1290	**0.1232**	-0.0209	**-0.0005**
Fall	0.2129	**0.1916**	0.1106	**0.1082**	-0.0191	-0.0015
Season Avg.	0.2415	**0.2258**	0.1243	**0.1204**	-0.0094	0.0111

Table 13: Step 1 versus the raw Stage IV RMSE, MAE, and MB for the non-mountainous region at all accumulation periods. Bolded numbers indicate an improvement in step 1 versus the raw Stage IV data.

adjustments are lower, and the magnitude of improvement is larger in the non-mountainous compared to the mountainous regions. The non-mountainous region RMSE and MAE improves for every accumulation period, but the MB only moves closer to zero in the three-year and summer accumulations. The MB in the mountainous region is

generally more negative and greater in magnitude than the non-mountainous region. When looking at the changes relative to the unflagged grid cells in table 14, the results show success in the algorithm. Every metric shows that the step 1 adjustment errors

Step 1 Change in Adjustment Relative to Unflagged Grid Cell in Non-Mountainous

Accumulation Period	No Flag RMSE	RMSE Change	No Flag MAE	MAE Change	No Flag MB	MB Change
Three-Year	0.1402	0.9812	0.0779	0.5918	0.0316	0.3626
One-Year	0.1564	0.7529	0.0877	0.5555	0.0306	0.4922
Winter	0.2825	0.0603	0.1356	0.4606	0.0255	0.7500
Spring	0.1975	0.8000	0.1070	0.4736	0.0379	0.6986
Summer	0.2032	0.7788	0.1169	0.4793	0.0349	0.3655
Fall	0.1813	0.6740	0.0973	0.1804	0.0065	0.6875
Season Avg.	0.2164	0.6254	0.1141	0.3823	0.0260	0.5790

Table 14: Unflagged grid cell RMSE, MAE, and MB versus the magnitude of the improvement after the step 1 adjustment relative to the unflagged grid cells in the non-mountainous region for all periods.

become closer to the unflagged grid cell errors. Overall, the beam blockage algorithm leads to larger improvements in magnitude and better flagging performance in the non-mountainous versus the mountainous region.

Overall, the beam blockage adjustments generally lead to improvements over the raw Stage IV data by reducing the error and positively increasing negative biases. Figure 4.1 shows the radar estimated precipitation before and after step 1 is applied in the southeast US for the three-year accumulation through November 2020. The left image is before the adjustment, the middle image is after the adjustment, and the right image shows the magnitude of the adjustments for each grid cell. A noteworthy feature is that

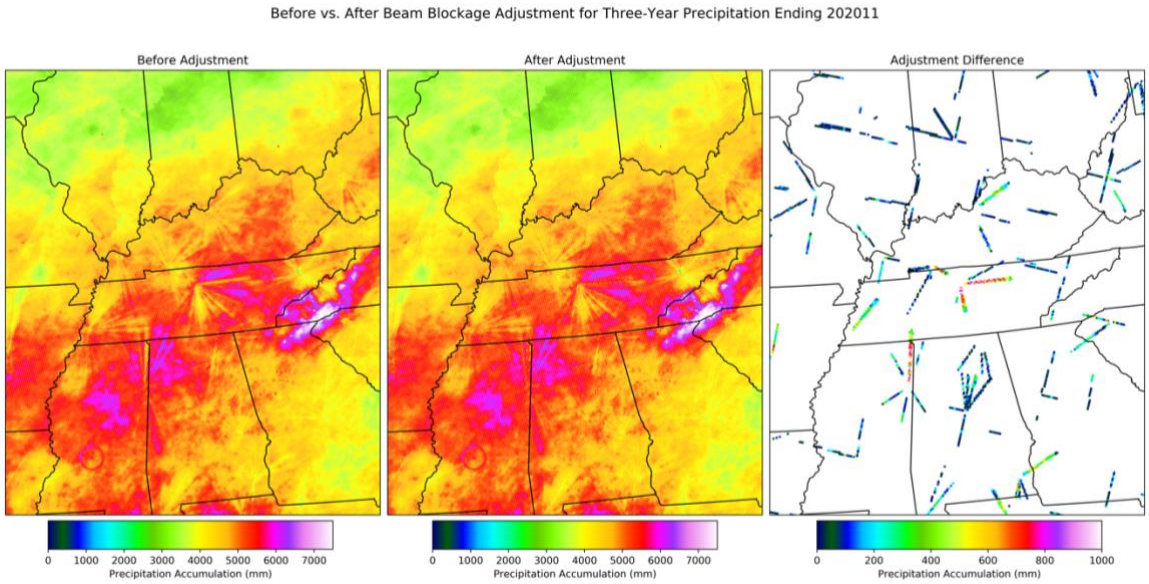

Figure 4.1: Radar estimates before step 1(left), after step 1 (middle), and the magnitude of the adjustment for each blocked beam for the three-year accumulation period through November 2020.

most of the beams that appear to be blocked are adjusted. There are numerous beam blockage examples, like the ones in north-central Tennessee, the northeast corner of Mississippi, and northern Georgia. After step 1 is applied, these blockages appear to be corrected as shown in the middle and right images. A weakness of the algorithm, however, is the adjustment of wide beam blockages. There are a handful of examples of wide beam blockages in figure 4.1, especially in Tennessee and Kentucky. In the adjustments, there are still faint reductions of estimated precipitation within the wide beams relative to the neighboring grid cells, but the errors appear to be smaller in magnitude. The wide beams generally appear in areas with complex topography, and the mountainous region results suggest that the algorithm struggles with this. Finally, figure 4.2 shows the difference in gauge-radar biases from after and before step 1 is applied for

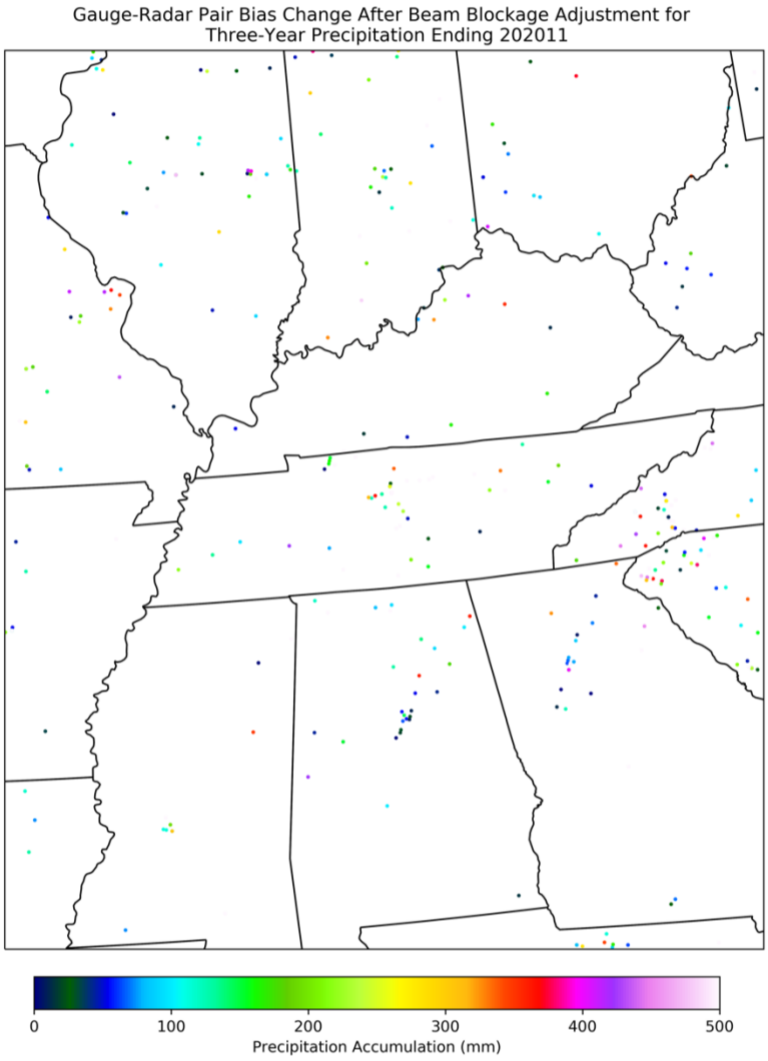

Figure 4.2: The difference between the gauge-radar bias after and before step 1 for three-year precipitation through November 2020.

the same region and accumulation period as figure 4.1. Only the pairs that change from the adjustment are shown. Every point in the map is a positive value, which indicates that the radar estimates are all increased. The remaining biases in the beam blockage adjusted dataset can be addressed in the later steps of the bias adjustment methodology.

4.2. Mean Field and Range Dependent Adjustment Results

The results of the performance of step 2, the mean field and range dependent adjustments, are discussed in this section. Because rain gauges are used in the adjustment calculations, data withholding tests are conducted to determine the error metrics. The gauges are first quality controlled, then 20% of them are randomly withheld and the RMSE, MAE, and MB are calculated between the withheld gauge PoN and the original analysis PoN at those locations. The original analysis for this step of the bias correction algorithm is the output from step 1. With the remaining gauges, the mean field and range dependent analysis is run for every grid point, then the error metrics are calculated again at the withheld locations with the newly calculated analysis. This section will first show the differences between step 1 and step 2 for the entire study area for the three-year, one-year, and seasonal accumulations. Then, the results of the performance in the similar climatological characteristics, mountainous versus non-mountainous, and effective observation density subregions are shown for each accumulation. Finally, the strengths and weaknesses of the algorithm are discussed.

The performance of step 2 is first tested for the entire study region at all accumulations. Table 15 shows the step 1 versus the step 2 RMSE, MAE, and MB using the withheld gauges. For every accumulation, the RMSE and MAE improve after step 2 is applied. Additionally, the MB is improved for every accumulation. The adjustments in step 1 led to increases in MB, but with the incorporation of gauge data, some of the adjustments in step 2 led to decreases in MB. Despite the adjustments being tested and

Step 2 vs Step 1 Errors for the Full Study Area

Accumulation Period	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
Three-Year	0.1451	**0.1275**	0.0738	**0.0697**	-0.0182	**-0.0046**
One-Year	0.1696	**0.1566**	0.0847	**0.0807**	-0.0245	**-0.0023**
Winter	0.2314	**0.2296**	0.1347	**0.1204**	0.0398	**0.0027**
Spring	0.1830	**0.1690**	0.1122	**0.0991**	0.0352	**-0.0013**
Summer	0.1908	**0.1795**	0.1171	**0.1089**	0.0296	**0.0176**
Fall	0.1710	**0.1676**	0.1035	**0.0953**	-0.0035	**-0.0012**
Season Avg.	0.1940	**0.1864**	0.1169	**0.1039**	0.0253	**0.0045**

Table 15: Step 2 versus step 1 RMSE, MAE, and MB for the full study area at all accumulation periods. Bolded numbers indicate an improvement in the step 2 versus step 1.

applied seasonally, the magnitude of the improvements is similar for each accumulation. The greatest improvement in RMSE is in the summer, and the smallest is in the winter. This is similar to how the step 1 adjustments performed. Errors in wintertime adjustments can be related to frozen precipitation within certain gauges. The winter, however, has a much larger improvement in MAE relative to the step 1 adjustments.

The performance of step 2 for the different subregions selected based on similar climatological characteristics is shown. Performance is expected to vary between the regions due to their different climatological features. Table 16 shows the results between the step 1 and step 2 data at the withheld gauge locations within the climatological subregions for the three-year accumulation. Like the results in table 15, the errors in step 2 are improved over the errors in step 1. Additionally, the MB is reduced for every region except for the southwest. The plains have the highest error out of every

Three-Year Step 2 vs Step 1 Errors in Climate Regions

Three-Year	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
Plains	0.1659	**0.1325**	0.0985	**0.0775**	0.0609	**-0.0127**
Southwest	0.1212	**0.1159**	0.0663	**0.0645**	0.0043	0.0053
Northeast	0.1396	**0.1303**	0.0756	**0.0688**	0.0100	**-0.0098**
Southeast	0.1229	**0.1218**	0.0711	**0.0643**	-0.0198	**0.0057**

Table 16: Step 2 versus step 1 RMSE, MAE, and MB for the subregions selected based on climatological characteristics for the three-year accumulation period. Bolded numbers indicate an improvement in step 2 versus step 1.

subregion, but they also have the greatest improvement following the adjustment. The southwest has the least amount of improvement out of the regions. For the one-year accumulation in table 17, the results follow a similar pattern as those in table 16.

One-Year Step 2 vs Step 1 Errors in Climate Regions

One-Year	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
Plains	0.1837	**0.1493**	0.1077	**0.0846**	0.0700	**-0.0022**
Southwest	0.1444	**0.1400**	0.0789	**0.0781**	0.0069	0.0102
Northeast	0.1558	**0.1453**	0.0870	**0.0796**	0.0184	**-0.0016**
Southeast	0.1414	**0.1401**	0.0804	**0.0789**	-0.0169	**0.0070**

Table 17: Step 2 versus step 1 RMSE, MAE, and MB for the subregions selected based on climatological characteristics for the one-year accumulation period. Bolded numbers indicate an improvement in the adjusted versus unadjusted data.

Interestingly, the MB in the southwest increases in both the three-year and one-year accumulations. Finally, the seasonal accumulation testing results are shown in table 18. Generally, the results show that the adjustments improve the estimates, however, the

Seasonal Step 2 vs Step 1 Errors in Climate Regions

Winter	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
Plains	0.3055	**0.2950**	0.2075	**0.1589**	0.1336	**0.0129**
Southwest	0.1768	**0.1758**	0.1066	**0.1012**	-0.0351	**0.0079**
Northeast	0.2156	**0.1911**	0.1346	**0.1180**	0.0515	**-0.0102**
Southeast	0.1714	**0.1705**	0.1037	**0.0970**	-0.0435	**-0.0047**
Spring	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
Plains	0.2037	**0.1703**	0.1256	**0.1030**	0.0816	**-0.0003**
Southwest	0.1837	**0.1803**	0.0997	**0.0973**	0.0144	**0.0079**
Northeast	0.1703	**0.1609**	0.1020	**0.0979**	0.0322	**-0.0145**
Southeast	0.1651	**0.1644**	0.0988	**0.0968**	-0.0089	**0.0080**
Summer	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
Plains	0.1998	**0.1730**	0.1220	**0.1046**	0.0564	**0.0015**
Southwest	0.1917	**0.1851**	0.1195	**0.1148**	0.0487	**0.0316**
Northeast	0.1842	**0.1747**	0.1116	**0.1050**	-0.0034	0.0166
Southeast	0.1878	**0.1852**	0.1156	**0.1151**	0.0196	0.0292
Fall	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
Plains	0.1856	**0.1805**	0.1092	**0.1034**	0.0307	**-0.0063**
Southwest	0.1655	0.1661	0.0923	**0.0908**	-0.0097	**0.0059**
Northeast	0.1588	**0.1529**	0.0951	**0.0893**	-0.0152	**-0.0050**
Southeast	0.1730	**0.1723**	0.0993	**0.0964**	-0.0309	**0.0053**

Table 18: Step 2 versus step 1 RMSE, MAE, and MB for the subregions selected based on climatological characteristics for the seasonal accumulation periods. Bolded numbers indicate an improvement in step 2 versus step 1.

magnitude of the improvements is seasonally dependent. The plains in the winter have the highest error of any region at any time. The RMSE, MAE, and MB are improved for

every season, but the improvements in the warmer months are much larger than in the colder months. The southwest sees improvements for all metrics except for the fall RMSE. Additionally, the MB is decreased in the warmer months, but increased in the cooler months. The northeast and southeast MB, like in the southwest, does not consistently decrease across the seasons. The northeast MB increases in the summer and fall, and the southeast MB decreases in all seasons except for the winter. The RMSE and MAE is improved for every accumulation in these regions as well.

The step 2 adjustments are tested in the mountainous versus the non-mountainous subregions using the same methods as the prior tests. Table 19 shows the results of step

Step 2 vs Step 1 Errors in Mountainous Region

Accumulation Period	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
Three-Year	0.1349	**0.1284**	0.0758	**0.0722**	-0.0157	**-0.0042**
One-Year	0.1568	**0.1419**	0.0861	**0.0843**	-0.0111	**-0.0008**
Winter	0.2132	**0.1976**	0.1343	**0.1232**	0.0073	**-0.0068**
Spring	0.1687	**0.1545**	0.1007	0.1019	0.0049	**-0.0042**
Summer	0.1784	**0.1765**	0.1129	**0.1063**	-0.0013	0.0106
Fall	0.1655	**0.1600**	0.0983	**0.0961**	-0.0265	**0.0015**
Season Avg.	0.1815	**0.1722**	0.1113	**0.1066**	-0.0039	**0.0003**

Table 19: Step 2 versus step 1 RMSE, MAE, and MB for the mountainous region for all accumulation periods. Bolded numbers indicate an improvement in step 2 versus step 1.

1 versus step 2 for the mountainous region. There is a reduction of error between step 1 and step 2 for every accumulation except for the spring MAE, and the MB is improved for all periods except for the summer. Like the other tests, the winter has the highest

errors and the greatest improvements. In the summertime, the MB has a large increase. This was also observed for the northeast and southeast in the summer, which is where the entirety of the mountainous region is located. The non-mountainous region results are shown in table 20. Overall, the errors are higher in the non-mountainous regions. Unlike the mountainous region, the winter RMSE increases, and the MB is

Step 2 vs Step 1 Errors in Non-Mountainous Region

Accumulation Period	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
Three-Year	0.1502	**0.1270**	0.0785	**0.0681**	0.0023	**-0.0046**
One-Year	0.1759	**0.1589**	0.0893	**0.0791**	0.0279	**0.0033**
Winter	0.2404	0.2454	0.1355	**0.1203**	0.0428	**0.0039**
Spring	0.2051	**0.1894**	0.1156	**0.0952**	0.0425	**0.0017**
Summer	0.2036	**0.1817**	0.1258	**0.1119**	0.0367	**0.0190**
Fall	0.1834	**0.1771**	0.1073	**0.1008**	0.0143	**-0.0019**
Season Avg.	0.2081	**0.1969**	0.1204	**0.1091**	0.0341	**0.0057**

Table 20: Step 2 versus step 1 RMSE, MAE, and MB for the non-mountainous region for all accumulation periods. Bolded numbers indicate an improvement in step 2 versus step 1.

improved for every accumulation. Additionally, the MB for every accumulation is reduced. Another noteworthy feature is that every MB is more positive in the non-mountainous region compared to the mountainous. This shows that the gauge estimates of precipitation are higher than the radar in the mountainous regions and the radar estimates are higher in the non-mountainous regions.

The final tested subregions are based upon the five effective observation density classes. It is expected that the regions with more gauge observations will have lower

Three-Year Step 2 vs Step 1 Errors in Gauge Density Regions

Three-Year Effective Obs. Density	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
0.0-12.0	0.1880	**0.1461**	0.1044	**0.0930**	0.0477	**-0.0049**
12.01-16.00	0.1379	**0.1326**	0.0836	**0.0697**	0.0335	**0.0124**
16.01-28.00	0.1427	**0.1251**	0.0786	**0.0705**	0.0215	**-0.0044**
28.01-64.0	0.1329	**0.1206**	0.0742	**0.0655**	0.0123	**-0.0062**
64.01+	0.1237	**0.1224**	0.0702	**0.0640**	-0.0015	-0.0052

Table 21: Step 2 versus step 1 RMSE, MAE, and MB for the subregions selected based on effective observation density for the three-year accumulation period. Bolded numbers indicate an improvement in step 2 versus step 1.

error and greater improvements compared to the areas with less observations. Table 21 shows the results of the errors between step 1 and step 2 at the withheld gauge locations

One-Year Step 2 vs Step 1 Errors in Gauge Density Regions

One-Year Effective Obs. Density	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
0.0-12.0	0.2197	**0.1742**	0.1269	**0.1054**	0.0632	**0.0031**
12.01-16.00	0.1628	**0.1516**	0.0933	**0.0829**	0.0381	**0.0158**
16.01-28.00	0.1572	**0.1417**	0.0873	**0.0801**	0.0262	**0.0002**
28.01-64.0	0.1494	**0.1374**	0.0830	**0.0736**	0.0160	**0.0004**
64.01+	0.1429	**0.1403**	0.0803	**0.0795**	0.0137	**0.0102**

Table 22: Step 2 versus step 1 RMSE, MAE, and MB for the subregions selected based on effective observation density for the one-year accumulation period. Bolded numbers indicate an improvement in step 2 versus step 1.

for the three-year accumulation for each effective observation density class. Every metric shows improvement between step 1 and step 2 except for the MB in the final

Seasonal Step 2 vs Step 1 Errors in Gauge Density Regions

Winter	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
0.0-12.0	0.3047	0.3191	0.1717	**0.1696**	0.0153	0.0193
12.01-16.00	0.2498	0.2634	0.1387	**0.1258**	0.0089	**0.0033**
16.01-28.00	0.2320	**0.2196**	0.1374	**0.1162**	0.0316	**-0.0005**
28.01-64.0	0.2108	**0.1882**	0.1223	**0.1066**	0.0172	**-0.0015**
64.01+	0.1925	**0.1711**	0.1140	**0.1086**	0.0009	0.0041
Spring	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
0.0-12.0	0.2504	**0.2146**	0.1562	**0.1266**	0.0771	**0.0169**
12.01-16.00	0.2014	**0.1854**	0.1187	**0.1107**	0.0485	**0.0107**
16.01-28.00	0.1840	**0.1658**	0.1080	**0.0960**	0.0418	**0.0003**
28.01-64.0	0.1677	**0.1572**	0.0984	**0.0919**	0.0255	**-0.0029**
64.01+	0.1652	**0.1640**	0.1035	**0.0949**	0.0122	**-0.0048**
Summer	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
0.0-12.0	0.2307	**0.2004**	0.1445	**0.1259**	0.0806	**-0.0056**
12.01-16.00	0.1989	**0.1894**	0.1268	**0.1173**	0.0560	**0.0259**
16.01-28.00	0.1959	**0.1816**	0.1170	**0.1083**	0.0299	**0.0217**
28.01-64.0	0.1851	**0.1715**	0.1131	**0.1046**	0.0246	**0.0119**
64.01+	0.1741	**0.1730**	0.1035	**0.1013**	0.0148	0.0261
Fall	Step 1 RMSE	Step 2 RMSE	Step 1 MAE	Step 2 MAE	Step 1 MB	Step 2 MB
0.0-12.0	0.2263	**0.2101**	0.1304	**0.1227**	0.0282	**0.0066**
12.01-16.00	0.1989	**0.1894**	0.1268	**0.1173**	0.0560	**0.0259**
16.01-28.00	0.1669	**0.1665**	0.0948	**0.0935**	-0.0019	-0.0037
28.01-64.0	0.1647	**0.1631**	0.0941	**0.0908**	-0.0118	**-0.0032**
64.01+	0.1551	0.1573	0.0883	**0.0871**	-0.0031	**0.0024**

Table 23: Step 2 versus step 1 RMSE, MAE, and MB for the seasonal effective observation density regions. Bolded numbers indicate an improvement in step 2.

class. The grid points in the first class expectedly have the highest errors, but they also benefit the most from the improvements. The magnitudes of the improvements in the fifth class are the smallest, but their overall errors are the lowest. When looking at the one-year accumulation in table 22, the results follow the same pattern. As the effective observation density increases, the amount of improvement decreases. Additionally, every MB is reduced in both the three-year and one-year accumulations. Finally, when considering the seasons, there are differences with the performance of the algorithm. Like many of the prior adjustments, the winter errors are the highest and do not always improve between the original and the adjusted data. Gauge errors in the winter can be related to frozen precipitation. The RMSE increases within the first two classes instead of decreases like the three-year and one-year accumulations. The MB also only improves in the second, third, and fourth classes. Unlike the other accumulations, the MB increases in the first and fifth class. The spring is the only season where every metric shows improvement with the adjustments. It is also the only season where every MB decreases. For the step 2 adjustments as a whole, the spring performs well. The errors in the summer follow a similar pattern as the spring, except the MB does not improve for the fifth class. Finally, the fall is the season with the least amount of error. Like the other accumulations with the least amount of error, the magnitudes of the improvements are small compared to the higher error accumulations.

Overall, the adjustments in step 2 result in improvements over step 1. When considering the errors for the entire study region, the adjustments successfully reduce both the RMSE and MAE for all accumulations. The MB is improved in most cases as

well. Figure 4.3 shows maps of before and after the adjustment, and the difference between step 2 and step 1 for the southeast US for the three-year precipitation through November 2020. The adjustments are equal in magnitude within the same ranges around

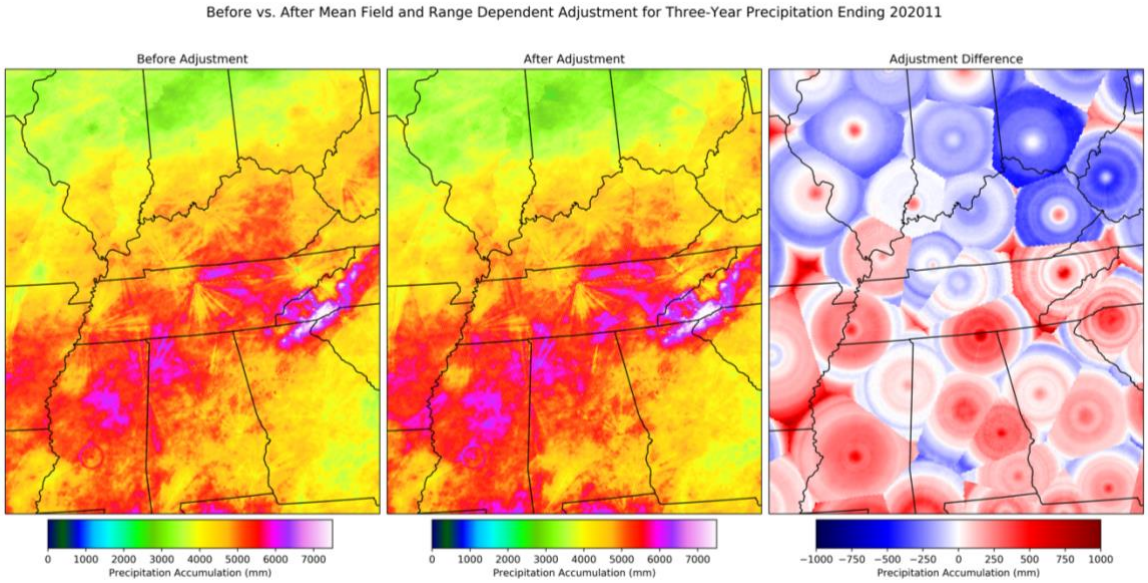

Figure 4.3: Radar estimates before step 2(left), after step 2 (middle), and the magnitude of the adjustments for the three-year accumulation period through November 2020.

the radar, which appear as circles. The mean field and range dependent biases are roughly equal with range, so this shows success in the algorithm. One weakness with the algorithm is the development of lines of discontinuity at the edges of the radar domain boundaries. A prominent example in the figure occurs in southeast Ohio, where there is a sharp discontinuity between green and orange colors. These lines occur because the adjustments are performed independently for each radar, and this error will be adjusted in the next step of the methodology. Finally, figure 4.4 shows the difference between after and before the step 2 adjustment in gauge-radar pair biases. The blue and red colors indicate that the radar estimated value increases and decreases respectively relative to

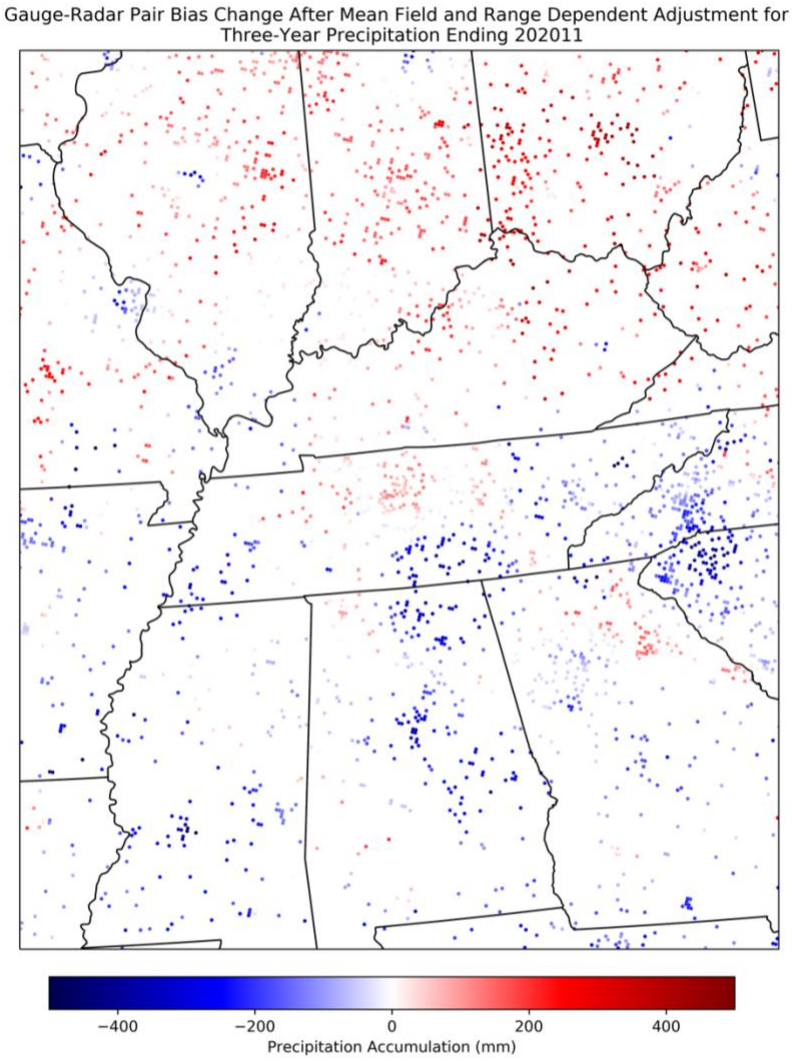

Figure 4.4: The difference between the gauge-radar bias after and before step 2 for three-year precipitation through November 2020.

the gauge value. It can be seen that alike colors tend to cluster in regions corresponding to radar domains. For example, the large area of red points indicating that the step 2 adjustment led to decreases in the radar estimated precipitation in southwest Ohio corresponds to the KILN radar domain. Overall, many of the differences are large. The remaining adjustments in the methodology focus on radar domain boundaries and the precipitation field as whole.

4.3. Radar Domain Discontinuity Adjustment Results

The results of step 3, the radar domain discontinuity adjustments, are discussed in this section. The RMSE, MAE, and MB are calculated for the three-year, one-year, and seasonal accumulations for the entire study area, climatological subregions, and mountainous versus non-mountainous regions. The effective observation density classes are not considered because the adjustments are entirely radar based. Additionally, no gauges are withheld in the error calculations. Finally, the errors are only calculated at the locations of the grid cell and gauge pairs that were adjusted, which are those within 25 km of every radar domain boundary.

The errors between the step 2 and step 3 adjustments are calculated for the full study area for all three accumulation periods. It is expected that the adjustments will result in improvements, but the changes will be small because of the IDW method. Table 24 shows the results of these calculations at the locations of the grid points that were

Step 3 vs Step 2 Errors for the Full Study Area

Accumulation Period	Step 2 RMSE	Step 3 RMSE	Step 2 MAE	Step 3 MAE	Step 2 MB	Step 3 MB
Three-Year	0.1183	**0.1167**	0.0691	**0.0680**	0.0004	**0.0003**
One-Year	0.1350	**0.1335**	0.0790	**0.0780**	0.0030	**0.0028**
Winter	0.2062	**0.2052**	0.1177	**0.1163**	0.0055	0.0061
Spring	0.1713	**0.1694**	0.0972	**0.0964**	0.0008	0.0014
Summer	0.1811	**0.1790**	0.1079	**0.1071**	0.0104	**0.0097**
Fall	0.1723	**0.1712**	0.0977	**0.0969**	-0.0001	-0.0002
Season Avg.	0.1825	**0.1810**	0.1050	**0.1040**	0.0041	0.0042

Table 24: Step 3 versus step 2 RMSE, MAE, and MB for all accumulation periods for the full study area. The bolded numbers indicate an improvement in step 3 versus step 2.

adjusted. Overall, the changes in error and bias are small between step 2 and step 3. The three-year and one-year accumulations show improvement for every metric. Additionally, the RMSE and MAE for every season and the summer MB are improved. The changes in MB for the winter, spring, and fall move further away from zero, but the differences from the unadjusted data are small. Unlike the prior steps of the bias adjustment procedure, the changes across all time periods are similar in magnitude.

The results of the step 3 adjustments for the climatological subregions are discussed for each accumulation. Table 25 shows the results between step 2 and step 3 for the three-year accumulation period. Every metric shows improvement for every

Three-Year Step 3 vs Step 2 Errors in Climate Regions

Three-Year	Step 2 RMSE	Step 3 RMSE	Step 2 MAE	Step 3 MAE	Step 2 MB	Step 3 MB
Plains	0.1219	**0.1194**	0.0710	**0.0694**	-0.0090	**-0.0089**
Southwest	0.1137	**0.1106**	0.0636	**0.0628**	0.0120	**0.0114**
Northeast	0.1209	**0.1187**	0.0725	**0.0716**	-0.0017	**-0.0003**
Southeast	0.1158	**0.1152**	0.0678	**0.0677**	0.0056	**0.0050**

Table 25: Step 3 versus step 2 RMSE, MAE, and MB for the three-year accumulation period for the climatological subregions. The numbers in bold indicate an improvement in step 3 versus step 2.

subregion. The magnitudes of the improvements are smallest in the southeast, and the other regions have similarly sized improvements. Additionally, the plains region has the highest RMSE, but the northeast has the highest MAE. This is different from the pattern of the plains always having the highest error in the prior adjustments. The northern regions are also the only negatively biased regions. The one-year adjustments are shown

One-Year Step 3 vs Step 2 Errors in Climate Regions

One-Year	Step 2 RMSE	Step 3 RMSE	Step 2 MAE	Step 3 MAE	Step 2 MB	Step 3 MB
Plains	0.1394	**0.1374**	0.0838	**0.0827**	-0.0055	**-0.0046**
Southwest	0.1299	**0.1289**	0.0746	**0.0740**	0.0168	**0.0149**
Northeast	0.1349	**0.1330**	0.0819	**0.0803**	-0.0010	**-0.0005**
Southeast	0.1316	**0.1308**	0.0766	**0.0765**	0.0067	**0.0058**

Table 26: Step 3 versus step 2 RMSE, MAE, and MB for the one-year accumulation period for the climatological subregions. The numbers in bold indicate an improvement in step 3 versus step 2.

in table 26. Like the three-year accumulation, every metric for every region shows improvement between step 2 and step 3. The plains region has the highest RMSE and MAE, which follows the pattern observed in the prior adjustments. The magnitudes of the adjustments are also more similar than in the three-year accumulation between the regions. Like in the three-year accumulation, the MB is negative in the north and positive in the south. The seasonal results are shown in table 27. In the winter, the RMSE and MAE for all of the regions and the MB in the northeast are improved. The plains region has the highest overall error by a wide margin and the southeast has the lowest. Interestingly, the positive biases are in the western and the negative biases are in the eastern regions. Similarly, the spring shows improvement for every RMSE and MAE. The MB is improved in two of the regions as well. The differences in the unadjusted and adjusted errors are similar across the regions, though the overall errors are much closer than in the winter. In the summer and fall, improvements occur in every region except for the northeast MB. The summer is the only season where the southwest,

Seasonal Step 3 vs Step 2 Errors in Climate Regions

Winter	Step 2 RMSE	Step 3 RMSE	Step 2 MAE	Step 3 MAE	Step 2 MB	Step 3 MB
Plains	0.2668	**0.2658**	0.1713	**0.1706**	0.0470	0.0481
Southwest	0.1770	**0.1764**	0.1044	**0.1039**	0.0091	0.0105
Northeast	0.1945	**0.1930**	0.1203	**0.1192**	-0.0075	**-0.0069**
Southeast	0.1649	**0.1636**	0.0957	**0.0947**	-0.0095	-0.0100
Spring	Step 2 RMSE	Step 3 RMSE	Step 2 MAE	Step 3 MAE	Step 2 MB	Step 3 MB
Plains	0.1779	**0.1760**	0.1048	**0.1037**	0.0045	**0.0043**
Southwest	0.1763	**0.1742**	0.0960	**0.0947**	0.0062	0.0069
Northeast	0.1588	**0.1566**	0.0967	**0.0943**	-0.0095	-0.0099
Southeast	0.1657	**0.1646**	0.0967	**0.0956**	0.0074	**0.0065**
Summer	Step 2 RMSE	Step 3 RMSE	Step 2 MAE	Step 3 MAE	Step 2 MB	Step 3 MB
Plains	0.1771	**0.1748**	0.1066	**0.1061**	-0.0147	**-0.0145**
Southwest	0.1822	**0.1792**	0.1098	**0.1088**	0.0306	**0.0279**
Northeast	0.1750	**0.1732**	0.1058	**0.1040**	0.0139	0.0144
Southeast	0.1839	**0.1831**	0.1135	**0.1134**	0.0244	**0.0241**
Fall	Step 2 RMSE	Step 3 RMSE	Step 2 MAE	Step 3 MAE	Step 2 MB	Step 3 MB
Plains	0.1876	**0.1861**	0.1093	**0.1085**	-0.0083	**-0.0074**
Southwest	0.1655	**0.1653**	0.0920	**0.0918**	0.0120	**0.0118**
Northeast	0.1549	**0.1526**	0.0949	**0.0932**	-0.0016	-0.0026
Southeast	0.1693	**0.1684**	0.0983	**0.0982**	0.0059	**0.0048**

Table 27: Step 3 versus step 2 RMSE, MAE, and MB for the seasonal accumulation periods for the climatological subregions. The numbers in bold indicate an improvement in step 3 versus step 2.

instead of the plains, has the highest overall error. Additionally, the plains region is negatively biased for both seasons and the only positive northeast bias is in the summer.

The mountainous versus the non-mountainous subregions are the final tested

Step 3 vs Step 2 Errors in Mountainous Region

Accumulation Period	Step 2 RMSE	Step 3 RMSE	Step 2 MAE	Step 3 MAE	Step 2 MB	Step 3 MB
Three-Year	0.1256	**0.1240**	0.0762	**0.0757**	-0.0071	**-0.0061**
One-Year	0.1413	**0.1398**	0.0854	**0.0837**	-0.0035	-0.0051
Winter	0.1938	**0.1918**	0.1227	**0.1203**	-0.0102	**-0.0101**
Spring	0.1676	**0.1660**	0.1013	**0.0991**	-0.0084	-0.0087
Summer	0.1797	**0.1783**	0.1112	**0.1103**	0.0055	**0.0042**
Fall	0.1653	**0.1637**	0.0990	**0.0988**	-0.0053	-0.0064
Season Avg.	0.1766	**01749**	0.1085	**0.1071**	-0.0046	-0.0053

Table 28: Step 3 versus step 2 RMSE, MAE, and MB for all accumulation periods for the mountainous region. The numbers in bold indicate an improvement in step 3 versus step 2.

subregion for the step 3 adjustments. Tables 28 and 29 show the results between step 2 and step 3 for all of the accumulation periods in the mountainous and non-mountainous

Step 3 vs Step 2 Errors in Non-Mountainous Region

Accumulation Period	Step 2 RMSE	Step 3 RMSE	Step 2 MAE	Step 3 MAE	Step 2 MB	Step 3 MB
Three-Year	0.1169	**0.1153**	0.0678	**0.0667**	0.0010	**0.0009**
One-Year	0.1337	**0.1322**	0.0780	**0.0771**	0.0041	**0.0039**
Winter	0.2085	**0.2076**	0.1168	**0.1156**	0.0085	0.0094
Spring	0.1715	**0.1696**	0.0965	**0.0959**	0.0025	0.0030
Summer	0.1806	**0.1784**	0.1074	**0.1069**	0.0132	**0.0124**
Fall	0.1733	**0.1722**	0.0976	**0.0971**	0.0012	**0.0011**
Season Avg.	0.1835	**0.1820**	0.1046	**0.1039**	0.0064	0.0065

Table 29: Step 3 versus step 2 RMSE, MAE, and MB for all accumulation periods for the non-mountainous region. The numbers in bold indicate an improvement in step 3 versus step 2.

regions respectively. The RMSE and MAE are improved for every accumulation in both regions, and the magnitudes of their improvements are similar. The adjustments and periods of improvement differ when considering the MB between the two regions. The three-year and summer MB are improved for both regions, and the improvements differ for the other time periods. Additionally, the MB is negative in the mountainous and positive in the non-mountainous regions for every accumulation. This pattern is also observed in the prior bias adjustments. Overall, the adjustments help improve the discontinuities in both the mountainous and non-mountainous regions. Remaining biases can be addressed in the two-dimensional adjustment.

The step 3 adjustments help remove the discontinuities caused by the beam blockage and mean field and range dependent bias adjustments. For every region and accumulation, the RMSE and MAE are improved between steps 2 and 3. The MB is

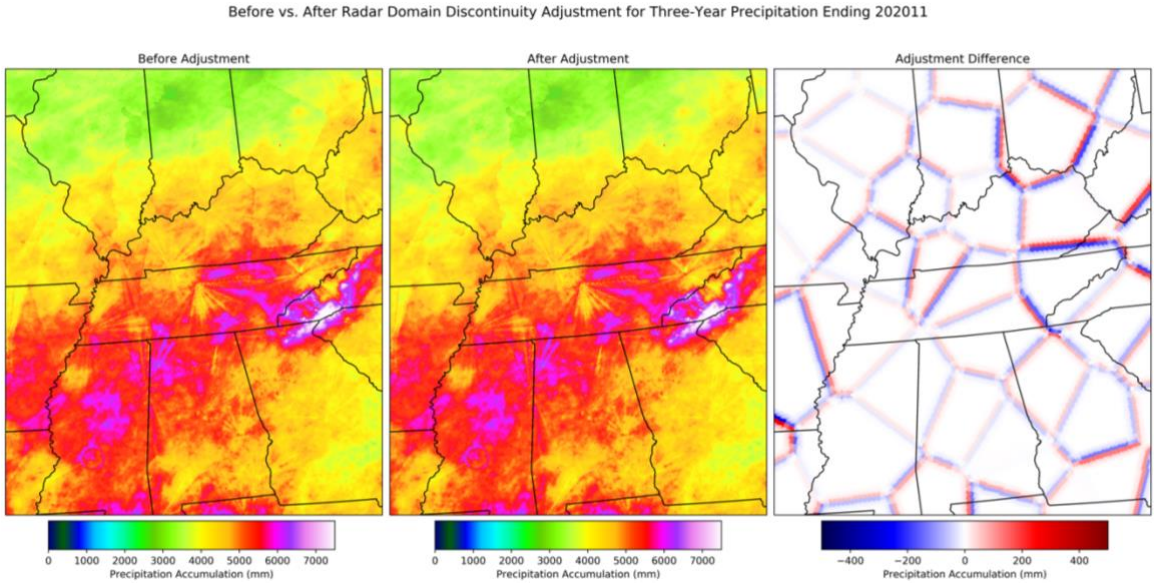

Figure 4.5: Radar estimates before step 3 (left), after step 3 (middle), and the magnitude of the adjustments for the three-year accumulation period through November 2020.

improved in most cases as well. Variations in the magnitude of the improvements between the seasons can be attributed to the adjustments being calculated independently for each accumulation. Figure 4.5 shows an example of the three-year precipitation estimates through November 2020 before and after the step 3 adjustment, and the magnitude of each adjustment. The before adjustment image on the left has multiple prominent radar domain discontinuities. One example is the discontinuity in southeast

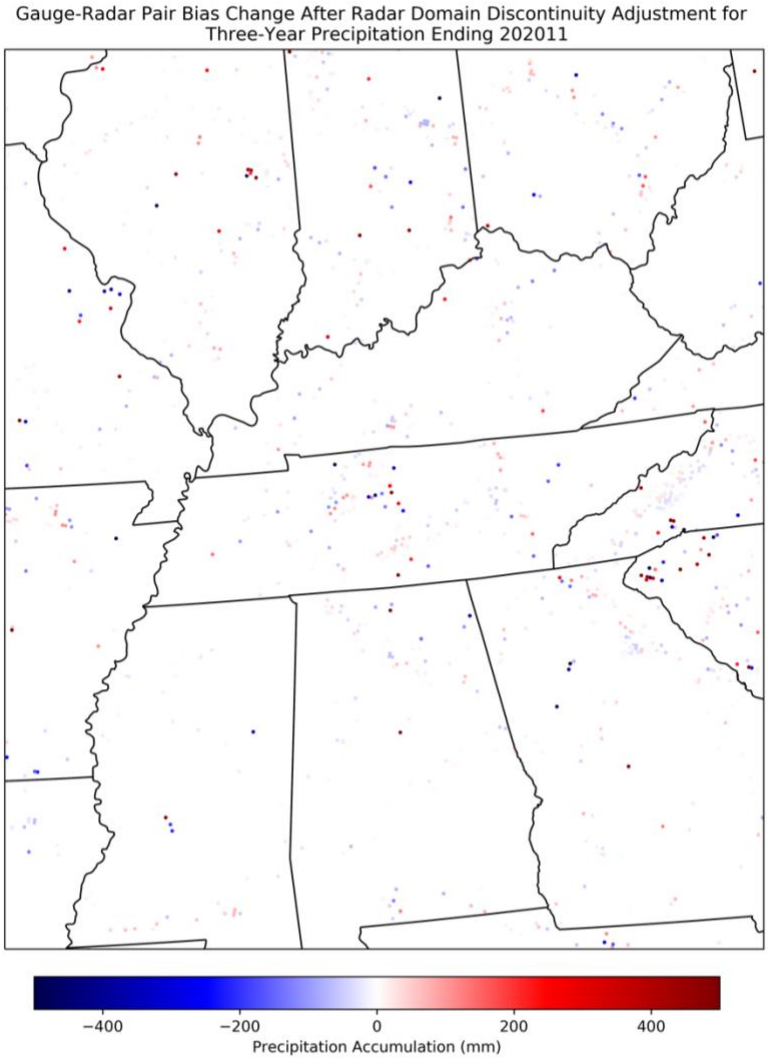

Figure 4.6: The difference between the gauge-radar bias after and before step 3 for three-year precipitation through November 2020.

Ohio in the southwest to northeast direction. The middle image, which shows the precipitation estimates after the step 3 adjustment, no longer shows prominent discontinuities, including the one in Ohio. The right image shows how the adjustments work. The side that is less than the other is increased and the side that is greater than the other is decreased at an inverse distance. The adjustments are greatest near the boundary, then they quickly fade away at longer distances. The process shown in the right image is done for every accumulation and every radar domain boundary within the study area. Finally, figure 4.6 shows how the difference after and before the step 3 adjustment in the gauge-radar pair biases. It can be seen that the adjustments follow radar domain boundaries, where the greatest changes in the negative direction are next to the greatest changes in the positive direction. Additionally, the magnitude of the adjustments is generally smaller than those in the previous steps. The remaining biases in the Stage IV precipitation dataset will be addressed for the precipitation field as whole by incorporating gauge data through interpolation.

4.4. Two-Dimensional Adjustment Results

The calculations and results of the performance of step 4, the two-dimensional adjustments, are discussed in this section. First, the three interpolation methods, IDW, IDW2, and OK are tested in order to determine the optimally performing method to use in the step 4 algorithm. Then, using the optimal interpolation method, the step 4 algorithm is tested for whether the adjustments should be calculated independently for each accumulation period or aggregated from the seasonal accumulations. Finally, the

step 4 algorithm is run for the full study region using the optimal accumulation period calculations. Data withholding is used for all of the analyses in this section, which are the determination of the optimal interpolation method, testing of the accumulation period calculations, and performance of the step 4 algorithm because rain gauges are used in the adjustments. This section will first show the results of the interpolation method errors, and the results of the accumulation period calculations. Then, the performance of the step 4 adjustments will be discussed for the full study area, climatological subregions, mountainous versus non-mountainous regions, and the different effective observation density classes for the three-year, one-year, and seasonal accumulations.

The performance of the IDW, IDW2, and OK interpolation methods are discussed. Because these methods consider the values of the closest points in the estimation of a given grid point, the number of neighboring points to use in the calculations are tested. The selected points are from 10 to 50 in increments of 10. The range of values of 10 to 50 are used because they are the least computationally expensive

RMSE for Interpolation Methods Using Different Numbers of Points

RMSE	IDW	IDW2	OK Exponential	OK Spherical	OK Linear	OK Hole-Effect	OK Power
10 points	0.1121	0.1177	0.1145	0.1146	0.1149	0.1146	0.1149
20 points	0.1104	0.1157	**0.1140**	**0.1142**	**0.1145**	**0.1141**	**0.1141**
30 points	0.1099	0.1150	**0.1140**	0.1149	0.1149	0.1142	0.1142
40 points	**0.1097**	0.1145	0.1142	0.1153	0.1153	0.1144	0.1144
50 points	0.1098	**0.1143**	0.1144	0.1156	0.1156	0.1145	0.1145

Table 30: RMSE for each tested interpolation method and model for 10 to 50 points for the three-year accumulation period. Bolded numbers indicate the lowest error in each column.

and the increments of 10 are used because there are small differences in error in shorter increments. Additionally, there are five different OK models tested: exponential, spherical, linear, hole-effect, and power. Tables 30, 31, and 32 show the RMSE, MAE, and MB respectively for the interpolation methods using between 10 and 50 points for

MAE for Interpolation Methods Using Different Numbers of Points

MAE	IDW	IDW2	OK Exponential	OK Spherical	OK Linear	OK Hole-Effect	OK Power
10 points	0.0653	0.0667	0.0657	**0.0656**	**0.0657**	**0.0655**	**0.0657**
20 points	0.0645	0.0658	0.0657	0.0661	0.0664	0.0659	0.0664
30 points	0.0646	**0.0652**	0.0658	0.0664	0.0667	0.0658	0.0667
40 points	**0.0644**	0.0654	**0.0654**	0.0661	0.0665	0.0657	0.0665
50 points	0.0645	0.0654	0.0658	0.0665	0.0668	0.0659	0.0667

Table 31: MAE for each tested interpolation method and model for 10 to 50 points for the three-year accumulation period. Bolded numbers indicate the lowest error in each column.

the three-year accumulation. Only the three-year accumulation is shown because the results are similar for the other accumulations. Table 30 shows that IDW has the lowest

MB for Interpolation Methods Using Different Numbers of Points

MB	IDW	IDW2	OK Exponential	OK Spherical	OK Linear	OK Hole-Effect	OK Power
10 points	**-0.0107**	**-0.0086**	-0.0081	-0.0081	-0.0081	-0.0078	-0.0081
20 points	-0.0114	-0.0093	-0.0088	-0.0086	-0.0085	-0.0084	-0.0086
30 points	-0.0117	-0.0095	-0.0083	-0.0082	-0.0083	-0.0083	-0.0083
40 points	-0.0114	-0.0097	-0.0081	-0.0078	-0.0082	-0.0082	-0.0081
50 points	-0.0115	-0.0098	**-0.0076**	**-0.0077**	**-0.0076**	**-0.0076**	**-0.0076**

Table 32: MB for each tested interpolation method and model for 10 to 50 points for the three-year accumulation period. Bolded numbers indicate the bias closest to zero.

overall error of all the tested methods. Additionally, the use of 40 points in the calculations perform the best. The IDW results are similar in table 31. It has the lowest overall MAE, and its optimal number of points is 40. The OK models, however, each perform best using 10 points rather than 20 points when considering table 31 versus table 30. Finally, the results in table 32 show that OK is the least biased interpolation method and IDW is the most biased. Each OK model has nearly the same MB as well. When considering all three metrics, IDW using 40 points has the lowest RMSE and MAE, but OK using 50-points is the least biased. The least biased IDW calculation uses only 10 points. With the given information, the selected interpolation method for the step 4 adjustments is IDW using 30-points. This method is selected because it is a compromise of having low overall error while minimizing the effects of the greater negative bias. Additionally, the computational expense of IDW is much less than OK.

The step 4 accumulation periods are tested using two different methods. The first method applies the adjustment to the seasonal accumulations, then aggregates them to obtain the one-year and three-year accumulations. The second method calculates and applies the adjustments independently for each accumulation. In the first method, the gauge-radar biases are calculated for every seasonal accumulation, then they are interpolated to the entire study area and added to the existing radar estimates. The one-year and three-year accumulations are obtained by summing the results of the seasonal analyses. For the second method, the gauge-radar biases and interpolations are performed independently for each accumulation. The one-year and three-year accumulations are obtained using the results of their respective interpolations rather than

the sum of the seasonal analyses. The goal of testing the accumulation periods is to determine the best performing method. The reduction of computational expense is a benefit of aggregating the shortest accumulations to obtain the longer accumulation estimates. Table 33 shows the results between the aggregated and independently run data

Aggregated vs Independent Two-Dimensional Analysis Errors

Accumulation Period	Aggregated RMSE	Independent RMSE	Aggregated MAE	Independent MAE	Aggregated MB	Independent MB
Three-Year	0.1170	0.0800	0.0371	0.0358	0.0004	-0.0100
One-Year	0.1031	0.0807	0.0389	0.0384	-0.0029	-0.0077

Table 33: RMSE, MAE, and MB of the aggregated versus independent accumulation period estimates for the three-year and one-year periods.

for the three-year and one-year accumulations. For both accumulations, the independently run adjustments have a lower RMSE and MAE compared to the aggregated calculations. The aggregated calculations, however, are less biased overall. Because the error reductions are larger than the increase in bias, the independent accumulation period calculations are used for the step 4 adjustment.

The step 4 adjustments are performed using the data from step 3 as the input, and the interpolation method is IDW using 30 points. The results between steps 3 and 4 for the full study area at all accumulations is shown in table 34. It can be seen that the RMSE and MAE are reduced for all accumulations. The magnitudes of the improvements are also greater than those prior to this step. When considering the bias, it becomes more negative for every accumulation period. This is likely due to the gauge data being incorporated. The only MB that moves closer to zero occurs in the summer. However, the magnitude by which the bias moves further away from zero is generally

Step 4 vs Step 3 Errors for the Full Study Area

Accumulation Period	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
Three-Year	0.1165	**0.1005**	0.0687	**0.0561**	-0.0035	-0.0079
One-Year	0.1338	**0.1162**	0.0788	**0.0654**	0.0032	-0.0076
Winter	0.2064	**0.1801**	0.1158	**0.0960**	0.0025	-0.0086
Spring	0.1684	**0.1464**	0.0966	**0.0794**	0.0007	-0.0039
Summer	0.1787	**0.1565**	0.1075	**0.0886**	0.0138	**-0.0025**
Fall	0.1684	**0.1454**	0.0965	**0.0770**	-0.0023	-0.0047
Season Avg.	0.1805	**0.1571**	0.1036	**0.0852**	0.0037	-0.0049

Table 34: Step 4 versus step 3 RMSE, MAE, and MB for all accumulation periods for the full study area. The bolded numbers indicate an improvement in step 4 versus step 3.

small. Overall, the decreases in error are large across the full study area.

The performance of the step 4 adjustments within the different climatological subregions are discussed. The results comparing steps 3 and 4 for the three-year accumulation for the climatological subregions is shown in table 35. Like in the full

Three-Year Step 4 vs Step 3 Errors in Climate Regions

Three-Year	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
Plains	0.1222	**0.0982**	0.0755	**0.0534**	-0.0111	**-0.0090**
Southwest	0.1086	**0.0951**	0.0638	**0.0536**	0.0034	-0.0097
Northeast	0.1175	**0.1057**	0.0688	**0.0581**	-0.0022	-0.0049
Southeast	0.1120	**0.1006**	0.0647	**0.0601**	0.0039	-0.0090

Table 35: Step 4 versus step 3 RMSE, MAE, and MB for the three-year accumulation period for the climatological subregions. The bolded numbers indicate an improvement step 4 versus step 3.

study area results, the RMSE and MAE are both substantially reduced for each region. However, a difference is that the MB increases for the plains, which is the only instance where the MB becomes more positive after the adjustment. The plains, like in the prior results, has the largest improvement. Additionally, it has the lowest MAE following the adjustment. For the one-year climatological subregion results in table 36,

One-Year Step 4 vs Step 3 Errors in Climate Regions

One-Year	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
Plains	0.1384	**0.1140**	0.0836	**0.0650**	-0.0038	-0.0090
Southwest	0.1293	**0.1140**	0.0763	**0.0642**	0.0108	**-0.0070**
Northeast	0.1334	**0.1188**	0.0795	**0.0679**	0.0005	**-0.0052**
Southeast	0.1293	**0.1158**	0.0756	**0.0665**	0.0084	-0.0085

Table 36: Step 4 versus step 3 RMSE, MAE, and MB for the one-year accumulation period for the climatological subregions. The bolded numbers indicate an improvement in step 4 versus step 3.

the RMSE and MAE improvements behave similarly as those in table 36, but the MB changes are different. The MB in every region becomes more negative after the adjustments, and the southwest and northeast see improvement. The seasonal results are shown in table 37. In the winter, the plains have the greatest improvement in error and MB. In the spring, the southwest has the highest overall error instead of the plains. However, the plains still benefit the greatest from the adjustments. Additionally, the MB in the northeast increases, which goes against the pattern seen in the other accumulations. Every metric shows improvement in the summer except for the plains MB. This season also has the most positively biased unadjusted data. Finally, the fall

Seasonal Step 4 vs Step 3 Errors in Climate Regions

Winter	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
Plains	0.2664	**0.2262**	0.1659	**0.1283**	0.0284	**-0.0073**
Southwest	0.1761	**0.1498**	0.1043	**0.0803**	0.0064	-0.0124
Northeast	0.1888	**0.1697**	0.1113	**0.0985**	-0.0071	-0.0076
Southeast	0.1672	**0.1538**	0.0962	**0.0834**	-0.0072	**-0.0052**
Spring	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
Plains	0.1709	**0.1418**	0.1092	**0.0788**	0.0058	**-0.0026**
Southwest	0.1761	**0.1578**	0.0969	**0.0843**	0.0075	-0.0044
Northeast	0.1585	**0.1364**	0.0954	**0.0759**	-0.0105	**-0.0038**
Southeast	0.1626	**0.1478**	0.0969	**0.0847**	0.0007	-0.0096
Summer	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
Plains	0.1731	**0.1451**	0.1054	**0.0828**	-0.0032	-0.0038
Southwest	0.1809	**0.1633**	0.1102	**0.0924**	0.0236	**-0.0022**
Northeast	0.1725	**0.1513**	0.1054	**0.0888**	0.0184	**0.0005**
Southeast	0.1861	**0.1677**	0.1154	**0.0993**	0.0263	**-0.0059**
Fall	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
Plains	0.1780	**0.1452**	0.1034	**0.0778**	-0.0080	**-0.0060**
Southwest	0.1611	**0.1451**	0.0911	**0.0790**	0.0086	**0.0006**
Northeast	0.1524	**0.1321**	0.0915	**0.0737**	-0.0057	-0.0059
Southeast	0.1724	**0.1563**	0.0962	**0.0838**	0.0008	-0.0066

Table 37: Step 4 versus step 3 RMSE, MAE, and MB for the seasonal accumulation periods for the climatological subregions. The bolded numbers indicate an improvement in step 4 versus step 3.

adjustments behave similarly as those in the winter and spring. When considering the adjustments overall, the magnitudes of improvement for the RMSE and MAE increase as

the length of the accumulation period decreases. The bias also becomes more negative in nearly every accumulation period and region.

The results for the performance of the step 4 adjustments for the mountainous versus the non-mountainous regions are discussed. Tables 38 and 39 show the results of

Step 4 vs Step 3 Errors in Mountainous Region

Accumulation Period	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
Three-Year	0.1210	**0.1083**	0.0709	**0.0611**	-0.0080	**-0.0052**
One-Year	0.1370	**0.1208**	0.0841	**0.0695**	-0.0015	-0.0052
Winter	0.1892	**0.1715**	0.1157	**0.0995**	-0.0180	**-0.0062**
Spring	0.1617	**0.1431**	0.0966	**0.0845**	-0.0109	**-0.0034**
Summer	0.1808	**0.1602**	0.1086	**0.0967**	0.0119	**-0.0045**
Fall	0.1616	**0.1420**	0.0938	**0.0808**	-0.0105	-0.0108
Season Avg.	0.1733	**0.1542**	0.1037	**0.0904**	-0.0069	-0.0062

Table 38: Step 4 versus step 3 RMSE, MAE, and MB for all accumulation periods for the mountainous region. The numbers in bold indicate an improvement in step 4 versus step 3.

the mountainous and non-mountainous step 3 versus step 4 data for all accumulation periods respectively. The RMSE and MAE are reduced for every accumulation period for both regions. The error is higher in the mountainous regions for the three-year, one-year, and summer accumulations, and the remaining seasons have higher errors in the non-mountainous regions. When considering MB, the mountainous values are more negative than the non-mountainous values. The adjustments in the full study area and the climatological subregions lead to more negatively biased data, but this is not the case for the three-year, winter, and summer accumulations in the mountainous region. The non-

Step 4 vs Step 3 Errors in Non-Mountainous Region

Accumulation Period	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
Three-Year	0.1158	**0.0991**	0.0683	**0.0554**	-0.0027	-0.0084
One-Year	0.1332	**0.1153**	0.0781	**0.0649**	0.0041	-0.0081
Winter	0.2091	**0.1815**	0.1160	**0.0958**	0.0056	-0.0093
Spring	0.1691	**0.1467**	0.0968	**0.0789**	0.0028	-0.0042
Summer	0.1781	**0.1556**	0.1071	**0.0874**	0.0138	**-0.0022**
Fall	0.1692	**0.1457**	0.0949	**0.0765**	-0.0011	-0.0039
Season Avg.	0.1814	**0.1574**	0.1037	**0.0847**	0.0053	**-0.0049**

Table 39: Step 4 versus step 3 RMSE, MAE, and MB for all accumulation periods for the non-mountainous region. The numbers in bold indicate an improvement in step 4 versus step 3.

mountainous MB, however, becomes more negatively biased in every accumulation following the adjustments. Overall, the magnitudes of the adjustments between the two regions are similar. The errors between the two regions following all of the adjustments are closer in value compared to the raw Stage IV data as well.

The results for the performance of the step 4 adjustments within the effective observation density subregions are discussed. It is expected that the regions having a higher gauge density will have the lowest error. Tables 40, 41, and 42 show the results of the step 3 versus step 4 adjustments within the effective observation density subregions for the three-year, one-year, and seasonal accumulations respectively. Like in the other regions, every RMSE and MAE is improved after the adjustment. Additionally, the MB is improved in the first and second class, and for all classes in the summer. The overall

Three-Year Step 4 vs Step 3 Errors in Gauge Density Regions

Three-Year Effective Obs. Density	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
0.0-12.0	0.1447	**0.1151**	0.0986	**0.0685**	0.0076	**-0.0026**
12.01-16.00	0.1176	**0.1028**	0.0700	**0.0594**	0.0037	**-0.0028**
16.01-28.00	0.1136	**0.0997**	0.0673	**0.0567**	-0.0039	-0.0078
28.01-64.0	0.1115	**0.0972**	0.0653	**0.0528**	-0.0081	-0.0118
64.01+	0.1118	**0.0944**	0.0674	**0.0536**	-0.0046	-0.0150

Table 40: Step 4 versus step 3 RMSE, MAE, and MB for the three-year accumulation period for the effective observation density subregions. The bolded numbers indicate an improvement in step 4 versus step 3.

errors tend to decrease as the effective observation density increases, however, there are instances where the fourth class performs better than the fifth. The overall bias tends to

One-Year Step 4 vs Step 3 Errors in Gauge Density Regions

One-Year Effective Obs. Density	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
0.0-12.0	0.1678	**0.1380**	0.1058	**0.0806**	0.0135	**-0.0079**
12.01-16.00	0.1364	**0.1202**	0.0795	**0.0670**	0.0141	**-0.0041**
16.01-28.00	0.1324	**0.1157**	0.0792	**0.0665**	0.0026	-0.0060
28.01-64.0	0.1238	**0.1090**	0.0730	**0.0616**	0.0006	-0.0087
64.01+	0.1311	**0.1111**	0.0784	**0.0608**	0.0043	-0.0094

Table 41: Step 4 versus step 3 RMSE, MAE, and MB for the one-year accumulation period for the effective observation density subregions. The bolded numbers indicate an improvement in step 4 versus step 3.

become more negative as the effective observation density increases. This is expected because gauges are negatively biased compared to radar, so the incorporation of more

Seasonal Step 4 vs Step 3 Errors in Gauge Density Regions

Winter	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
0.0-12.0	0.2908	**0.2478**	0.1820	**0.1420**	0.0265	**0.0033**
12.01-16.00	0.2159	**0.1800**	0.1299	**0.1054**	0.0399	**0.0067**
16.01-28.00	0.2043	**0.1837**	0.1167	**0.1019**	0.0040	-0.0050
28.01-64.0	0.1878	**0.1625**	0.1071	**0.0893**	-0.0002	-0.0110
64.01+	0.1749	**0.1487**	0.1099	**0.0802**	0.0046	-0.0106
Spring	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
0.0-12.0	0.2159	**0.1800**	0.1299	**0.1054**	0.0399	**0.0067**
12.01-16.00	0.1788	**0.1590**	0.1013	**0.0878**	0.0162	**0.0016**
16.01-28.00	0.1673	**0.1465**	0.0962	**0.0800**	0.0011	-0.0027
28.01-64.0	0.1538	**0.1339**	0.0916	**0.0745**	-0.0060	-0.0088
64.01+	0.1632	**0.1379**	0.0951	**0.0711**	-0.0040	-0.0070
Summer	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
0.0-12.0	0.1999	**0.1744**	0.1253	**0.1031**	-0.0003	**-0.0065**
12.01-16.00	0.1859	**0.1666**	0.1127	**0.0989**	0.0226	**0.0109**
16.01-28.00	0.1792	**0.1546**	0.1078	**0.0877**	0.0134	**-0.0025**
28.01-64.0	0.1711	**0.1508**	0.1027	**0.0851**	0.0117	**-0.0048**
64.01+	0.1740	**0.1506**	0.1049	**0.0833**	0.0225	**0.0011**
Fall	Step 3 RMSE	Step 4 RMSE	Step 3 MAE	Step 4 MAE	Step 3 MB	Step 4 MB
0.0-12.0	0.2106	**0.1762**	0.1265	**0.0944**	0.0096	**-0.0031**
12.01-16.00	0.1694	**0.1485**	0.0967	**0.0854**	0.0130	**0.0028**
16.01-28.00	0.1632	**0.1430**	0.0947	**0.0772**	-0.0032	-0.0043
28.01-64.0	0.1602	**0.1378**	0.0893	**0.0734**	-0.0063	-0.0067
64.01+	0.1566	**0.1318**	0.0892	**0.0718**	-0.0014	-0.0094

Table 42: Step 4 versus step 3 seasonal RMSE, MAE, and MB for the effective observation density subregions. The bolded numbers indicate an improvement in step 4.

gauges will make the bias more negative. For every accumulation period, the first class benefits the most from the adjustments. This pattern is also observed with the step 2 results. One difference between the step 4 versus the step 2 adjustments is the magnitude of the error reduction as the effective observation density increases. In the step 4 adjustments, the magnitude of the improvements remains roughly the same in classes two through five while the magnitudes decrease as the classes increase in the step 2 adjustments. Finally, similar seasonal variations persist, where the winter has the highest

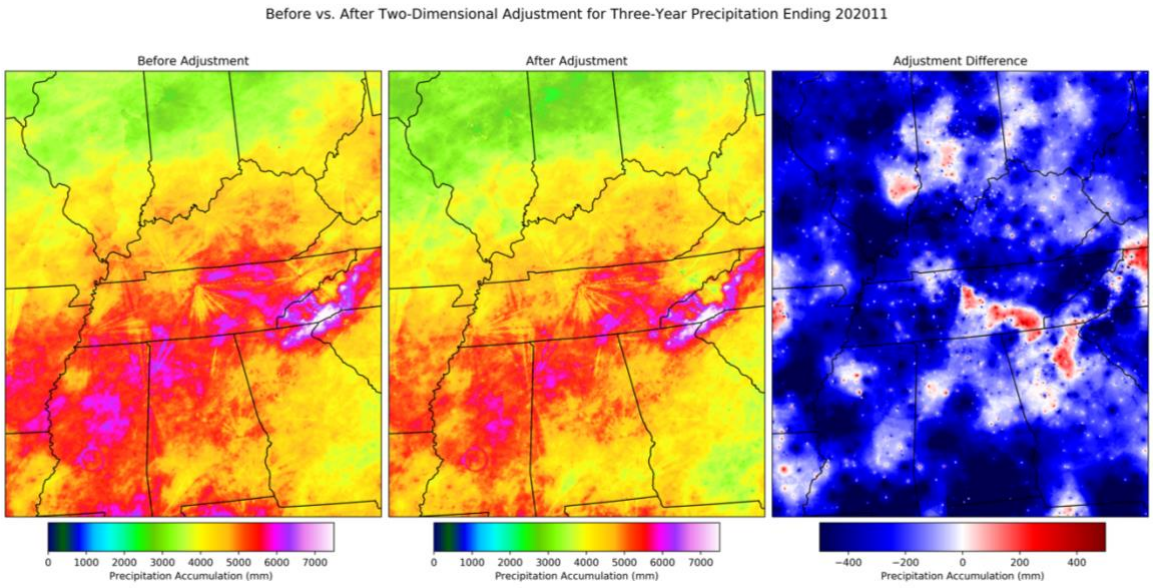

Figure 4.7: Radar estimates before step 4 (left), after step 4 (middle), and the magnitude of the adjustments for the three-year accumulation period through November 2020.

error, the spring has the lowest, and the summer and fall behave similarly.

The step 4 adjustments led to large improvements over the input data for all study regions. The selected interpolation method is less computationally expensive compared to OK, and it has a lower RMSE and MAE, but a more negative bias. The

decreasing MB in most adjustments reflects that feature of the IDW, but the greater error improvements over OK makes the IDW selection justifiable. Figure 4.7 shows the three-

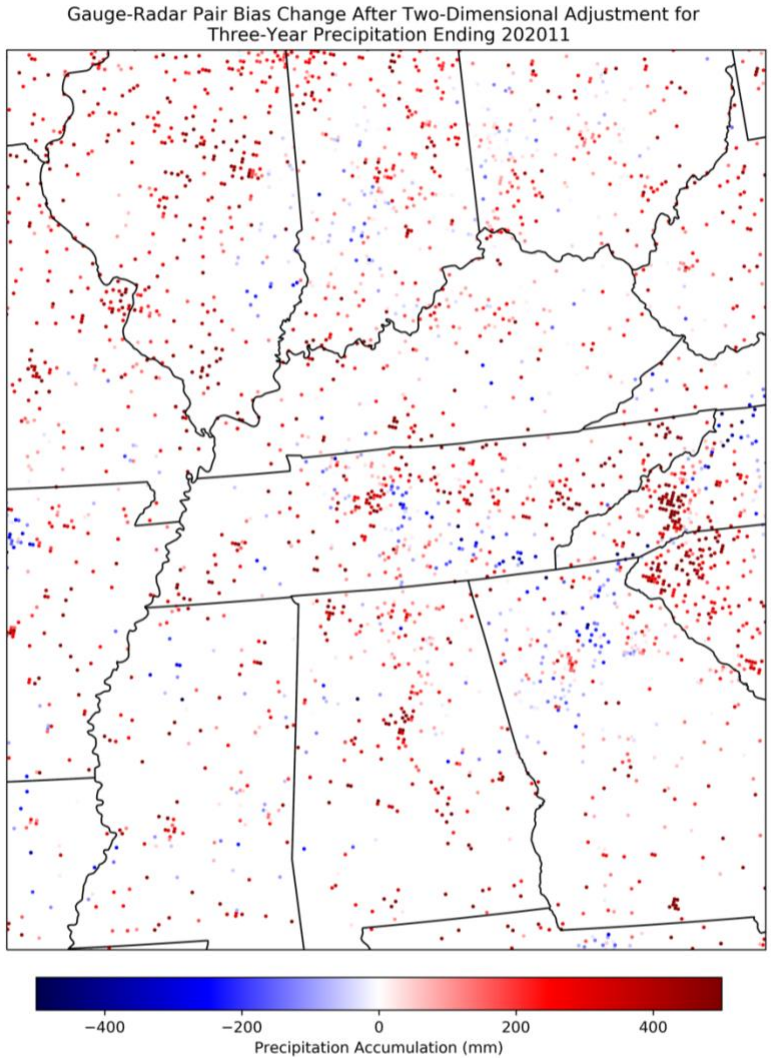

Figure 4.8: The difference between the gauge-radar bias after and before step 4 for three-year precipitation through November 2020.

year radar estimated precipitation images through November 2020 for before and after step 4, and the magnitude of the adjustments. The adjustment difference image shows that step 4 generally decreased the radar estimates. This is seen in the results by the decreasing biases after the adjustment. Additionally, the adjustments are large compared

to the prior steps. Finally, figure 4.8 shows the gauge-radar pair bias differences between the adjustment image in figure 4.7, the gauge-radar pairs generally have decreasing radar values. The output data after the step 4 adjustment is an improved Stage IV dataset over the raw data.

5. CONCLUSIONS

This study introduces and evaluates a revised methodology for adjusting biases in Stage IV precipitation data for the purpose of drought monitoring in the U.S. east of the Rocky Mountains based on work done by McRoberts and Nielsen-Gammon (McRoberts and Nielsen-Gammon 2017, McRoberts 2014). First, a new method that corrects for beam blockages is introduced using image filtering and ridge detection. The aim of this was to improve upon potential flagging errors in the original adjustment method. Next, a mean field and range dependent bias adjustment methodology based upon seasonal calculations and gauge accumulations is discussed. Then, discontinuities that develop at the edges of each radar domain due to the first two steps are adjusted using an IDW method. Finally, after extensive testing of IDW, IDW2, and OK, the remaining two-dimensional biases are adjusted using interpolated gauge-radar bias data. The performance of each step of the bias adjustment methodology is assessed by calculating errors of the Stage IV estimates relative to reliable gauge data for the full study region, subregions based on similar climatological characteristics, mountainous versus non-mountainous areas, and for the mean field and range dependent and two-dimensional analyses, subregions based on the effective observation density of gauges.

The four hypotheses that are tested in this study are discussed in section 2. In short, they are (1) QPE data will be improved with an optimally selected interpolation method, (2) OK will be the best performing interpolation method, (3) regions with low gauge density and high topographic relief will benefit the most from the adjustments, and (4) adjustments using gauge-based adjustments will perform better than those only

using radar. The results of each one is discussed in the following paragraphs.

Hypothesis 1 is tested in the two-dimensional step of the bias adjustment methodology by calculating the RMSE, MAE, and MB for the adjusted data versus the radar domain discontinuity adjusted data. Every single RMSE and MAE show improvement after the two-dimensional adjustment, and many of the biases are reduced as well. Hypothesis 1 is supported when considering the consistent decreases in error. When comparing the raw Stage IV data to the two-dimensional adjusted data with the optimally selected interpolation method, there is a large improvement. For example, in the adjusted three-year accumulation period in table 34 versus the original data in table 3, there is a decrease in the PoN RMSE by 0.0376, a decrease in MAE by 0.0364, and the MB moves closer to zero by 0.0108. The magnitudes of the improvements are similar for the other tested accumulation periods as well. For this, hypothesis 1 is supported because the QPE data is improved when considering every metric that is used in this study. The results in tables 30, 31, and 32 suggest that the other tested interpolation methods could have improved the data as well because the errors are close in value to those of the optimal IDW method. However, the greatest improvement in the QPE data is achieved through use of the IDW with 30 points.

Hypothesis 2 is tested using data withholding tests for IDW, IDW2, and OK and its different models for varying numbers of points to consider in the calculations. It was believed that OK would perform the best out of the three interpolation methods because of its ability to account for clustering and to take directional bias into consideration (Gentile et al. 2013). However, the results in tables 30, 31, and 32 generally do not

support OK being the best performing method. When considering the RMSE and MAE in tables 30 and 31 respectively, hypothesis 2 is not supported because IDW consistently performs better than OK. However, hypothesis 2 can be partially supported because OK is consistently less biased than IDW and IDW2 as shown in table 32. In addition to the slightly higher errors seen in OK versus IDW, the other drawback is the computational expense of OK. For this reason, OK is not the optimal interpolation method for the two-dimensional adjustment in this study.

Hypothesis 3 is tested using the different subregions for this study. The first part of it, gauge density, is only tested in the mean field and range dependent and two-dimensional adjustment steps because these are the only adjustments that use gauge data. In the mean field and range dependent effective observation density subregion tests, hypothesis 3 is supported because the regions with the least number of gauges have the largest improvement. For the same test in the two-dimensional adjustments, the results follow a similar pattern, though, the magnitudes of improvement are greater across all regions compared to the mean field and range dependent adjustment tests. The topographic relief part of hypothesis 3 is tested in the mountainous versus non-mountainous results for each step of the bias adjustment methodology. The greatest improvements in both error and bias for all steps are generally seen in the non-mountainous region, which does not support hypothesis 3. However, there are still consistent improvements within the mountainous region for all steps of the bias adjustment methodology.

In addition to the mountainous versus non-mountainous and gauge effective observation density subregions, there are variations with the performance of the bias adjustment methodology within the climatological subregions. For most of the tests, the plains subregion has the largest error and a positive bias. This region also has the least number of gauges. Steps 2 and 4, the gauge-based correction steps, helped substantially decrease the errors and reduce the bias in this region. Compared to the other subregions, the plains benefitted the most from the adjustments. The northeast subregion also has some of the highest errors out of the subregions. Like the plains, its highest errors are seen in the warmer months. These two regions experience more frozen precipitation than the southwest and southeast subregions. The southern regions have their highest errors in the warmer months, which is opposite from the northern ones. The variations in climate features like precipitation type and intensity can cause errors to differ by region.

Hypothesis 4 is tested by calculating the error metrics for each step of the bias adjustment methodology using the same gauges. The radar-based adjustments are performed in steps 1 and 3, and the gauge-based adjustments are performed in steps 2 and 4. Figure 5.1 shows the RMSE, MAE, and MB for each step of the bias adjustment methodology for all three accumulation periods. The gauges used for comparison are constant for each calculation. The selected gauges are withheld gauges used in the testing of steps 2 and 4. In the RMSE and MAE graphs, there is a decrease in error after each step. However, the smallest changes occur after steps 1 and 3, and the largest occur after steps 2 and 4. This helps support hypothesis 4 because the gauge-based adjustments have the largest decreases in error in these calculations. When considering the MB, the

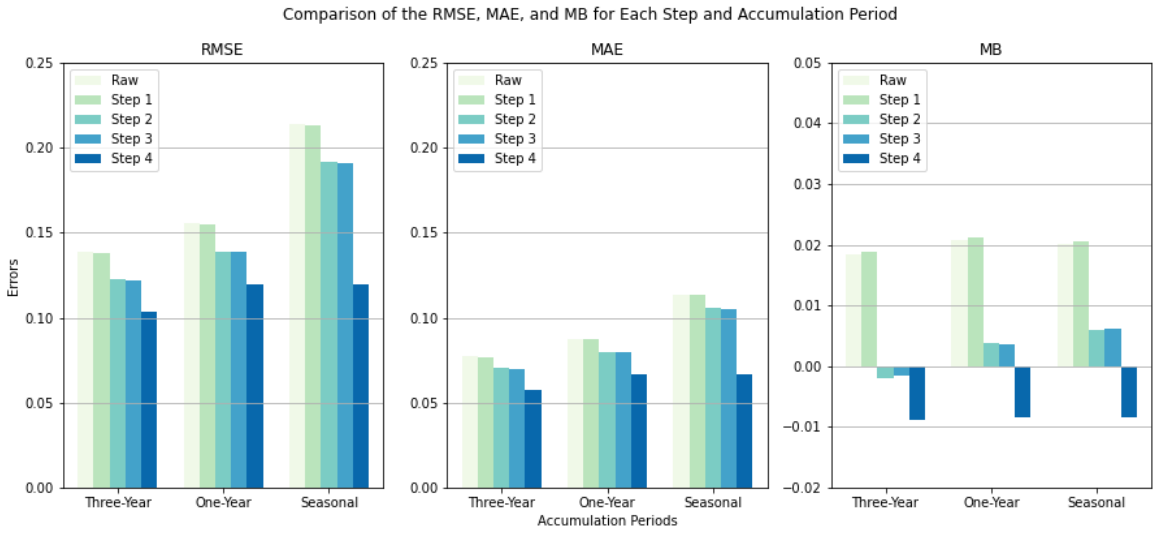

Figure 5.1: RMSE, MAE, and MB for each step and accumulation period for the full study period and area. The gauges used for comparison are the same for each calculation.

results are mixed. Every MB moves further from 0 after step 1 is applied. However, only the seasonal MB moves further from zero after step 3. For the gauge-based adjustments, the largest improvement in MB occurs after step 2 is applied. However, the step 4 adjustment results in the largest movement away from 0, but the data is less biased than the raw data. For this, hypothesis 4 is partially supported. The gauge-based improvements result in the largest decreases in error, and overall, a decrease in bias.

The 4-step bias adjustment methodology presented in this study can be useful to drought monitoring applications because it helps to improve the accuracy of long-term Stage IV radar estimated precipitation. The performance of the adjusted data for the full study area is much improved over the raw Stage IV data. For the smaller subregions tested in this study, variations in performance exist. For decision makers, it is important to consider the region and accumulation period when using the bias adjusted dataset

because the estimates perform differently in certain situations. Despite the variations in performance that exist, the bias adjusted dataset performs better in every tested circumstance than the raw dataset.

Future work for this study can address shortcomings and expand upon the region. One main shortcoming in the algorithm is with wide beam blockages. Overall, beam blockage adjustments result in improvements, but the wider beam blockages still faintly appear in the radar estimated precipitation images. Another shortcoming that can be improved upon is the negative biases after the two-dimensional adjustment. The testing and incorporation of different interpolation methods can help address this issue because IDW is negatively biased in this study. Future work can also aim to expand the extent of the study area. The current size is large, but much of the country is left out and many of the western regions are prone to severe drought. Finally, future work should rerun the entire an analysis using the PRISM normals from 1991-2020 in order to account for some of the effects of climate change.

REFERENCES

AghaKouchak A., A. Farahmand., F. S. Melton, J. Teixeira., M. C. Anderson, B. D. Wardlow, C. R. Hain, 2015: Remote sensing of drought: Progress, challenges and opportunities. *Rev. Geophys,* **53,** 452 -480, https://doi.org/10.1002/2014RG000456.

AMS, 2019: drought. Accessed 4 June 2021, https://glossary.ametsoc.org/wiki/Drought.

Anagnostou, E. N., W. F. Krajewski, D. Seo, E. R. Johnson, 1998: Mean-Field rainfall bias studies for WSR-88D. *J. Hydrol. Eng.,* **3,** 149-159, https://doi.org/10.1061/(ASCE)1084-0699(1998)3:3(149).

Baeck, M. L., and J. A. Smith, 1998: Rainfall estimation by the WSR-88D for heavy rainfall events. Wea. Forecasting, **13**, 416-436, https://doi.org/10.1175/1520-0434(1998)013<0416:REBTWF>2.0.CO;2.

Blankenship C. B., J. L. Case, B. T. Zavodsky, W. L. Crosson, 2016: Assimilation of SMOS Retrievals in the Land Information System. *IEEE Trans. Geosci. Electron.,* **54,** 6320 -6332, https://doi.org/10.1109/TGRS.2016.2579604.

Bech, J., B. Codina, J. Lorente, D. Bebbington, 2003: The Sensitivity of Single Polarization Weather Radar Beam Blockage Correction to Variability in the Vertical Refractivity Gradient. *J. Atmos. Oceanic. Technol.* **20.6,** 845-855, https://doi.org/10.1175/1520-0426(2003)020<0845:TSOSPW>2.0.CO;2.

Case, J. L., S. V. Kumar, R. J. Kuligowski, and C. Langston, 2013: Comparison of four precipitation forcing datasets in Land Information System simulations over the Continental U.S. Preprints. *27th Conf. on Hydrology*, Austin, TX, Amer. Meteor. Soc., P69.

Cecinati F., A. M. Moreno-Ródenas, M. A. Rico-Ramirez, M. T. Veldhuis, J. G. Langeveld, 2017: Considering rain gauge uncertainty using kriging for uncertain data. *J. Atmo.,* **9,** 446, https://doi.org/10.3390/atmos9110446.

Ciach, C. J., and W. F. Krajewski, 1999: On the estimation of radar rainfall error variance. *Adv. Water Resour.*, **22**, 585-595, https://doi.org/10.1016/S0309-1708(98)00043-8.

Code of Federal Regulations: IFR Altitudes. Special Federal Aviation Regulation No. 97, 3 pp., https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-95.

Daly, C., R. P. Neilson, D. L. Phillips, 1994: A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain. *J. Appl. Meteor. Climatol*., **33**, 140 -158, https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2.

Durham, B., J. Haviland, 2020: Reassessment of Title 14 CFR Volume 2 Part 95 Subpart B Designated Mountainous Areas. FAA, 08/02/2021, https://www.faa.gov/air_traffic/flight_info/aeronav/acf/media/Presentations/20-02-Designated-Mountainous-Areas.pdf.

Eldardiry, H., E. Habib, Y. Zhang, J. Graschel, 2015: Artifacts in Stage IV NWS Real-Time Multisensor Precipitation Estimates and Impacts on Identification of Maximum Series. *J. Hydrol. Eng.,* **22,** E4015003, https://doi:10.1061/(ASCE)HE.1943-5584.0001291.

ESRI, 2022: Data classification methods. Accessed February 2022, https://pro.arcgis.com/en/pro-app/latest/help/mapping/layer-properties/data-classification-methods.html.

Foehn, A., J. G. Hernández, B. Schaefli, G. D. Cesare, 2018: Spatial interpolation of precipitation from multiple rain gauge networks and weather radar data for operational applications in Alpine catchments. *J. Hydrol.,* **563,** 1092-1110, https://doi.org/10.1016/j.jhydrol.2018.05.027.

Fulton, R. A., J. P. Breidenbach, D. -J. Seo, D. A. Miller, T. O'Bannon, 1998: The WSR-88D Rainfall Algorithm. *Wea. Forecasting,* **13,** 377 -395, https://doi.org/10.1175/1520-0434(1998)013<0377:TWRA>2.0.CO;2.

Gentile, M., F. Courbin, G. Meylan, 2013: Interpolating point spread function anisotropy. *Astron. Astrophys.,* **549,** 1 -20, https://doi.org/10.1051/0004-6361/20121979.

Hao, Z., X. Yuan, Y. Xia, F. Hao, V. P. Singh, 2017: An Overview of Drought Monitoring And Prediction Systems at Regional and Global Scales. *Bull. Amer. Meteor. Soc.,* **98,** 1879 -1896, https://doi.org/10.1175/BAMS-D-15-00149.1.

Hayes, M. J., M. D. Svoboda, B. D. Wardlow, M. C. Anderson, F. Kogan, 2012: Drought Monitoring: Historical and Current Perspectives. *Remote Sensing of Drought: Innovative Monitoring Approaches, B.* D. Wardlow, M. C. Anderson, J. P. Verdin, CRC Press/Taylor & Francis, 1 -16.

Heim, R. R., 2002: A Review of Twentieth-Century Drought Indices Used in the United States. *Bull. Amer. Meteor. Soc.,* **83**, 1149 -1166, https://doi.org/10.1175/1520-0477-83.8.1149.

Houborg, R., M. Rodell, B. Li, R. Reichle, B. F. Zaitchik, 2012: Drought Indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resour. Res.,* **48,** W07525, https://doi.org/10.1029/2011WR011291.

Hunter, S. M., 1996: WSR-88D Radar Rainfall Estimation: Capabilities, Limitations, and Potential Improvements. *Natl. Wea. Dig,* **20,** 26 -38.

Jayakrishnan, R., R. Srinivasan, J. G. Arnold, 2004: Comparison of raingage and WSR-88D Stage III precipitation data over the Texas-Gulf basin. *J. Hydrol.,* **292,** 135 -152, https://doi.org/10.1016/j.jhydrol.2003.12.027.

Krajewski, W. F., A. Ntelekos, and R. Goska, 2006: A GIS-based methodology for the for the assessment of weather radar beam blockage in mountainous regions: Two examples from the U.S. NEXRAD network. Comput. Geosci., **32**, 283–302, https://doi.org/10.1016/j.cageo.2005.06.024

Krajewski, W. F., B. Vignal, B.-C. Seo, and G. Villarini, 2011: Statistical model of the range-dependent error in radar-rainfall estimates due to the vertical profile of reflectivity. *J. Hydrol*., **402,** 306-316, https://doi.org/10.1016/j.jhydrol.2011.03.024.

Legates, D. R., 2000: Real-Time Calibration of Radar Precipitation Estimates. *Prof. Geogr.*, **52,** 235 -246, https://doi.org/10.1111/0033-0124.00221.

Li, B., and Coauthors, 2019: Global GRACE Data Assimilation for Groundwater and Drought Monitoring: Advances and Challenges. *Water Resour. Res.,* **55,** 7564 -7586, https://doi.org/10.1029/2018WR024618.

Lin, Y., and K. E. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses: Development and applications. Preprints, *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2.

Liu, D., A. K. Mishra, Z. Yu, C. Yang, G. Konapala, T. Vu, 2017: Performance of SMAP, AMSR-E, and LAI for weekly agricultural drought forecasting over the continental United States. *J. Hydrol.,* **553,** 88 -104, http://dx.doi.org/10.1016/j.jhydrol.2017.07.049.

McRoberts, D. B., 2014: Minimizing Biases in Radar Precipitation Estimates. Ph.D. dissertation, Dept of Atmospheric Sciences, Texas A&M University, 220 pp.

Maddox R. A., J. Zhang, J. J. Gourley, K. W. Howard, 2002: Weather radar coverage over the contiguous united states. *Wea. Forecasting,* **17,** 927-934, https://doi.org/10.1175/1520-0434(2002)017<0927:WRCOTC>2.0.CO;2

NIDIS, 2006: National Integrated Drought Information System Act of 2006. Pub. L. 109-430, 15 U.S.C. 311 and 313d. [Available online at www.gpo.gov/fdsys/pkg/PLAW-109publ430/pdf/PLAW-109publ430.pdf.]

Mishra, A., T. Vu, A. V. Veettil, D. Entekhabi, 2017: Drought Monitoring with soil moisture active passive (SMAP) measurements. *J. Hydrol.,* **552,** 620 -632, http://dx.doi.org/10.1016/j.jhydrol.2017.07.033.

Mladenova, I. E., J. D. Bolten, W. T. Crow, N. Sazib, M. H. Cosh, C. J. Tucker, C. Reynolds, 2019: Evaluating the Operational Application of SMAP for Global Agricultural Drought Monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **12,** 3387 -3397, 10.1109/JSTARS.2019.2923555.

Mukherjee, S., A. Mishra, K. E. Trenberth, 2018: Climate Change and Drought: a Perspective on Drought Indices. *Curr. Clim. Change. Rep.,* **4,** 145 -163, https://doi.org/10.1007/s40641-018-0098-x.

Nelson, B. R., O. P. Pratt, D.-J. Seo, E. Habib, 2015: Assessment and Implications of NCEP Stage IV Quantitative Precipitation Estimates for Product Intercomparisons. *Wea. Forecasting,* **31,** 371 -394, 10.1175/WAF-D-14-00112.

Overeem, A., I. Holleman, and A. Buishand, 2009: Derivation of a 10-year radar-based climatology of rainfall. *J. Appl. Meteorol. Climatol.,* **48,** 1448-1463, https://doi.org/10.1175/2009JAMC1954.1

Quiring, S. M., 2009: Developing Objective Operational Definitions for Monitoring Drought. *J. Appl. Meteor. Climatol.,* **48,** 1217 -1229, 10.1175/2009JAMC2088.1.

Rabiei, E., U. Haberlandt, 2015: Applying bias correction for merging rain gauge and radar data. *J. Hydrol.,* **522,** 544-557, http://dx.doi.org/10.1016/j.jhydrol.2015.01.020.

Sato, Y., and Coauthors, 1998: Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. *Medical Image Analysis,* **2**, 143-168, https://doi.org/10.1016/S1361-8415(98)80009-1.

Shellito, P. J., and Coauthors, 2016: SMAP Soil Moisture drying more rapid than observed in situ following rainfall events. *Geophys. Res. Lett.,* **43,** 8069 -8075, 10.1002/2016GL069946.

Steiner, M., J. A. Smith, S. J. Burges, C. V. Alonso, R. W. Darden, 1999: Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation. *Water Resour. Res.,* **35,** 2487 - 2503, https://doi.org/10.1029/1999WR900142.

Thomas, A. C., J. T. Reager, J. S. Famiglietti, M. Rodell, 2014: A GRACE Based water storage deficit approach for hydrological drought characterization. *Geophys. Res. Lett.,* **41,** 1537 - 1545, 10.1002/2014GL059323.

United States Bureau of Transportation Statistic. United States. Bureau of the Census Geography Division (2014). State Boundaries. United States and Territories, 2012. [Shapefile].

Villarini, G., and W. F. Krajewski, 2009: Empirically based modelling of radar-rainfall uncertainties for a C-band radar at different time-scales. *Q.J.R. Meteorol. Soc.,* **135,** 1424-1438, https://doi.org/10.1002/qj.454.

Villarini, G., and W. F. Krajewski, 2010a: Sensitivity studies of the models of radar-rainfall uncertainties. *J. Appl. Meteor. Climatol.*, **49**, 288-309, https://doi.org/10.1175/2009JAMC2188.1

Villarini, G., and W. F. Krajewski, 2010b: Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall. *Surv. Geophys*., **31**, 107-129, 10.1007/s10712-009-9079-x.

Westcott, N. E., H. V. Knapp, S. D. Hilberg, 2008: Comparison of gage and multi-sensor precipitation estimates over a range of spatial and temporal scales in the Midwestern United States. *J. Hydrol.,* **351,** 1 -12, 10.1016/j.jhydrol.2007.10.057.

Yilmaz, K. K., T. S. Hogue, K-L. Hsu, S. Sorooshian, H. V. Gupta, T. Wagener, 2005: Intercomparison of Rain Gauge, Radar, and Satellite-Based Precipitation Estimates with Emphasis on Hydrologic Forecasting. *J. Hydrometeor.,* **6,** 497 - 517, https://doi.org/10.1175/JHM431.1.

Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.,* **117,** D03139, 10.1029/2011JD016048.

Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *J. Geophys. Res.,* **117,** D03110, 10.1029/2011JD0.