

ABSTRACT

WOOTTEN, ADRIENNE MARI. Evaluating Weather Generator Techniques for Downscaling Precipitation for Seasonal Forecasting. (Under the direction of Ryan P. Boyles.)

Precipitation has an influence on multiple sectors including agriculture, water resource management and natural resource management. Monthly and seasonal estimates of precipitation from global climate models (GCMs) such as the Climate Forecast System model are used to provide seasonal forecasts of precipitation. However, the spatial resolution of most GCMs are several hundred kilometers which does not allow for detailed local estimates needed for crop modeling, streamflow modeling, and decision making for irrigation and water resources management. Downscaling is used to translate GCM information to local scales. Dynamic downscaling techniques are the most computationally expensive of these techniques, while empirical statistical downscaling (ESD) techniques are less expensive and can provide similar results to dynamic techniques. One technique to downscale GCMs is the use of weather generators, which can provide local estimates in areas with sparse observed data and a virtually unlimited number of local daily realizations of seasonal forecasts. However, weather generators often do not account for the spatial structure of precipitation or capture the variability associated with extreme events such as tropical cyclones and droughts.

In this study, the WGEN weather generator (Richardson and Wright, 1984), the geo-spatial temporal weather generator (GiST, Baigorria and Jones, 2010), and a new version of GiST with additional consideration for extreme events (GiSTR) are evaluated against observations for 1971-2000. Both GiST and GiSTR are shown to replicate the spatial structure of precipitation better than WGEN, while GiSTR better replicates the observed variability of precipitation in this time period as compared to GiST and WGEN. While GiSTR replicates the variability of precipitation it overestimates the average total precipitation in the period. GiST provides the best replication of the average total precipitation. Given the ability to replicate both the spatial and temporal structure along with

the variability of precipitation, GiSTR is compared with WGEN to evaluate several variations to downscaling with weather generators.

Downscaling with weather generators involves perturbing the input parameters of the generator to reflect the changes projected by the GCM in the forecast period based on a scaling relationship. A second set of analyses focuses on evaluating four different approaches to using the scaling relationship to downscale with weather generators based on two assumptions used in this relationship. The first assumption involves the interpolation of current day parameters at each station to the GCM grid points, and the second assumption involves the estimation of the parameters of the two-parameter gamma distribution used for estimating precipitation amounts. Most generators commonly use area weighted averaging for parameter interpolation and method of moments to estimate the gamma distribution. This study compares this variation against three new variations using GiSTR and Climate Forecast System Reanalysis (CFSR) in two separate time periods. The results of this analysis suggest that for both time periods the commonly used variation has larger error than all of the new variations when applied in the Southeast U.S. The results also indicate that variations using locally weighted regression or Greenwood and Durand approximation have lower error than other variations.

Finally, WGEN and GiSTR are used with one of the downscaling variations to test the performance of each generator when they are used for downscaling in a forecast mode. The results of this additional analysis indicate that GiSTR has lower error than WGEN when used for downscaling in the Southeast U.S. However, it is noted that WGEN may be sensitive to the choice of downscaling variation. Future work in this area includes additional improvements to GiSTR with regards to extreme precipitation, evaluating each combination of downscaling variations and weather generators for the Southeast U.S., and applying weather generator downscaling to seasonal forecasting.

Evaluating Weather Generator Techniques for Downscaling Precipitation
For Seasonal Forecasting

by
Adrienne Mari Wootten

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Marine, Earth and Atmospheric Sciences

Raleigh, North Carolina

2011

APPROVED BY:

Ryan P. Boyles
Committee Chair

Guillermo A. Baigorria

Montserrat Fuentes

Fredrick Semazzi

BIOGRAPHY

Adrienne Wootten was born near New York, NY and raised outside of Baltimore, MD. She is the daughter of two supportive parents, Peter and Liidia, and has an exceptional sister, Jennifer. She graduated from Fallston High School in 2004 and enrolled in the Department of Marine, Earth, and Atmospheric Sciences at North Carolina State University (NCSU) in the fall of 2004. During her undergraduate education, she was actively involved with the NCSU Women in Science and Engineering (WISE) program, where she served on the student council and received the Lifesaver Award for service to program in 2005. In addition, she was actively involved with the NCSU student chapter of the American Meteorological Society (AMS). These activities led to her involvement in the Environmental Statistics Practicum program under the direction of William F. Hunt Jr. from Spring 2007 through Spring 2008. Under this program, Adrienne was involved in consulting projects with various agencies including the North Carolina Department of Environment and Natural Resources, the New Jersey Department of Environmental Protection, and the Maryland Department of the Environment. Her involvement in AMS also led to her involvement in research with Dr. Sethu Raman and the State Climate Office of North Carolina during the summer of 2007. Adrienne graduated Magna Cum Laude in May 2008 from NCSU with a Bachelor of Science Degree in Meteorology and a minor in Statistics. Following graduation, she was hired by the U.S. Environmental Protection Agency Atmospheric Modeling and Analysis Division (U.S. EPA-AMAD) as a student contractor while continuing her work at the North Carolina State Climate Office. In fall of 2009, Adrienne was accepted into the Master of Science in Atmospheric Science program at NCSU as a graduate research assistant under the guidance of Dr. Ryan Boyles. Adrienne continues to work closely with the State Climate Office of North Carolina, where she actively participates in research, extension and outreach activities.

ACKNOWLEDGMENTS

First and foremost, I thank God for giving me the strength, wisdom, and ability to pursue my thesis research in the face of adversity. I would also like to extend my sincerest thanks to my committee chair, Dr. Ryan Boyles, for the opportunity to pursue my graduate education. Ryan, thank you for your encouragement and patience for those times when there has been “one more thing” to discuss. I am also extremely grateful to Guillermo Baigorria, who provided a wealth of knowledge and advice about his own weather generator and what I have implemented as a new addition. A huge thank you to my committee members, Dr. Montserrat Fuentes and Dr. Fredrick Semazzi, for providing advice and knowledge in both statistical and climate modeling, along with ideas for improvements. In addition, I also thank the Southeast Climate Consortium for funding my research, and Dr. Vasu Misra for the additional advice he provided.

My sincerest thanks also go to the staff and students at the State Climate Office of North Carolina for their support and ideas for this project and several others in the past four years. In addition, I am also extremely grateful to all my friends for their patience and support in this process.

Last and certainly not least, my biggest thanks goes to my parents, my sister (and brother-in-law), and all the rest of my family. All of you have been the biggest support throughout this process and much of my life. There are no words to express my gratitude for what all of you have done. For being the biggest cheerleaders and support, in this and much more, I am extremely grateful. THANK YOU!!!!

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	xvii
Chapter 1 . Introduction	1
1.1. Downscaling Background.....	1
1.1.1. Dynamic Downscaling.....	2
1.1.2. Empirical Statistical Downscaling	3
1.2. Seasonal Precipitation Patterns in the Southeast U.S.	12
1.3. Motivation and Objectives.....	13
Chapter 2 . Methods and Experimental Design	21
2.1. Overview of GiST.....	21
2.1.1. Initial Event Generation	22
2.1.2. Subsequent Event Generation	24
2.1.3. Amount Generation.....	24
2.2. Potential Improvements to the GiST weather generator.....	25
2.2.1. Extreme Precipitation Generation.....	25
2.2.2. Probable Results of the altered GiST technique	26
2.3. Potential Improvements to Weather Generator Downscaling	27
2.3.1. The General Concept of Weather Generator Downscaling	27
2.3.2. Changes and Improvements to the General Approach.....	31
2.4. Experimental Design.....	34
2.4.1. Design for Evaluating Weather Generator Improvements.....	34
2.4.2. Design for Evaluating Downscaling Variations.....	36
Chapter 3 . Weather Generator Evaluation	45
3.1. North Carolina Domain	45
3.1.1. Temporal Parameters	46
3.1.2. Spatial Structure Parameters	47
3.1.3. Precipitation Amount Parameters	49

3.1.4. Probability Distribution Function Analysis	53
3.1.5. Dry and Wet Spell Frequency Analysis	54
3.2. Southeast United States Evaluation	56
3.2.1. Temporal Parameters	57
3.2.2. Spatial Structure Parameters	58
3.2.3. Precipitation Amount Parameters	60
3.2.4. Probability Distribution Function Analysis	66
3.2.5. Dry and Wet Spell Frequency Analysis	68
3.3. Summary Discussion	70
Chapter 4 . Evaluation of Downscaling Variations – 1979-2000	149
4.1. Downscaling Simulation Comparison	151
4.1.1. Temporal Parameters	151
4.1.2. Spatial Structure Parameters	153
4.1.3. Precipitation Amount Parameters	156
4.1.4. Probability Distribution Function Analysis	162
4.1.5. Dry and Wet Spell Frequency Analysis	165
4.2. Weather Generator Comparison - Downscaling Context	167
4.2.1. Temporal Parameters	168
4.2.2. Spatial Structure Parameters	169
4.2.3. Precipitation Amount Parameters	171
4.2.4. Probability Distribution Function Analysis	174
4.2.5. Changes with Outliers Removed	176
4.2.6. Dry and Wet Spell Frequency Analysis	178
4.3. Summary Discussion	179
4.3.1. Downscaling Simulation Comparison	179
4.3.2. Weather Generator Comparison – Downscaling Context.....	181
Chapter 5 . Evaluation of Downscaling Variations: 2001-2009	269
5.1. Downscaling Simulation Comparison	270
5.1.1. Temporal Parameters	270
5.1.2. Spatial Structure Parameters	272

5.1.3. Precipitation Amount Parameters	275
5.1.4. Probability Distribution Function Analysis	283
5.1.5. Dry and Wet Spell Frequency Analysis	287
5.2. Weather Generator Comparison – Downscaling Context.....	289
5.2.1. Temporal Parameters	290
5.2.2. Spatial Structure Parameters	291
5.2.3. Precipitation Amount Parameters	293
5.2.4. Probability Distribution Function Analysis	297
5.2.5. Changes with Outliers Removed	298
5.2.6. Dry and Wet Spell Frequency Analysis	301
5.3. Summary Discussion	303
5.3.1. Downscaling Simulation Comparison	303
5.3.2. Weather Generator Comparison – Downscaling Context.....	307
Chapter 6 . Conclusions and Future Work	408
6.1. Summary and Conclusions	408
6.2. Recommended Usage	422
6.3. Potential Applications.....	424
6.4. Recommendations for Future Work	425
REFERENCES.....	435

LIST OF TABLES

Table 1.1. Advantages and Disadvantages of the three categories of ESD techniques	16
Table 2.1. Description of parameters used to evaluate the each weather generator.	40
Table 3.1. RMSE comparison of generated Markov Transition probabilities (P_{01} and P_{11}) values for the NC domain.	77
Table 3.2. Summary of Results from the two-sample T-tests with the 1st order Markov Transition Probabilities (P_{01} and P_{11}) for the NC domain. P-values in red indicate that the difference in RMSE between the two generators compared is significantly less than zero. The Difference in average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.	77
Table 3.3. RMSE comparison of generated unconditional probability of rain (π) and persistence (γ) values for the NC domain.	78
Table 3.4. Summary of Results from the two-sample T-tests with unconditional probability of rain (π) and persistence (γ) for the NC domain. P-values in red indicate that the difference in RMSE between the two generators compared is significantly less than zero. The Difference in average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.	78
Table 3.5. Summary of results of two-sample T-tests for each correlation matrix (p-values only) for the NC domain. P-values in red indicate that RMSE of the first generator is significantly larger than the second generator.	79
Table 3.6. Summary of results of the two-sample T-test on the generated values of the mean daily nonzero precipitation (μ) for the North Carolina domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.	79
Table 3.7. Summary of results of the two-sample T-test on the generated values of the variance of daily nonzero precipitation (σ^2) for the NC domain. P-values in red	

indicate RMSE of one generator is larger than the other. The Difference in Average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.....	79
Table 3.8. Summary of results of the two-sample T-test on the generated values of the average total precipitation ($E[S(T)]$) for the NC domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.....	80
Table 3.9. Summary of results of the two-sample T-test on the generated values of the inter-annual variability ($Var[S(T)]$) for the NC domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.	80
Table 3.10. RMSE comparison of generated Markov Transition probabilities (P_{01} and P_{11}) values for the Southeast U.S.	81
Table 3.11. Summary of Results from the two-sample T-tests with the Markov Transition Probabilities for the Southeast U.S. domain. P-values in red indicate that the difference in RMSE between the two generators compared is significantly less than zero.....	81
Table 3.12. RMSE comparison of generated unconditional probability of rain (π) and persistence (γ) values for the Southeast U.S. domain.	82
Table 3.13. Summary of Results from the two-sample T-tests with unconditional probability of rain (π) and persistence (γ) for the Southeast US domain. P-values in red indicate that the difference in RMSE between the two generators compared is smaller or larger than zero, depending on the Difference in the Average RMSE.....	82
Table 3.14. Summary of results of one tailed two-sample T-tests for each correlation matrix (p-values only) for the Southeast US domain. P-values in red indicate that RMSE of the first generator is significantly larger than the second generator.	83

Table 3.15. Summary of results of the two-sample T-test on the generated values of the mean daily nonzero precipitation (μ) for the Southeast U.S. domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.....	83
Table 3.16. Summary of results of the two-sample T-test on the generated values of the variance of daily nonzero precipitation (σ^2) for the Southeast U.S. domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.	83
Table 3.17. Summary of results of the two-sample T-test on the generated values of the average total precipitation ($E[S(T)]$) for the Southeast U.S. domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE is the average RMSE of the second generator subtracted from the Average RMSE of the first generator.....	84
Table 3.18. Summary of results of the two sample T-test on the generated values of the inter-annual variability ($Var[S(T)]$) for the Southeast U.S. domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE (mm^2) is the average RMSE of the second generator subtracted from the first generator.	84
Table 3.19 .Summary results of which generator has the lowest error on average for each parameter tested.	85
Table 4.1. RMSE Comparison for each downscaling simulation across the Southeast U.S. for value of P_{01}	185
Table 4.2. RMSE Comparison for each downscaling simulation across the Southeast U.S. for value of P_{11}	185
Table 4.3. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the Markov Transition Probabilities (P_{01} and P_{11}). P-values in red	

indicate that either the first simulation has a significantly larger or smaller RMSE than the second..... 186

Table 4.4. Two-sample T-test results comparing each downscaling simulation to the Control for the Markov Transition Probabilities. The difference in the average RMSE is the Simulation – Control. P-values in red indicate that RMSE for that simulation is significantly larger than the Control. 186

Table 4.5. RMSE Comparison for each downscaling simulation across the Southeast U.S. for value of the unconditional probability of rain (π)..... 187

Table 4.6. RMSE Comparison for each downscaling simulation across the Southeast U.S. for value of persistence (γ)..... 187

Table 4.7. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the unconditional probability of rain (π) and persistence (γ). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second..... 188

Table 4.8. Two-sample T-test results comparing each downscaling simulation to the Control for the unconditional probability of rain (π) and persistence (γ). The difference in the average RMSE is the Simulation – Control. P-values in red indicate that RMSE for that simulation is significantly less than the Control. 188

Table 4.9. Average RMSE for each downscaling simulation for each correlation matrix. .. 188

Table 4.10. Two-sample T-test results comparing each downscaling simulation to the Control for the correlation matrix of precipitation events (ρ_{ev}). The difference in the average RMSE is the Simulation – Control. P-values in red indicate that RMSE for that simulation is significantly larger than the Control..... 189

Table 4.11. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the Markov transition probabilities for extreme events. P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second simulation..... 189

Table 4.12. Two-sample T-test results comparing each downscaling simulation to the Control for the correlation matrix for precipitation extreme events (ρ_{ex}). The difference in the

average RMSE is the Simulation – Control. P-values in red indicate that RMSE for that simulation is significantly larger than the Control..... 190

Table 4.13. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the mean daily nonzero precipitation (μ). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second. 190

Table 4.14. Summary of Two-sample T-test results comparing downscaling simulations to the Control for the RMSE of the mean daily nonzero precipitation (μ). P-values in red indicate the simulation has a significantly larger RMSE than the Control..... 190

Table 4.15. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the variance of daily nonzero precipitation (σ^2). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second. 191

Table 4.16. Summary of Two-sample T-test results comparing downscaling simulations to the Control for the RMSE of the variance of daily nonzero precipitation (σ^2). P-values in red indicate that either that the simulation has a significantly larger RMSE than the Control..... 191

Table 4.17. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the average total precipitation ($E[S(T)]$). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second. 191

Table 4.18. Summary of Two-sample T-test results comparing downscaling simulations to the Control for the RMSE of the average total precipitation ($E[S(T)]$). P-values in red indicate that either that the simulation has a significantly larger RMSE than the Control. 192

Table 4.19. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the inter-annual variability ($Var[S(T)]$). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second. 192

Table 4.20. Summary of Two-sample T-test results comparing downscaling simulations to the Control for the RMSE of the inter-annual variability ($Var[S(T)]$). P-values in red indicate that either that the simulation has a significantly larger RMSE than the Control.	192
Table 4.21. Comparison of the RMSE of Markov Transition probabilities (P_{01} and P_{11}) between the GiSTR simulation and the WGEN simulation.....	193
Table 4.22. Comparison of the RMSE of the unconditional probability of rain (π) and persistence (γ) between the GiSTR simulation and the WGEN simulation.....	194
Table 4.23. Summary of the two-sample T-test results for Markov Transition probabilities(P_{01} and P_{11}), the unconditional probability of rain (π), and persistence (γ). P-values in red indicate that the RMSE of that parameter for WGEN is significantly larger than the RMSE of that parameter for GiSTR.	194
Table 4.24. Summary of two-sample T-test results for the RMSE of the correlation matrix of precipitation (ρ), the correlation matrix of precipitation events(ρ_{ev}),the correlation matrix of precipitation amounts(ρ_{am}), and the correlation matrix of precipitation extreme events (ρ_{ex}). P-values in red indicate that the RMSE of that parameter for WGEN is significantly larger than the RMSE for GiSTR.....	195
Table 4.25. Summary of the two-sample T-test results for the mean daily nonzero precipitation (μ), the variance of the daily nonzero precipitation (σ^2), the average total precipitation ($E[S(T)]$), and the inter-annual variability($Var[S(T)]$). P-values in red indicate that the RMSE of that parameter for WGEN is significantly larger than the RMSE of that parameter for GiSTR.	195
Table 4.26. Summary of the two-sample T-test results for the mean daily nonzero precipitation (μ), the variance of the daily nonzero precipitation (σ^2), the average total precipitation ($E[S(T)]$), and the inter-annual variability($Var[S(T)]$). after removing outlier stations. P-values in red indicate that the RMSE of that parameter for WGEN is significantly larger than the RMSE of that parameter for GiSTR.....	196
Table 4.27. Summary of Best Downscaling Variation on Average for the different parameters evaluated.	196

Table 4.28. Comparison Table between the weather generators (full details in Chapter 3) and downscaling simulations in the Southeast U.S. Comparison includes GiSTR and WGEN only.	197
Table 5.1. RMSE Comparison for each 2001-2009 downscaling simulation across the Southeast U.S. for value of P_{01}	309
Table 5.2. RMSE Comparison for each 2001-2009 downscaling simulation across the Southeast U.S. for value of P_{11}	309
Table 5.3. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the Markov Transition Probabilities (P_{01} and P_{11}). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second. The Difference in the Average RMSE indicates is the difference between the average RMSE of the first simulation and the average RMSE of the second simulation.	310
Table 5.4. Summary of Two-sample T-test results on the RMSE of the Markov Transition Probabilities (P_{01} and P_{11}) for each 2001-2009 downscaling simulation with the corresponding 1979-2000 downscaling simulation. P-values in red indicate that the RMSE of the 2001-2009 simulation is significantly larger than the 1979-2000 simulation.	310
Table 5.5. RMSE Comparison for each 2001-2009 downscaling simulation across the Southeast U.S. for values of the conditional probability of rain (π).	311
Table 5.6. RMSE Comparison for each 2001-2009 downscaling simulation across the Southeast U.S. for values of persistence (γ).	311
Table 5.7. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the unconditional probability of rain (π) and persistence (γ). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second. The Difference in the Average RMSE is the difference between the average RMSE of the first simulation and the average RMSE of the second simulation.	312

Table 5.8. Summary of Two-sample T-test results on the RMSE of the unconditional probability of rain (π) and persistence (γ) for each 2001-2009 downscaling simulation with the corresponding 1979-2000 downscaling simulation. P-values in red indicate that the RMSE of the 2001-2009 simulation is significantly larger than the 1979-2000 simulation.....	312
Table 5.9. Average RMSE for each 2001-2009 downscaling simulation for each correlation matrix.	313
Table 5.10. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the mean daily nonzero precipitation (μ) and the variance of the daily nonzero precipitation(σ^2). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second. The Difference in the Average RMSE is the difference between the average RMSE of the first simulation and the average RMSE of the second simulation.	313
Table 5.11. Summary of Two-sample T-test results on the RMSE the mean daily nonzero precipitation (μ) and the variance of the daily nonzero precipitation(σ^2). P-values in red indicate that the RMSE of the 2001-2009 simulation is significantly larger than the corresponding RMSE of the 1979-2000 simulation.	314
Table 5.12. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the average total precipitation ($E[S(T)]$) and the inter-annual variability($Var[S(T)]$). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second. The Difference in the Average RMSE indicates is the difference between the average RMSE of the first simulation and the average RMSE of the second simulation.	314
Table 5.13. Summary of Two-sample T-test results on the RMSE of the average total precipitation ($E[S(T)]$) and the inter-annual variability($Var[S(T)]$)for each 2001-2009 simulation with the corresponding 1979-2000 simulation. P-values in red indicate that the RMSE of the 2001-2009 simulation is significantly larger than the 1979-2000 simulation.....	315

Table 5.14. Comparison of the RMSE of the Markov Transition Probabilities (P_{01} and P_{11}) between the GiSTR downscaling simulation and the WGEN downscaling simulation. Time period of 2001-2009. 315

Table 5.15. Comparison of the RMSE of the unconditional probability of rain (π) and persistence (γ) between the GiSTR downscaling simulation and the WGEN downscaling simulation. Time period of 2001-2009. 316

Table 5.16. Summary of the two-sample T-test results for the Markov Transition Probabilities (P_{01} and P_{11}), the unconditional probability of rain (π), and persistence (γ). P-values in red indicate that the RMSE of that parameter for WGEN is significantly larger than the RMSE of that parameter for GiSTR. Time period of 2001-2009. 316

Table 5.17. Summary of two-sample T-test results for the RMSE of the correlation matrix of precipitation (ρ), the correlation matrix of precipitation events (ρ_{ev}), the correlation matrix of precipitation amounts(ρ_{am}), and the correlation matrix of precipitation extreme events(ρ_{ex}). P-values in red indicate that the RMSE of that parameter for WGEN is significantly larger than the RMSE for GiSTR..... 317

Table 5.18. Summary of two-sample T-test results from the RMSE of the correlation matrix of precipitation (ρ), the correlation matrix of precipitation events (ρ_{ev}), the correlation matrix of precipitation amounts(ρ_{am}), and the correlation matrix of precipitation extreme events(ρ_{ex})between each 2001-2009 simulation and its Control from Chapter 3. The Difference in the Average RMSE is the Simulation – Control. P-values in red indicate that the RMSE of that simulation is larger than the Control simulations. .. 317

Table 5.19. Summary of the two-sample T-test results for the mean daily nonzero precipitation (μ), the variance of the daily nonzero precipitation (σ^2), the average total precipitation ($E[S(T)]$), and the inter-annual variability ($Var[S(T)]$). P-values in red indicate that the RMSE of that parameter for the WGEN simulation is significantly larger than the RMSE of that parameter for GiSTR simulation. 318

Table 5.20. Summary of the two-sample T-test results for the mean daily nonzero precipitation (μ), the variance of the daily nonzero precipitation (σ^2), the average total

precipitation ($E[S(T)]$), and the inter-annual variability ($Var[S(T)]$). P-values in red indicate that the RMSE of that parameter for the WGEN simulation is significantly larger than the RMSE of that parameter for GiSTR simulation. Outlier stations removed..... 318

Table 5.21. Summary of results from Chapters 4 and 5 for the downscaling simulation comparison by parameter..... 319

Table 5.22. Comparison Table between the Control (full details in Chapter 3), 1979-2000 downscaling simulations, and 2001-2009 downscaling simulations in the Southeast U.S. Comparison includes GiSTR and WGEN only. 320

Table 5.23. Comparison Table between downscaling simulations in the Southeast U.S. Comparison includes GiSTR and WGEN only, with outlier stations removed..... 321

Table 6.1. Description of parameters used to evaluate the each weather generator. 428

Table 6.2. Average RMSE Comparison for values of the mean daily nonzero precipitation (μ) and the average total precipitation ($E[S(T)]$) between time periods and downscaling simulations. 428

LIST OF FIGURES

Figure 1.1. Flowchart for daily precipitation generation using Markov chain generators. Adapted from Wilks and Wilby (1999).	17
Figure 1.2. Flowchart for daily precipitation generation using spell-length generators. Adapted from Wilks and Wilby (1999)	18
Figure 1.3. Histogram of Historical Daily Nonzero Precipitation at Raleigh-Durham Airport.	19
Figure 1.4. 0555Z 12/26/2010 NEXRAD Southeast U.S. Radar Reflectivity, showing frontal precipitation.	20
Figure 1.5. 2254Z 5/13/2011 NEXRAD Southeast U.S. Radar Reflectivity, showing unstructured convection.	20
Figure 2.1. Example of Theoretical PDFs produced by GiST and GiSTR.....	41
Figure 2.2. Locations of the 91 COOP stations in the North Carolina domain used in the weather generator evaluation	42
Figure 2.3. Locations of the 319 COOP stations in the Southeast United States domain used in the weather generator evaluation.	43
Figure 2.4. NWS COOP stations (black) circles and CFSR grid points (red dots) used in the downscaling evaluation.....	44
Figure 3.1. RMSE of each weather generator by month for the correlation matrix of precipitation (ρ) (a), the correlation matrix of precipitation events (ρ_{ev})(b), the correlation matrix of precipitation amounts (ρ_{am})(c), and the correlation matrix of precipitation extreme events (ρ_{ex})(d) for North Carolina domain.....	86
Figure 3.2. Average Decay Function Correlogram for each generator and observations for January for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). North Carolina domain.	87

Figure 3.3. Average Decay Function Correlogram for each generator and observations for July for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). North Carolina domain.	88
Figure 3.4. RMSE (mm) comparison for the generated values of mean daily nonzero precipitation (μ) for North Carolina domain.....	89
Figure 3.5. Difference between generated and observed values of mean daily nonzero precipitation (μ) across the North Carolina domain for the month of August for WGEN(a), GiST(b), and GiSTR (c)	90
Figure 3.6. RMSE (mm ²) comparison for the generated values of the variance of the daily nonzero precipitation (σ^2) for North Carolina domain.....	93
Figure 3.7. Difference between generated and observed values of the variance of daily nonzero precipitation (σ^2) across the North Carolina domain for the month of September for WGEN(a), GiST(b), and GiSTR (c)	94
Figure 3.8. RMSE (mm) comparison for generated values of the average total precipitation ($E[S(T)]$) for North Carolina domain.....	97
Figure 3.9. Boxplots of the difference between generated and observed values of the average total precipitation ($E[S(T)]$) for WGEN (a), GiST (b), and GiSTR(c) for North Carolina domain.....	98
Figure 3.10. RMSE (mm ²) comparison for generated values of the inter-annual variability ($Var[S(T)]$) for North Carolina domain.	99
Figure 3.11. Generated vs. Observed PDF examples for January for station 315838 in Morganton, NC for WGEN (a), GiST (b), and GiSTR (c).	100
Figure 3.12. Generated vs. Observed PDF examples for September for station 319457 in Wilmington, NC for WGEN (a), GiST (b), and GiSTR (c).....	103
Figure 3.13. Monthly PDF examples, generated vs. observed for GiSTR for January (a) and September (b) for all stations in the domain.	106
Figure 3.14. Boxplots of the frequency of January Wet Spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for North Carolina domain.....	107

Figure 3.15. Boxplots of the frequency of January dry spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for North Carolina domain.	108
Figure 3.16. Boxplots of the frequency of July Wet Spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for North Carolina domain.	109
Figure 3.17. Boxplots of the frequency of July dry spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for North Carolina domain.	110
Figure 3.18. RMSE of each weather generator by month for the correlation matrix of precipitation (ρ) (a), the correlation matrix of precipitation events (ρ_{ev})(b), the correlation matrix of precipitation amounts (ρ_{am})(c), and the correlation matrix of precipitation extreme events (ρ_{ex})(d) for the Southeast US domain.	111
Figure 3.19. Average Decay Function Correlogram for each generator and observations for January for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). Southeast U.S. domain.	112
Figure 3.20. Average Decay Function Correlogram for each generator and observations for July for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). Southeast U.S. domain.	113
Figure 3.21. RMSE (mm) comparison for the generated values of mean daily nonzero precipitation (μ) for the Southeast US domain.	114
Figure 3.22. Difference between generated and observed values of mean daily nonzero precipitation (μ) (mm) across the Southeast U.S. domain for the month of August for WGEN(a), GiST(b), and GiSTR (c).	115
Figure 3.23. RMSE (mm^2) comparison for the generated values of the variance of daily nonzero precipitation (σ^2)for the Southeast US domain.	119
Figure 3.24. Difference between generated and observed values of the variance of daily nonzero precipitation (σ^2)(mm^2) across the Southeast U.S. domain for the month of August for WGEN (a), GiST (b), and GiSTR (c).	120

Figure 3.25. RMSE (mm) comparison for the generated values of the average total precipitation($E[S(T)]$) for the Southeast U.S. domain.....	124
Figure 3.26. RMSE (mm^2) comparison for the generated values of the inter-annual variability ($Var[S(T)]$) for the Southeast US domain.....	125
Figure 3.27. Generated vs. Observed PDF examples for January for station 310301 in Asheville, NC for each weather generator for WGEN (a), GiST (b), and GiSTR (c).	126
Figure 3.28. Generated vs. Observed PDF examples for January for station 460102 in Alderson, WV for each weather generator for WGEN (a), GiST (b), and GiSTR (c).	129
Figure 3.29. Generated vs. Observed PDF examples for January for station 089176 in Venice, FL for each weather generator for WGEN (a), GiST (b), and GiSTR (c)...	132
Figure 3.30. Generated vs. Observed PDF examples for September for station 310301 in Asheville, NC for each weather generator for WGEN (a), GiST (b), and GiSTR (c).	135
Figure 3.31. Generated vs. Observed PDF examples for September for station 460102 in Alderson, WV for each weather generator for WGEN (a), GiST (b), and GiSTR (c).	138
Figure 3.32. Generated vs. Observed PDF examples for September for station 089176 in Venice, FL for each weather generator for WGEN (a), GiST (b), and GiSTR (c)...	141
Figure 3.33. Boxplots of the frequency of January Wet Spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for the Southeast US domain.	144
Figure 3.34. Boxplots of the frequency of January Dry Spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for the Southeast US domain.	145
Figure 3.35. Boxplots of the frequency of July Wet Spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for the Southeast US domain.....	146

Figure 3.36. Boxplots of the frequency of July Dry Spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for the Southeast US domain.....	147
Figure 3.37. Observed values of the variance of daily nonzero precipitation (σ^2) (mm ²) across the Southeast U.S. domain for September.	148
Figure 4.1. RMSE of each downscaling simulation by month for the correlation matrix of precipitation (ρ) (a), the correlation matrix of precipitation events(ρ_{ev}) (b), the correlation matrix of precipitation amounts (ρ_{am}) (c), and the correlation matrix of precipitation extreme events (ρ_{ex}) (d).....	198
Figure 4.2. Average Decay Function Correlogram for each simulation and observations for January for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 1979-2000 time period.	199
Figure 4.3. Average Decay Function Correlogram for each simulation and observations for July for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 1979-2000 time period.	200
Figure 4.4. RMSE Comparison between downscaling simulations for mean daily nonzero precipitation (μ).	201
Figure 4.5. Difference between generated and observed values of mean daily nonzero precipitation (μ) for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for January.....	202
Figure 4.6. Difference between generated and observed values of mean daily nonzero precipitation (μ) for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for June.....	207
Figure 4.7. Difference between generated and observed values of mean daily nonzero precipitation (μ) for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for September.	212
Figure 4.8. RMSE Comparison between downscaling simulations for the variance of daily nonzero precipitation (σ^2).	217

Figure 4.9. Difference between generated and observed values of the variance of daily nonzero precipitation (σ^2) for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for April.....	218
Figure 4.10. Difference between generated and observed values of the variance of daily nonzero precipitation (σ^2) for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for October.	223
Figure 4.11. RMSE Comparison between downscaling simulations for the RMSE of the average total precipitation ($E[S(T)]$).....	228
Figure 4.12. Comparison between downscaling simulations for the RMSE of the inter-annual variability ($Var[S(T)]$).	229
Figure 4.13. Generated vs. Observed PDF examples for a station 083163 in Ft. Lauderdale, FL for January for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).....	230
Figure 4.14. Generated vs. Observed PDF examples for a station 311975 in Concord, NC for January for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d)..	233
Figure 4.15. Generated vs. Observed PDF examples for a station 283516 in Greenwood Lake, NJ for January for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).....	236
Figure 4.16. Generated vs. Observed PDF examples for a station 083163 in Ft. Lauderdale, FL for September for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).....	239
Figure 4.17. Generated vs. Observed PDF examples for a station 311975 in Concord, NC for September for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).	242
Figure 4.18. Generated vs. Observed PDF examples for a station 283516 in Greenwood Lake, NJ for September for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).....	245
Figure 4.19. Boxplots of January Wet Spell Frequencies for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).....	248

Figure 4.20. Boxplots of January Dry Spell Frequencies for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).....	249
Figure 4.21. Boxplots of July Wet Spell Frequencies for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).	250
Figure 4.22. Boxplots of July Dry Spell Frequencies for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).	251
Figure 4.23. RMSE of the WGEN and GiSTR simulations by month for the correlation matrix of precipitation (ρ) (a), the correlation matrix of precipitation events(ρ_{ev}) (b), the correlation matrix of precipitation amounts (ρ_{am}) (c), and the correlation matrix of precipitation extreme events (ρ_{ex}) (d).....	252
Figure 4.24. Average Decay Function Correlogram for WGEN and GiSTR compared to observations for January for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 1979-2000 time period.	253
Figure 4.25. Average Decay Function Correlogram for WGEN and GiSTR compared to observations for July for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 1979-2000 time period.	254
Figure 4.26. RMSE Comparison between downscaling simulations for the mean daily nonzero precipitation (μ).	255
Figure 4.27. Difference between generated and observed values of the mean daily nonzero precipitation (μ) for the WGEN simulation in April.	256
Figure 4.28. RMSE Comparison between downscaling simulations for the variance of the daily nonzero precipitation (σ^2).	257
Figure 4.29. Difference between generated and observed values of the variance of the daily nonzero precipitation (σ^2) for the WGEN simulation in September.	258
Figure 4.30. RMSE Comparison between downscaling simulations for the average total precipitation ($E[S(T)]$).	259
Figure 4.31. RMSE Comparison between downscaling simulations for the inter-annual variability ($Var[S(T)]$).	259

Figure 4.32. Difference between generated and observed PDF of nonzero precipitation for station 097847 in Savannah, GA for the WGEN simulation in April.	260
Figure 4.33. Difference between generated and observed PDF of nonzero precipitation for a station 080945 in Bradenton, FL for the WGEN simulation in September.....	261
Figure 4.34. Difference between generated and observed values of the mean daily nonzero precipitation (μ) for the WGEN downscale simulation in January.	262
Figure 4.35. RMSE Comparison between downscaling simulations for the mean daily nonzero precipitation (μ) with outliers removed.	263
Figure 4.36. RMSE Comparison between downscaling simulations for the average total precipitation ($E[S(T)]$) with outliers removed.	263
Figure 4.37. RMSE Comparison between downscaling simulations for the variance of the daily nonzero precipitation (σ^2) with outliers removed.	264
Figure 4.38. RMSE Comparison between downscaling simulations for the inter-annual variability ($Var[S(T)]$) with outliers removed.	264
Figure 4.39. Boxplots of the frequency of wet spells of various lengths for January for observations (a), WGEN simulation (b), and GiSTR simulation (c).....	265
Figure 4.40. Boxplots of the frequency of dry spells of various lengths for January for observations (a), WGEN simulation (b), and GiSTR simulation (c).....	266
Figure 4.41. Boxplots of the frequency of wet spells of various lengths for July for observations (a), WGEN simulation (b), and GiSTR simulation (c).....	267
Figure 4.42. Boxplots of the frequency of dry spells of various lengths for July for observations (a), WGEN simulation (b), and GiSTR simulation (c).....	268
Figure 5.1. Cross Validation station map for 2001-2009.....	322
Figure 5.2. RMSE of each downscaling simulation by month for the correlation matrix of precipitation (ρ) (a), the correlation matrix of precipitation events (ρ_{ev}) (b), the correlation matrix of precipitation amounts(ρ_{am}) (c), and the correlation matrix of precipitation extreme events(ρ_{ex}) (d).....	323

Figure 5.3. Average Decay Function Correlogram for each simulation compared to observations for January for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 2001-2009 time period. 324

Figure 5.4. Average Decay Function Correlogram for each simulation compared to observations for July for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 2001-2009 time period. 325

Figure 5.5. RMSE Comparison between 2001-2009 downscaling simulations for mean daily nonzero precipitation (μ). 326

Figure 5.6. Difference between generated and observed values of mean daily nonzero precipitation (μ) for each of the 2001-2009 downscaling simulations: LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for January. 327

Figure 5.7. Difference between generated and observed values of mean daily nonzero precipitation (μ) for each of the 2001-2009 downscaling simulations: LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for June..... 332

Figure 5.8. Difference between generated and observed values of mean daily nonzero precipitation (μ) for each of the 2001-2009 downscaling simulations: LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for September..... 337

Figure 5.9. RMSE Comparison between 2001-2009 downscaling simulations for the variance of the daily nonzero precipitation (σ^2). 342

Figure 5.10. Difference between generated and observed values of the variance of the daily nonzero precipitation (σ^2) for each of the 2001-2009 downscaling simulations: LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for April. 343

Figure 5.11. Difference between generated and observed values of the variance of the daily nonzero precipitation (σ^2) for each of the 2001-2009 downscaling simulations: LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for September..... 348

Figure 5.12. Difference between generated and observed values of the variance of the daily nonzero precipitation (σ^2) for each of the 2001-2009 downscaling simulations: LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for October. 353

Figure 5.13. RMSE Comparison between 2001-2009 downscaling simulations for the average total precipitation ($E[S(T)]$).....	358
Figure 5.14. RMSE Comparison between 2001-2009 downscaling simulations for the inter-annual variability ($Var[S(T)]$).....	358
Figure 5.15. Observed vs. Generated PDFs of nonzero precipitation for station 090969 in Blairsville, GA for January for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).....	359
Figure 5.16. Observed vs. Generated PDFs of nonzero precipitation for station 319457 in Wilmington, NC for January for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).....	362
Figure 5.17. Observed vs. Generated PDFs of nonzero precipitation for station 286026 in Newark, NJ for January for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).....	365
Figure 5.18. Observed vs. Generated PDFs of nonzero precipitation for station 090969 in Blairsville, GA for September for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).....	368
Figure 5.19. Observed vs. Generated PDFs of nonzero precipitation for station 319457 in Wilmington, NC for September for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).....	371
Figure 5.20. Observed vs. Generated PDFs of nonzero precipitation for station 286026 in Newark, NJ for September for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).....	374
Figure 5.21. Observed vs. Generated PDFs of nonzero precipitation for station 092283 in Cornelia, GA for September for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).....	377
Figure 5.22. Boxplots of January Wet Spell Frequencies in 2001-2009 for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).....	380
Figure 5.23. Boxplots of July Wet Spell Frequencies in 2001-2009 for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).....	381

Figure 5.24. Boxplots of January Dry Spell Frequencies in 2001-2009 for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).	382
Figure 5.25. Boxplots of July Dry Spell Frequencies in 2001-2009 for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).	383
Figure 5.26. RMSE of the WGEN and GiSTR downscaling simulation by month for the correlation matrix of precipitation (ρ) (a), the correlation matrix of precipitation events (ρ_{ev}) (b), the correlation matrix of precipitation amounts(ρ_{am}) (c), and the correlation matrix of precipitation extreme events(ρ_{ex}) (d). Time period of 2001-2009.	384
Figure 5.27. Average Decay Function Correlogram for WGEN and GiSTR compared to observations for January for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 2001-2009 time period.	385
Figure 5.28. Average Decay Function Correlogram for WGEN and GiSTR compared to observations for July for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 2001-2009 time period.	386
Figure 5.29. RMSE Comparison between downscaling simulations for mean daily nonzero precipitation (μ). 2001-2009 time period.....	387
Figure 5.30. Difference between generated and observed values of mean daily nonzero precipitation (μ) for the WGEN simulation (a) and the GiSTR simulation (b) in April. 2001-2009 time period.....	388
Figure 5.31. RMSE Comparison between downscaling simulations for the variance of daily nonzero precipitation (σ^2). 2001-2009 time period.....	391
Figure 5.32. Difference between generated and observed values of the variance of daily nonzero precipitation (σ^2). for the WGEN simulation (a) and the GiSTR simulation (b) in September. 2001-2009 time period.	392
Figure 5.33. RMSE Comparison between downscaling simulations for the average total precipitation ($E[S(T)]$). 2001-2009 time period.....	395
Figure 5.34. RMSE Comparison between downscaling simulations for the inter-annual variability($Var[S(T)]$). 2001-2009 time period.....	395

Figure 5.35. Difference between generated and observed values of the mean daily nonzero precipitation(μ) (a) and the variance of daily nonzero precipitation (σ^2). (b) in January for the WGEN downscale simulation. 2001-2009 time period..... 396

Figure 5.36. Difference between generated and observed PDF of nonzero precipitation for station 091345 in Brunswick, GA for the WGEN simulation in April..... 399

Figure 5.37. Difference between generated and observed PDF of nonzero precipitation for a station 092485 near Lake Wales, FL for the WGEN simulation in September..... 399

Figure 5.38. RMSE Comparison between downscaling simulations for the mean daily nonzero precipitation (μ). 2001-2009 time period. Outlier stations removed. 400

Figure 5.39. RMSE Comparison between downscaling simulations for the average total precipitation ($E[S(T)]$). 2001-2009 time period. Outlier stations removed. 400

Figure 5.40. RMSE Comparison between downscaling simulations for the variance of daily nonzero precipitation (σ^2).. 2001-2009 time period. Outlier stations removed. 401

Figure 5.41. RMSE Comparison between downscaling simulations for the inter-annual variability($Var[S(T)]$). 2001-2009 time period. Outlier stations removed. 401

Figure 5.42. Boxplots of the frequency of wet spells of various lengths for January for observations (a), WGEN simulation (b), and GiSTR simulation (c). 2001-2009 time period. 402

Figure 5.43. Boxplots of the frequency of wet spells of various lengths for July for observations (a), WGEN simulation (b), and GiSTR simulation (c). 2001-2009 time period. 403

Figure 5.44. Boxplots of the frequency of dry spells of various lengths for January for observations (a), WGEN simulation (b), and GiSTR simulation (c). 2001-2009 time period. 404

Figure 5.45. Boxplots of the frequency of dry spells of various lengths for July for observations (a), WGEN simulation (b), and GiSTR simulation (c). 2001-2009 time period. 405

Figure 5.46. Observed values of the variance of daily nonzero precipitation (σ^2) for the 1979-2000 time period across the domain for September..... 406

Figure 5.47. Observed values of the variance of daily nonzero precipitation (σ^2) for the 2001-2009 time period across the domain for September..... 407

Figure 6.1. Observed values of the variance of daily nonzero precipitation (σ^2) for the 1979-2000 time period (a) and the 2001-2009 time period (b)..... 429

Figure 6.2. Observed values of the variance of daily nonzero precipitation (σ^2) for the 1991-2000 time period (a) and the 1981-1990 time period (b)..... 432

Chapter 1. Introduction

Precipitation has a strong influence on environmental resources, including water, agricultural, natural resources. Management of these resources benefits from forecasts of precipitation at seasonal or monthly time scales for decision making (particularly with regards to irrigation and decisions regarding water restrictions) and crop and streamflow modeling. Local daily precipitation data are used in crop yield models such as the CERES model (Tsvetsinskaya et al, 2003). In addition, accurate forecasts of seasonal precipitation can allow farmers and water managers to plan for potential severe droughts, such as the 2007 drought in the southeast United States which resulted in estimated crop losses of \$1.3 billion (Ding et al, 2008). However, most global climate models (GCMs) which are used to predict seasonal precipitation have resolutions of one to several degrees latitude and longitude (several hundred kilometers), which is too coarse to provide the local forecasts of precipitation needed for decision making and modeling. To address this limitation of GCMs, global scale information about precipitation is translated to the local scale through the downscaling.

1.1. Downscaling Background

Downscaling is used in many studies to translate global scale information to the local scale. A common definition of downscaling is described by Benestad (2008) – “the process of making the link between the state of some variable representing the large space and the state of some variable representing a much smaller space”. This general definition of downscaling has led to the development of several techniques for handling this process, which can be broken down into two basic groups. The first of these groups is dynamic or numerical downscaling, and the second group is known as empirical statistical downscaling (ESD).

1.1.1. Dynamic Downscaling

Dynamic or numerical downscaling makes use of all the model physics which are primarily used in numerical weather and climate modeling. The most common techniques for dynamic downscaling involve initializing regional climate models (RCMs, e.g. Barstad et al, 2009; Frei et al, 2003, Qian et al, 2009; Pesquero et al, 2010) or nested limited area models (LAMs, e.g. Giorgi, 1990; Frogner et al, 2006; Tudor and Termonia, 2010) with the output of the desired GCM to be downscaled (Varis et al, 2004). In addition, dynamic downscaling also includes the use of high resolution or variable resolution atmospheric GCMs (AGCMs, e.g. Fox-Rabinovitz et al, 2001; May and Roeckner, 2001; Fox-Rabinovitz et al, 2006). While among the most complex and computationally expensive techniques, the AGCMs are noted in several studies to provide the most globally consistent results. Although AGCMs are globally consistent, significant underlying errors can be present in the AGCM (Varis et al, 2004; Christensen et al, 2001). In contrast, the RCM and LAM approaches to dynamic downscaling are capable of producing multi-decadal simulations and describe climate feedback mechanisms acting on the regional scale with less computational expense than AGCMs (Varis et al, 2004). Both RCMs and nested LAMs are driven by the GCM to be downscaled, and have several limitations. First, the RCMs and LAMs inherit the systematic error in the driving fields provided by the GCM. Second, there is a lack of two-way interactions between the regional and global scale climate. Finally, there are algorithmic limitations of the lateral boundary interface between the two models (Varis et al, 2004; Christenson and Christenson, 2001). While several recent studies (e.g. Frei, 2003; Rummukainen et al, 2001; Qian et al, 2009) have worked to address these limitations and coupling with hydrological and earth system models, the primary limitations of dynamic downscaling remain the influence of systematic errors on the results and the computational expense. In addition, Benestad et al (2008) describe several other drawbacks of dynamic downscaling related to the numerical models involved. First, the parameterization schemes

used are created based on present climate and therefore may not be valid in future climates, particularly with regard to climate change studies. Second, numerical instabilities result from the discrete values on a discrete grid in a model, which give some of the error from dynamic downscaling. While dynamic downscaling has several advantages and disadvantages, empirical statistical downscaling contains multiple techniques which have their own advantages and disadvantages.

1.1.2. Empirical Statistical Downscaling

Empirical Statistical Downscaling (ESD) refers to a group of downscaling techniques which use statistical methods and observed climate data to build the relationship between local scale and global scale. While many techniques can be defined as ESD techniques, they share many common advantages. First, they are more computationally efficient than dynamic downscaling techniques. Second, they are based on standard and accepted statistical procedures. Third, ESD techniques may be flexibly crafted for specific purposes. Finally, ESD techniques are able to incorporate historical climate information for the desired region (Von Storch et al, 2000; Varis et al, 2004). There are also several disadvantages to ESD techniques. First, these techniques assume that the statistical relationships will be unchanged in a future climate. Second, they require a long (> 20 years) and reliable data series to allow the statistical relationships to be robust. Third, like dynamic downscaling techniques, ESD techniques are also affected by the errors in GCM used (Goodess et al, 2001; Varis et al, 2004). These advantages and disadvantages are common to all ESD techniques, which are broken down into three categories; transfer functions, weather typing and weather generators.

Transfer functions or regression methods are the simplest ESD techniques used to determine the relationship between large area and site specific surface climate data or large-scale upper air data and local surface climate data. These techniques include linear and multiple linear regression, canonical correlation analysis, principal components analysis (or empirical orthogonal functions), artificial neural networks and kriging

(Barrow, 2002; Wilby et al, 2004; e.g. Benestad, 2007; Busuioc et al, 2001; Heyen et al, 1996; Hoar and Nychka, 2008; Widmann et al, 2003).

Weather typing techniques function similarly to transfer functions. However, weather typing focuses on determining the relationship between atmospheric circulation types and local weather patterns. The circulation types or weather classes, can be determined objectively through cluster analysis (e.g. Huth, 2000; Kidson, 2000; Hewiston and Crane, 2002, Corte-Real et al, 1999) or subjectively through various published schemes (e.g. Bardossy and Caspary, 1990; Jones et al, 1993). Weather type techniques include analogue methods, self organizing maps and Monte Carlo techniques (Wilby et al, 2004. e.g. Conway and Jones 1998; Timbal and McAvaney, 2001; Zorita and Von Storch, 1999; Hughes et al, 1993).

Weather generators are models which replicate the statistical attributes of a local climate, but do not replicate the observed sequences of events (Wilby et al, 2004; Wilks and Wilby, 1999). Most commonly, these generator make use of first order Markov chains (e.g. Richardson and Wright, 1984) or the probability of dry and wet spells of various lengths (e.g. Racsco et al, 1991).

Each of the categories of ESD techniques have multiple advantages and disadvantages, which are summarized in Table 1.1 (Barrow, 2002; Wilby et al, 2004). Transfer functions are the least computationally expensive of downscaling techniques, can produce numerous ensembles easily, employ the full possible range of predictors and have several “off the shelf” solutions and software readily available to most users. Given these advantages, transfer functions also suffer from the largest number of disadvantages among downscaling techniques including the assumption of linearity or normality and poor representation of extreme events. Weather typing approaches also have several advantages including their versatility and the ability to yield the most physically interpretable linkages to surface climates. However, weather typing requires the added task of weather classification, the scenarios produced are insensitive to future climate forcing, and may not capture intra-type variations in surface climate. Like transfer functions, weather generators are also able to produce a large number of ensembles for uncertainty analysis, but can also provide

representative time series of unlimited length in regions with sparse data. In addition, weather generators are also capable of producing daily or sub-daily output and the parameters can be altered to reflect a future climate for changes in both the mean and variability. Most weather generators suffer from two distinct disadvantages; the lack of consideration of spatial structure and representation of extreme events. Given the ability to generate daily information for a large number of ensembles of infinite length along with the other advantages, weather generators are used for downscaling in this study. In addition, the two main disadvantages of weather generators are also addressed as part of this research. Further background on the structure and various weather generator techniques is discussed in the next section.

1.1.2.1. Overview of Weather Generators

While there are multiple basic structures of weather generators, most weather generators begin by generating precipitation events using Markov chains. A Markov chain is a common model used to represent time series of discrete variables. These chains can be imagined as being based on a collection of states of a system. For each time period, the length of which is equal to the separation between observations, the Markov chain can remain in the same state or change to another state (Wilks, 2006). For the modeling of precipitation events, the Markov chain approach is limited to two states, rain or no rain. The behavior of this Markov chain is governed by conditional probabilities known as transition probabilities. A transition probability, like all conditional probabilities, is the probability of an event given that another event has already occurred e.g. the probability of event A occurring given that event B has already occurred. Transition probabilities are conditional probabilities which refer to the states of the Markov chain for the current time step and the previous time step, or for example with precipitation events:

$$P_{01} = \Pr\{X_t = 1 \mid X_{t-1} = 0\} \quad (1.1)$$

Where X represents the state of the Markov chain for day t and day $t-1$, and a value of $X=1$ indicates rain for the given day and a value of $X=0$ indicates no rain for the given day.

Therefore, P_{01} is a transition probability, i.e. the probability that rain will occur on day t given that it did not rain the previous day. For transition probabilities which only consider the previous observation, such as in the example, there are three other probabilities involved: P_{00} , P_{10} and P_{11} which are the probability of a dry day following a dry day, the probability of a dry day following a rainy day, and the probability of a rainy day following a rainy day respectively. Markov chains with two states consider the previous time step (i.e. the transition probabilities are conditioned only on the previous time step) are known as two-state first order Markov chains. While some weather generators (Richardson and Wright, 1984; Wilks and Wilby, 1999) use these Markov chains to help with event generation, other generators use second or third order Markov chains, which consider the last two or three time steps respectively.

Regardless of what order is used, all Markov chain generators have a similar structure for generating precipitation. A flowchart describing this process is provided in Figure 1.1. Each generator begins by generating a uniform random number (u) between 0 and 1. This random number is compared to the appropriate transition probability (p_c , in this example), and if the $u < p_c$ the generator will produce a rain event at that location for that day. If a rain event is generated, the generator moves on to generate precipitation amounts using statistical distributions fit to all nonzero precipitation amounts. Among Markov chain generators one of the oldest generators is WGEN (Richardson and Wright, 1984). This generator makes use of the two state first order Markov chain to generate precipitation events and a two parameter gamma distribution for precipitation amounts, and is also used as a guide for building other Markov chain generators.

In contrast to Markov chain generators, spell length generators simulate precipitation events based on the probability of dry and wet spells of various lengths. Rather than

simulating individual days, the spell length generator alternates between generating dry and wet spells based on the relative frequency of wet and dry spell lengths. For instance, a spell length generator would generate a wet spell followed by a dry spell. Also, the spell length generator does not generate a new spell before finishing the previous spell generation (Wilks and Wilby, 1999). The oldest example of these weather generators is LARS-WG, which has been used for multiple downscaling studies (Racsko et al, 1991; Semenov and Barrow, 1997). A flowchart describing the generation process of a spell-length generator is shown in Figure 1.2. Spell length generators begin by using the probability of spells of various lengths to determine if a string of days to determine the spell length (L) and if the spell is a dry or wet spell. For each day of the spell L is reduced by one until L returns to zero. At this point the generator creates another value of L and generation proceeds with a spell of the opposite type. That is, if the previous spell was wet, the next spell will be dry. If a given day falls into wet spell, the generator will then proceed to the amount generation process for that day, otherwise no precipitation is generated for each day in a dry spell. Nonzero precipitation amounts are generally right skewed (such as the example shown in Figure 1.3). As such, weather generators do not model precipitation amounts with a normal distribution, but with exponential distributions. Specifically, most weather generators use either the exponential, mixed exponential or the two parameter gamma distribution. While the exponential distribution is the simplest of the three, the two parameter gamma distribution is used more often since it allows for more flexible accommodation of amount frequencies, improving the realism of the weather generator output (Wilks and Wilby, 1999). The amount generation process functions similarly to the event generation process for both types of weather generators. A standard normal random number is generated, which is converted to the precipitation amount using the cumulative distribution functions of the standard normal distribution and the chosen distribution fitted to observed precipitation amounts (Wilks and Wilby, 1999; Wilks, 2008; Richardson and Wright, 1984).

Most weather generators begin by generating precipitation, but also include processes to simulate temperature and solar radiation using auto-regressive (AR) modeling and assuming precipitation as a surrogate for cloud cover. While these approaches can produce

temperature and solar radiation values with little computational expense, recent studies have shown that AR modeling can produce negative diurnal temperature ranges and produce temperature values that are more normally distributed than observed temperature distributions (Schoof et al, 2005; Schoof, 2008; Harmel et al, 2002). While precipitation, temperatures and solar radiation can be downscaled using weather generators, the focus of this study will be on the primary function of most weather generators: generating and downscaling precipitation.

In general, weather generators have several input parameters which define the observed statistics of precipitation events and amounts. These parameters can vary by month or by season and for simulations in the current time period the observations from that period are used to estimate these parameters. For downscaling, these input parameters are conditioned upon information from the GCM. The conditioning process has two different scaling relationships which are used to condition the weather generator:

$$F(P_{GCM})_{future} - F(P_{GCM})_{present} = F(P_{station})_{future} - F(P_{station})_{present} \quad (1.2)$$

$$\frac{F(P_{GCM})_{future}}{F(P_{GCM})_{present}} = \frac{F(P_{station})_{future}}{F(P_{station})_{present}} \quad (1.3)$$

Where P is one of the parameters of the weather generator, either for the GCM grid point (P_{GCM}) or for an individual location ($P_{station}$), and $F(P)$ is some function of the parameters of the weather generator. The first relationship assumes an absolute change in the parameters, while the second assumes a relative change. In both cases, the change at each scale is assumed to be the same. In general, the event based parameters are conditioned using the first relationship and amount based parameters are conditioned using the second relationship (Chen et al, 2006). However, in many cases the present day information from the GCM is not used to represent the weather generator parameters in the large scale for the present day. Given all the approximations used in forming any GCM, the GCM modeled climate is

slightly different than the observed climate as shown by observations, even when compared to area averages. Therefore, the present climate at the large scale is often not the GCM data, but an area average or another extrapolation of the station data to each GCM grid point in the domain (Wilks, 1999; Wilby et al, 1998). The stations used in the area average are matched to the appropriate GCM grid point that each of the relationship can be solved for each individual location for the future time (Chen et al, 2006). While the parameters for individual weather generators are slightly different the downscaling process is similar. However, while the process is similar among different weather generators, this study will highlight and evaluate alternatives to some of the key assumptions made in the general approach to weather generator downscaling.

WGEN, like most weather generators, generates precipitation for individual locations, with no regard for the spatial structure of precipitation events and amounts. Therefore, WGEN is capable of generating the temporal structure of precipitation, but does not consider the spatial structure of precipitation. Recently, Baigorria and Jones (2010) developed the geo-spatial temporal weather generator (GiST), which is a Markov chain generator similar to WGEN. However, GiST considers the spatial structure of precipitation event and amounts and handles multiple locations. Several other weather generators are documented including Met&Roll (Dubrovsky, 1997; Wilks, 1992), SIMMETEO (Geng et al, 1986), CLIGEN (Nicks and Gander, 1994) and MARKSIM (Jones and Thornton, 2000). Each of these generator are Markov Chain generators, but either use multiple order chains or make use of additional predictors such as upper-air parameters. However, none of these weather generators consider both the spatial and temporal components of precipitation. While GiST accounts for both the spatial and temporal structure, the question of accuracy still remains across a broad region such as the southeastern US. In addition, several studies note that weather generators tend to underestimate the occurrence of heavy (or rare) precipitation events (Wilby et al, 2004; Barrow, 2002; Furrier and Katz, 2008). Therefore, this study will focus on evaluating GiST against WGEN and a proposed improvement to the GiST weather generator to address the heavy precipitation issue.

1.1.2.2. The Main Concerns with ESD

Regardless of which ESD techniques are chosen for downscaling, there are four concerns which need to be considered when evaluating the results, particularly with regards to the uncertainty of the technique. These four concerns are best described by Lewiston and Crane (2006) and summarized below:

1. Synoptic-scale vs. Local-scale forcing – While synoptic-scale forcing has an influence on local climate, there is also a degree of fixed and variable local forcing. Fixed forcings include topography and land-water boundaries, while variable forcings include land use and land cover change. GCMs cannot capture this small-scale variability and may be parameterized depending on the RCM. However, for some RCMs and for ESD, these small-scale forcings become stochastic permutations of the response to a given synoptic state.
2. Stationarity – Given that there is a relationship between the global and local scales, ESD is feasible with present climates. However, ESD techniques assume that relationship between scales in the present climate is also true in future climates (a particular problem for transfer functions). Physically, this suggests that changes in a future climate are primarily in the timing, persistence and frequency of large-scale events. This is particular issue with climate change studies, but it is a reasonable assumption for seasonal forecasting.
3. Predictor variables – The choice of predictors is important to ESD. The predictors need to reflect a physical relationship rather than relationships with no clear physical relationship. Since the predictors are taken from the GCMs, additional details that are important include having a strong relationship, that the predictor is well represented by the GCM, and that predictor responds to change similarly to the predictand (Benestad et al, 2008)
4. Temporal resolution – In addition to downscaling time mean fields, there must be consideration given to the fact that the future climate may change the histogram of

events without changing the mean, i.e. the variability of precipitation amounts change with no change to the mean amounts.

While these concerns have some impact to all types of ESD, the descriptions of the various types suggest that some of these are more of a concern for specific techniques. As mentioned previously, downscaling with weather generators produce output on daily or sub-daily timescales and allow that both the mean and variance of the variable of interest (precipitation in this case) can be conditioned for the response to a future climate. This allows downscaling with weather generators to have less trouble with temporal resolution, than other ESD techniques. Stationarity and forcing issues are concerns in ESD regardless of the type to technique considered. However for most weather generators, stationarity becomes a concern during the conditioning process, which primarily influences temporal parameters, i.e. the timing, frequency and persistence of events. Benestad et al (2008) argue that stationarity is also a concern for dynamic downscaling (particularly in climate change studies) since the parameterization schemes used in GCMs and RCMs may not be valid in a changed climate. Predictor variables are a concern primarily for transfer functions and weather typing techniques, which often make use of multiple predictor variables. However, certain weather generators, such as Met&Roll (Dubrovsky, 1997), also make use of multiple predictors from the GCM.

While using GiST for downscaling raises the same concerns as all ESD techniques, the advantages it has compared to other techniques, such as the direct consideration of spatial structure (which addresses one of the main disadvantages of most weather generators), allow unique opportunities for downscaling. Before returning to the motivation for evaluating different weather generators and different downscaling approaches, the different spatial structures of precipitation and their relation to climate in southeast United States is discussed in the next section.

1.2. Seasonal Precipitation Patterns in the Southeast U.S.

The southeast United States is an area with diverse topography and a climate described as warm-temperate or subtropical with no dry season and mild winters. The exception to this is the southern tip of Florida which is described as a tropical savannah climate (Oliver and Hidore, 2002; White et al, 1998). This region is subject to some of the warmest climate conditions in the United States, with its highest annual precipitation along the Gulf of Mexico and in south Florida (Baigorria et al, 2007a). While the southeast (with the exception of Florida) has no distinct wet or dry seasons there are two distinct dominant processes which cause rainfall in this region. During the cold season, precipitation in the southeast is caused by frontal and large upper-level storm systems coming out of the Midwest and Great Plains. These processes produce precipitation over a wide area in a northeast-southwest direction, as shown by the spatial pattern of typical winter precipitation in Figure 1.4. The correlations among stations in the region follow a northeast to southwest direction perpendicular to the path of most fronts through the region. During warm months, the dominant causes of precipitation are tropical cyclones and largely unstructured convection. Unstructured convection produces precipitation in seemingly more random locations (for example, the spatial pattern depicted in Figure 1.5), while tropical cyclones produce intense precipitation over smaller areas than frontal processes and larger areas than unstructured convection. For these causes of precipitation, correlations between weather stations are concentric and decrease rapidly over short distances (Baigorria et al, 2007a, 2007b).

Forecasting seasonal precipitation at a local scale is a challenge for most GCMs because the two dominant precipitation processes in the region occur at a finer resolution than the GCM. Most GCMs have a resolution of one or more degrees (several hundred kilometers), while the spatial structure of precipitation (as shown by weather stations) is often less than 10 km. Therefore the processes that produce precipitation are represented by sub-grid parameterization schemes. As described previously, precipitation in warm months is primarily produced through convective (both structured and unstructured) processes. In

terms of numerical modeling (GCM or weather forecasting), this means repeated calls to these parameterization schemes which are the weakest part of the model. Multiple studies have shown the issues of such parameterization schemes, particularly with regards to convective precipitation and cloud radiation interaction (e.g. Palmer, 1996). In addition, Benestad et al (2008) note that the equations used in parameterizations provide bulk descriptions of sub-processes and land surface processes, but not an exact representation. They also point out that in GCMs and RCMs these schemes inhibit overly strong convection or filter out physically impossible solutions, resulting in model solutions that are close to the real world, but also imply a lack of physical consistency. These issues with GCM and RCM parameterizations are sources of error in forecasting precipitation at monthly and seasonal timescales at local scales. A finer resolution GCM would capture the details of processes that produce precipitation, but there would be a dramatic increase in the computational expense compared to current GCM runs.

1.3. Motivation and Objectives

Given the problems that are evident with forecasting precipitation from GCMs, downscaling becomes necessary to handle precipitation forecasts with finer detail than GCMs can provide. However, given that RCMs have similar parameterization schemes and the other disadvantages of dynamic downscaling described previously, ESD becomes an ideal choice. ESD techniques reduce the error with regard to parameterization schemes, and have been shown to produce results with similar skill to dynamic downscaling techniques (e.g. Hellstrom et al, 2001; Kidson and Thompson, 1998; Hanssen-Bauer et al, 2003), with less computational expense than dynamic techniques. Effective precipitation forecasts resulting from ESD can be used by hydrologists and water resource managers to model water quality and reservoir levels. In addition, crop modelers can make use of these downscaled forecasts to aid in projections of crop yields. These projections can allow for decision making with regards to the water management, irrigation and crop insurance. Since weather

generators produce output at the daily timescale they can provide direct inputs of precipitation for both applications for both climate change studies and monthly/seasonal forecasting.

Given the potential uses for multiple sectors it becomes necessary to evaluate the accuracy and uncertainty associated with all parts of weather generator downscaling. Baigorria and Jones (2010) have already shown that GiST produces similar results to WGEN for precipitation in limited areas of Florida and North Carolina. This study will expand that research to evaluate each generator for the state of North Carolina and for the Southeast United States. In addition, this study will evaluate an adjustment to the GiST generator to account for extreme precipitation. Finally, while the generators themselves will be evaluated, several variations on the assumptions used in the scaling relationship for downscaling weather generators will also be evaluated for the Southeast United States. Specifically the following questions will be address in this study:

- What are the strengths and weaknesses of each generator in each domain?
- What spatial patterns (if any) are present in the error of each generator?
- What are the strengths and weaknesses of different approaches used to downscaling with weather generators?
- Are there seasonal/spatial trends in the error of each approach?
- What are the possible physical causes for these errors? That is, what climatological phenomena may not be captured by each generator/approach?

Chapter 2 addresses the details of the GiST weather generator, the proposed improved version of GiST, and downscaling with weather generators in addition to the experiment design. Chapter 3 focuses on evaluating three weather generators in the state of North Carolina and across the Southeast United States. Chapters 4 and 5 discuss the evaluation of different downscaling approaches with weather generators, while Chapter 6 presents the conclusions of this study. The ultimate goal of this study is to determine which weather

generator and downscaling variation is most appropriate for downscaling seasonal forecasts of precipitation.

Table 1.1. Advantages and Disadvantages of the three categories of ESD techniques

Category	Advantages	Disadvantages
<i>Transfer Functions</i>	<ul style="list-style-type: none"> • Much less computationally demanding than dynamic downscaling • Ensembles of high resolution climate scenarios can be produced easily • Employs the full range of available predictors • ‘Off the shelf’ solutions and software available 	<ul style="list-style-type: none"> • Large amounts of obs. required • Specialist knowledge required to apply correctly • Relationships only valid during the calibration period • May not be possible to derive significant relationships for some variables • A predictor which may not appear as the most significant may be significant in a changed climate • Poor representation of observed variance and extreme events • May assume linearity or normality
<i>Weather Typing</i>	<ul style="list-style-type: none"> • Founded on sensible linkages between climate on the large scale and weather on the local scale • Versatile (e.g. applicable to surface climate, air quality, etc.) • Compositing for analysis of extreme events • Yields physically interpretable linkages to surface climate 	<ul style="list-style-type: none"> • Differences in the relationships between weather type and local weather have occurred at some sites during the observed record • Scenarios produced are insensitive to future climate forcing • Requires the added task of weather classification • May not capture intra-type variations in surface climate
<i>Weather Generators</i>	<ul style="list-style-type: none"> • Able to generate time series of unlimited length • Able to obtain representative weather time series in regions of sparse data • The parameters of the weather generator can be altered to reflect a future climate, for both variability and mean changes • Large ensembles for uncertainty analysis • Can generate daily and sub-daily information 	<ul style="list-style-type: none"> • Seldom able to describe all aspects of climate, especially persistent, rare or decadal or century-scale events • Designed to be used independently at individual location, meaning that few account for spatial correlation • Unanticipated effects to secondary variables of change precipitation parameters.

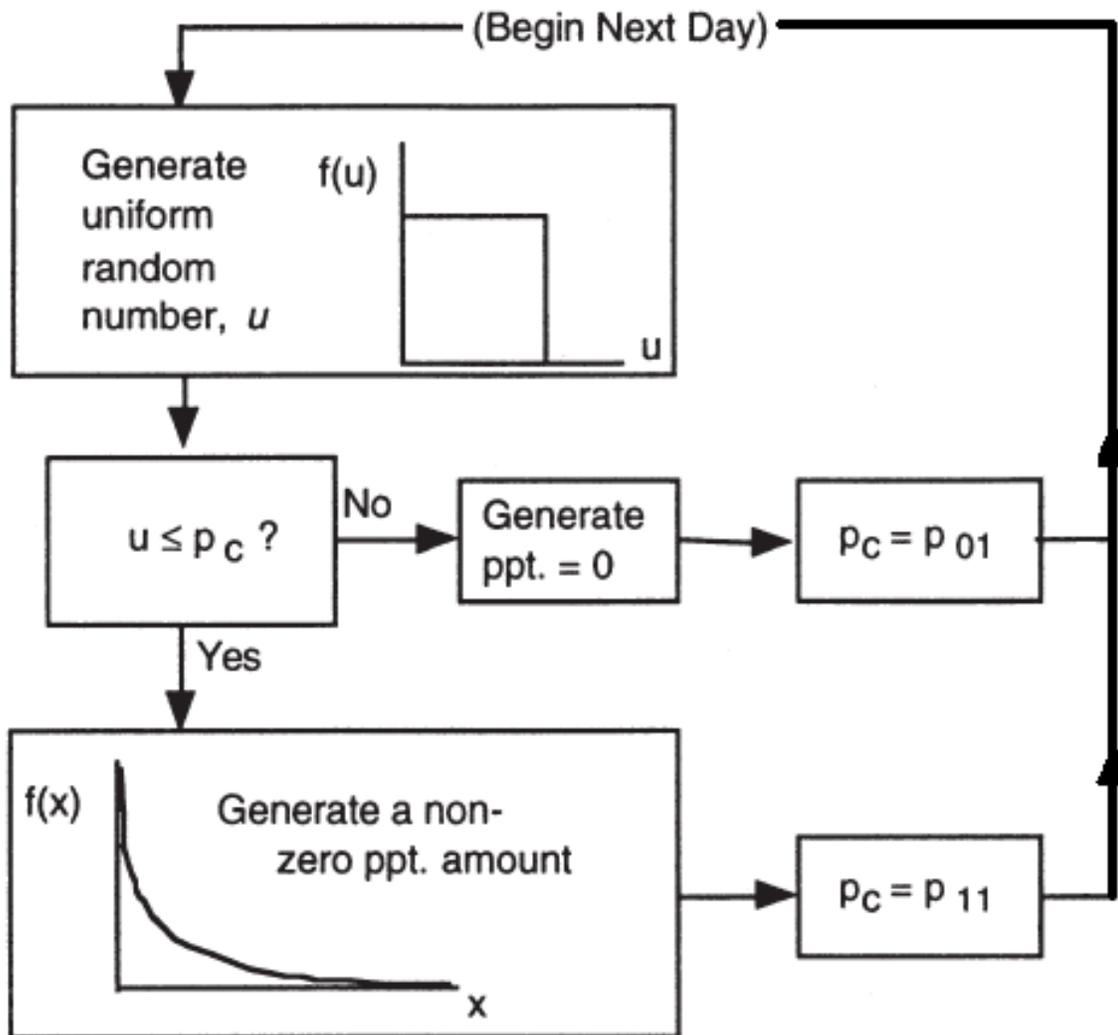


Figure 1.1. Flowchart for daily precipitation generation using Markov chain generators. Adapted from Wilks and Wilby (1999).

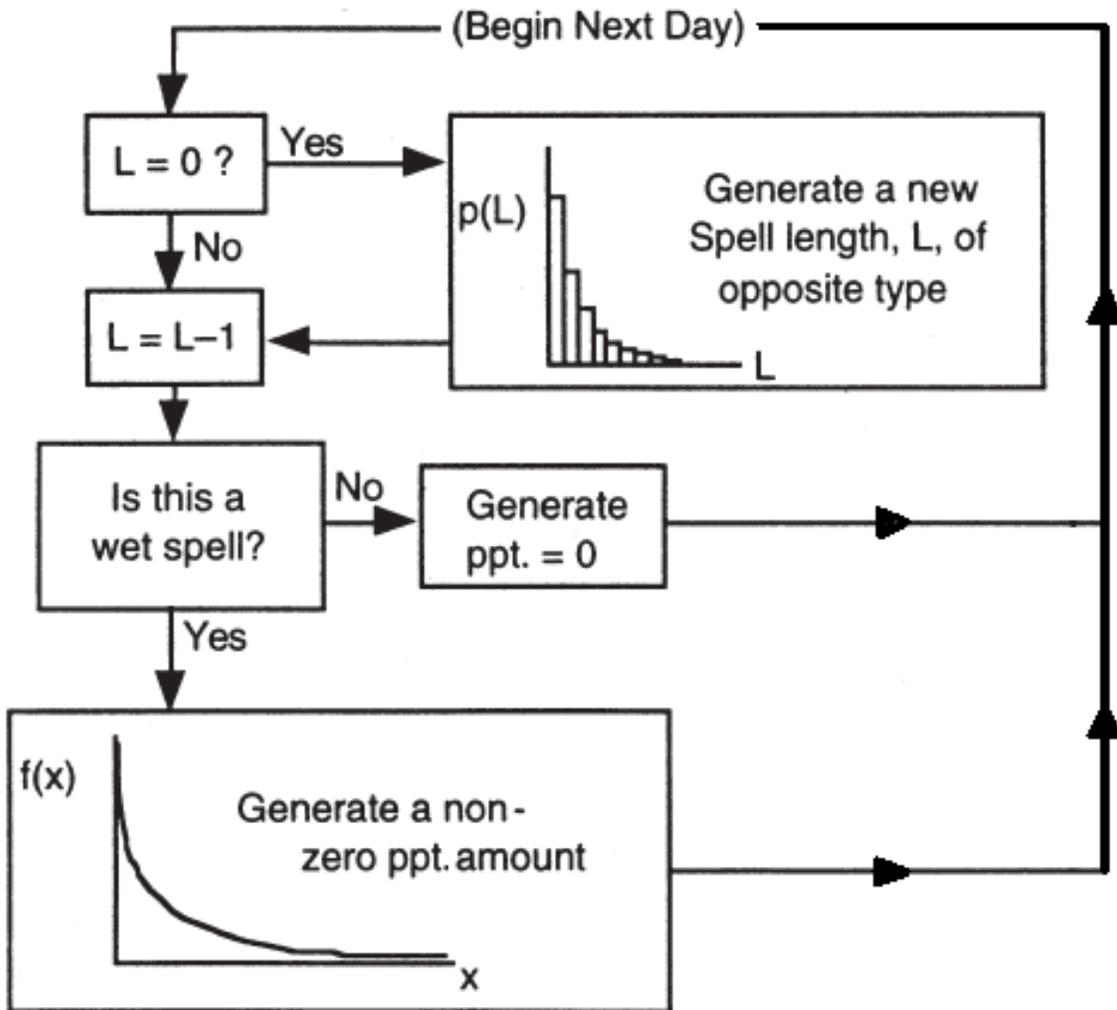


Figure 1.2. Flowchart for daily precipitation generation using spell-length generators. Adapted from Wilks and Wilby (1999).

Histogram of Historical Daily Nonzero Precipitation at Raleigh-Durham Airport (KRDU)

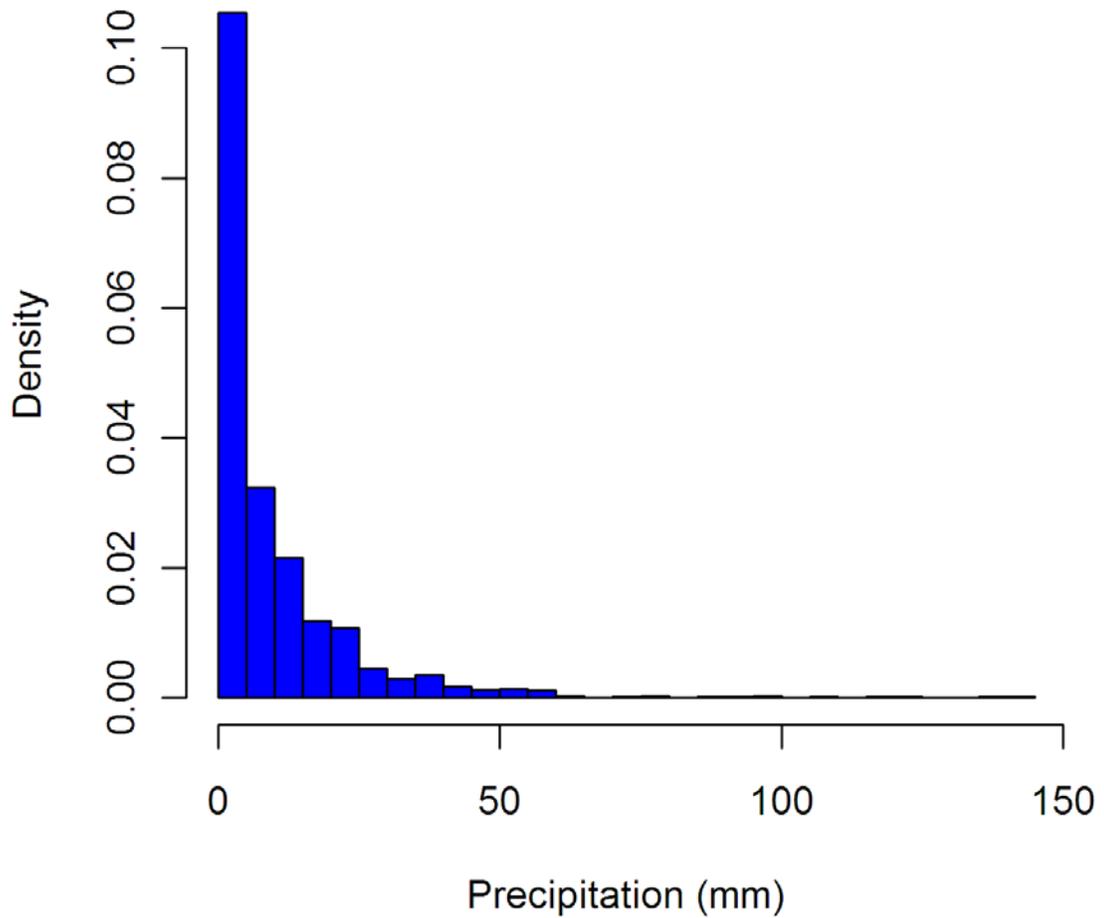


Figure 1.3. Histogram of Historical Daily Nonzero Precipitation at Raleigh-Durham Airport.

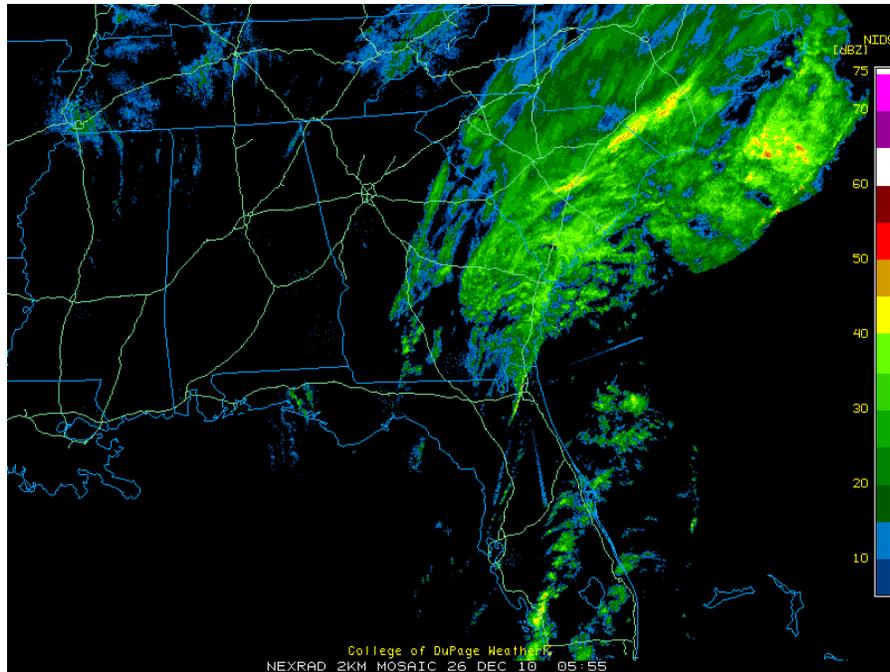


Figure 1.4. 0555Z 12/26/2010 NEXRAD Southeast U.S. Radar Reflectivity, showing frontal precipitation.

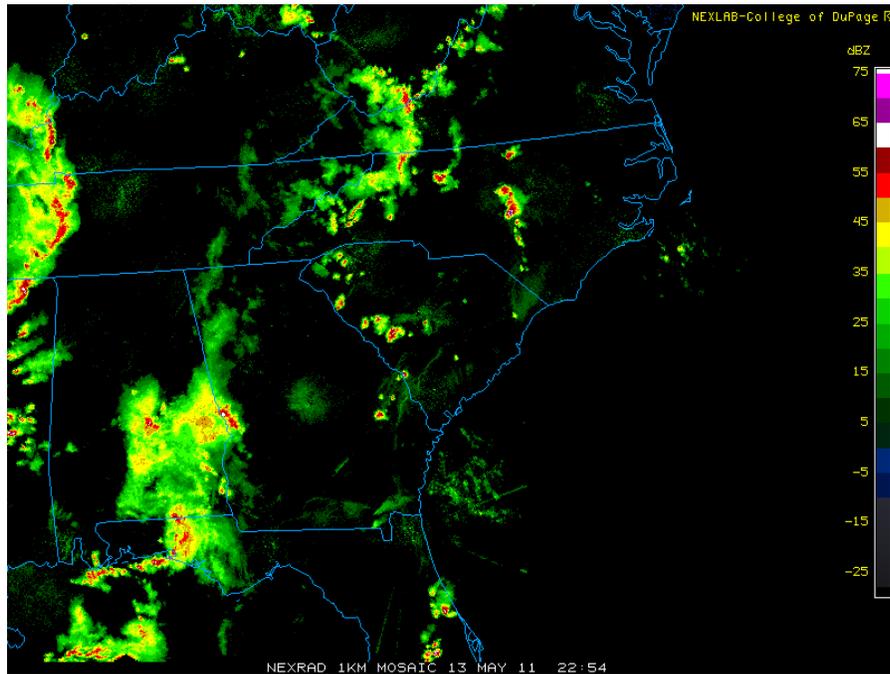


Figure 1.5. 2254Z 5/13/2011 NEXRAD Southeast U.S. Radar Reflectivity, showing unstructured convection.

Chapter 2. Methods and Experimental Design

The geospatial temporal weather generator, GiST, created by Baigorria and Jones (2010) is used to downscale the Climate Forecast System (CFS) model precipitation output and test the accuracy of changes to the various elements. In this chapter, the GiST weather generator process is reviewed briefly, followed by a discussion of downscaling with weather generators and what elements are varied, both in the GiST generator and in the downscaling process.

2.1. Overview of GiST

Full details of this technique are available in Baigorria and Jones (2010), but this section briefly reviews the differences of the GiST technique from traditional weather generators. The GiST weather generator as created by Baigorria and Jones (2010) is a Markov chain generator built similarly to the WGEN (Richardson and Wright, 1984). As mentioned in the previous chapter, most Markov Chain weather generators use first or second order chains to generate precipitation events for individual points independent of the relationship with other locations. In order to capture the spatio-temporal structure of precipitation events, GiST makes use of a two-state orthogonal Markov chain. Given any location and month an example of one of the transition probabilities in the Markov chain is:

$$P_{1|0,0,0} = \Pr(X_{i_t} = 1 \mid X_{j_t} = 0, X_{k_t} = 0, X_{i_{t-1}} = 0) \quad (2.1)$$

In equation 2.1, X reflects the state of the Markov chain, i.e. whether or not precipitation occurs (for rain $X=1$, no rain $X=0$). The location i refers to the location being generated and locations j and k refer to those locations most strongly correlated with location i , which have

already been generated. Therefore the transition probability in this example, $P_{1|0,0,0}$, is probability of precipitation occurring at station i and time t given that precipitation has not occurred at station j and time t , has not occurred at station k and time t , and did not occur at station i and time $t-1$. Given that there are three conditions used in each transition probability, there are eight transition probabilities used by GiST, while two state first order chains have two transition probabilities. However, the orthogonal Markov chain requires that two locations have already generated precipitation. GiST includes an initial process that generates precipitation for the first two locations simultaneously.

2.1.1. Initial Event Generation

The first two locations generated by the GiST process are those locations that are shown to have the highest correlation with all other locations through the use of the Euclidean N-correlational distance (G). This value is based upon the Pearson correlation of rainfall events and represents the strength of the association of one location with all other remaining locations. Given the value of G for each location, GiST determines the order that locations are generated by ranking the G values for each station. Smaller values of G indicate a location is more correlated with the remaining locations, while larger values indicate that the location is more independent from other locations. Therefore, the first two stations generated are those with the smallest values of G , and generated separately from the remaining locations. The remaining locations are generated from the smallest value of G to the largest value, which is the most independent location. In order to accurately reproduce the correlation structure of precipitation in each month, GiST determines the value of G at each location for each month using the observed correlations of rainfall events (Baigorria and Jones, 2010; Baigorria et al, 2007a).

In addition to determining the values of G for each month, GiST also determines the multi-annual average of the monthly number of rainy days (\bar{T}_m) and the multi-annual standard deviation of the monthly number of rainy days (σ_m) for each month (m). For every

location for each month, the expected number of rainy days (\hat{T}_m) is generated using the following equation:

$$\hat{T}_m = \bar{T}_m + \sigma_m r_{norm}^\psi . \quad (2.2)$$

In equation 2.2, r_{norm}^ψ is a set of correlated random numbers determined using a vector of normally distributed random numbers (\mathbf{R}_n) and a Cholesky factorization matrix (\mathbf{F}) calculated based on the correlation matrix of the monthly number of precipitation events. The vector of random numbers is matrix multiplied by the Cholesky matrix to produce a vector of correlated random numbers which are rescaled to the Gaussian distributed percentages of the population being included (-1 to 1 for 68% or -1.96 to 1.96 for 95% for instance) in the final calculation of \hat{T}_m . Although the values of \hat{T}_m are computed for each location, they are used only to provide the target number of rainy days for the first two locations generated.

Each day of the month for each of the starting locations is filled with rain and no rain days according to the value of \hat{T}_m . Both series are then reordered to best fit the relative frequency of spatial patterns between the two locations. Finally, to complete the event generation for the first two stations, both are reordered simultaneously to find the best fit for the 1st order Markov transition probabilities. From an ensemble of reordered datasets, the final event generation for both locations is the member which minimizes the root mean square error (RMSE) of the first order Markov transition probabilities for each location at the same time. Once the events are generated, the GiST generator proceeds to generate the remaining locations.

2.1.2. Subsequent Event Generation

The remaining locations for each month are generated in order from the smallest to the largest values of G , using the two-state orthogonal Markov chain described previously. Following a similar pattern to most Markov chain generators, a random number from the uniform distribution is compared to the appropriate orthogonal transition probability to determine rainfall at each of the remaining stations. Each of the locations generated are conditioned upon the two previously generated and most-correlated stations along with the previous state of the current location. When the first day of a new month is generated, the previous state refers to the last day of the previously generated month. However, for the first day of the generation, the unconditional probability of precipitation is used to determine the previous state for the location being generated. For each generated location another ensemble of simulations is created, and the member which minimizes the RMSE of 1st order transitional probabilities and the correlation of the location with the remaining stations is used in the final generation. However, in determining the RMSE it is important to note that only those locations which are not used to generate the current location are used to determine the RMSE for each ensemble.

2.1.3. Amount Generation

In addition to the correlation of precipitation events between locations, the correlation of nonzero precipitation amounts and the two parameter gamma distribution are used for generating precipitation amounts. GiST fits the shape (α) and scale (β) using the polynomial approximation proposed by Greenwood and Durand (1960). While the moment estimators for the distribution could be used to fit the parameters, several studies have shown that the moment estimators can give bad results for small values of α (Thom, 1958; Wilks, 1990; Wilks, 2006). To begin the amount generation process the correlation matrix of daily rainfall amounts transformed to the Gaussian distribution is determined. A vector of uniform random numbers distributed from 0 to 1 is matrix multiplied by the Cholesky factorization of the

correlation matrix. The resulting vector of correlated random numbers, which follows a Gaussian distribution, is then rescaled to the appropriate gamma distribution using the cumulative probability functions. The amount generation process is performed for each day and simplified so that the resulting vectors and matrices reflect the number of locations where rain events were generated. That is if K locations have generated rain events, then the vectors contain K elements and the Cholesky factorization matrix used is a $K \times K$ matrix. The results of this completed process are rainfall amounts and events that are spatially and temporally correlated. While the generation process in GiST has been summarized here, the full details of the generation process and algorithms involved are available in Baigorria and Jones (2010).

2.2. Potential Improvements to the GiST weather generator

As mentioned previously, both GiST and WGEN use the two-parameter gamma distribution to model nonzero precipitation amounts in their respective amount generation processes. While this distributions fit the majority of the distribution of precipitation, recent studies have shown that this distribution does not accurately represent precipitation in the tail of the gamma distribution, i.e. extremely heavy precipitation events (Furrer and Katz, 2008). Therefore, this study explores an alternative to model extreme precipitation events in the GiST weather generator.

2.2.1. Extreme Precipitation Generation

In order to generate extreme precipitation events, the GiST weather generator is altered to introduce an additional process to generate extreme events for each station. In this case extreme events are defined as those that fall into the top 10% of the observed precipitation distribution. The extreme event generation process functions similarly to the event generation process of GiST. In the extreme event generation process rain events are

those precipitation events where the observed precipitation is greater than the 90th percentile of the distribution.

In the amount generation process of the altered GiST generator, the gamma parameters for the top 10% of nonzero precipitation amounts are fitted in addition to the gamma parameters for the all nonzero precipitation amounts. While the process for generating extreme precipitation amounts functions similarly to the process for generating general precipitation amounts, extreme precipitation amounts are not generated if the precipitation itself never occurs, i.e. if no rain events are generated than no extreme events can occur.

2.2.2. Probable Results of the altered GiST technique

The result from this altered version of GiST is re-sampling the tail of the observed precipitation distribution, or causing more heavy precipitation events to occur than in the original GiST or WGEN generators. While this re-sampling approach increases the number of heavy precipitation events, since it is constrained to occur only when both the event generation simulates a rain event and the extreme event generation simulates an extreme event it is unlikely that heavy precipitation is over-sampled. Figure 2.1 shows an example of the distribution produced by the altered GiST approach in comparison to the original technique. From this point on this study, the altered version of GiST is referred to as GiSTR. As part of this paper, the accuracy of GiST, GiSTR and WGEN are compared to show the strengths and weakness of each generator.

2.3. Potential Improvements to Weather Generator Downscaling

2.3.1. The General Concept of Weather Generator Downscaling

As mentioned in the previous chapter, downscaling with weather generators makes use of two basic equations which condition the parameters of the generator based upon the information. These equations, known as scaling relationships (Chen et al, 2006), are shown again below:

$$F(P_{GCM})_{future} - F(P_{GCM})_{present} = F(P_{station})_{future} - F(P_{station})_{present} \quad (2.3)$$

$$\frac{F(P_{GCM})_{future}}{F(P_{GCM})_{present}} = \frac{F(P_{station})_{future}}{F(P_{station})_{present}} \quad (2.4)$$

For most Markov chain generators, the parameters related to precipitation events are perturbed using the equation 2.3 (the assumption of absolute change) and the parameters related to amounts are perturbed using the equation 2.4 (the assumption of relative change). For 1st order Markov chain generators these parameters include the mean of the daily nonzero precipitation (μ), the variance of the daily nonzero precipitation (σ^2), the unconditional probability of daily precipitation occurrence (π), and the lag 1 autocorrelation of the daily precipitation occurrence series (γ , also known as persistence). The parameters π and γ represent information regarding precipitation events and provide equivalent information to the 1st order transition probabilities through the following equations 2.5 and 2.6:

$$\pi = \frac{P_{01}}{1 + p_{01} - p_{11}} \quad (2.5)$$

$$\gamma = P_{11} - P_{01} \quad (2.6)$$

The remaining parameters, μ and σ^2 , are precipitation amount parameters and can be used to determine α and β through the moments of the gamma distribution, which are used by the generator for generating precipitation amounts. These four main parameters are determined for each month.

With regards to downscaling with weather generators, most studies consider the changes in climate through the monthly statistics from the GCM used (Wilks, 1999; Chen et al, 2006). However, while the monthly statistics are used (or what is available to use), weather generator parameters are daily statistics. Therefore in these cases, weather generator parameters are determined through the use of four monthly statistics;

- The average total precipitation ($E[S(T)]$)
- The variance of the total precipitation ($Var[S(T)]$), which also characterizes the inter-annual variability
- The average number of wet days ($E[N(T)]$)
- The variance of the number of wet days ($Var[N(T)]$)

These four monthly parameters can be translated into the daily counterparts for each month using the following set of equations:

$$E[S(T)] = E[N(T)]\mu \quad (2.7)$$

$$Var[S(T)] = E[N(T)]\sigma^2 + Var[N(T)]\mu^2 \quad (2.8)$$

$$E[N(T)] = T\pi \quad (2.9)$$

$$Var[N(T)] = \frac{T\pi(1-\pi)(1+\gamma)}{(1-\gamma)} \quad (2.10)$$

Using equations 2.7 through 2.10 most studies work to translate these available values for the future climate at each GCM grid point to the corresponding daily parameters of the weather generator. However, while the basic equations to determine these parameters are common to most studies, there are several different approaches presented in literature. The values of each of the generator parameters will be noted as π_{GCM} , γ_{GCM} , μ_{GCM} , and σ^2_{GCM} and represent the portion of the scaling relationship corresponding to the future climate portrayed at each GCM grid point.

While the present day values of each weather generator parameter at the GCM grid point may also be obtained in this way, most studies do not follow this approach. As best presented by Wilks (1999) and also argued by Benestad et al. (2008), all numerical climate models (both RCMs and GCMs) are a physical representation of the climate. Since both types of climate models contain multiple parameterizations and approximations they are different to some degree from the climate portrayed by observations. Therefore, in most cases the GCM data is not used to represent the present climate on the large scale, but an area average or a weighted area average of the parameters surrounding the GCM grid point is taken to represent the present values of each parameter at each GCM grid point in the present time. For reference in this discussion the parameters will be referred to as, $\bar{\pi}, \bar{\gamma}, \bar{\mu}, \bar{\sigma}^2$.

Given the determination of the GCM parameters and the values of these parameters in the present time for each location, the perturbed values of each parameter ($\hat{\pi}, \hat{\gamma}, \hat{\mu}, \hat{\sigma}^2$) can be determined through the use of the scaling relationship at each location. However, while the amount parameters are often perturbed directly, perturbing the event parameters directly can result in unrealistic values of the parameters. Therefore, the event parameters are transformed using the log odds and Fischer Z transformations before being used in the scaling relationship (Wilks, 1999; Chen et al, 2006). The log odds transformation is applied to all the values of π while the Fischer Z transformation is applied to all values of γ .

$$L(\pi) = \ln \left[\frac{\pi}{1 - \pi} \right] \quad (2.11)$$

$$Z(\gamma) = \frac{1}{2} \ln \left[\frac{1+\gamma}{1-\gamma} \right] \quad (2.12)$$

These transformed parameters are used in the scaling relationship which assumes an absolute change, and solving the relationship for the transformed value of $\hat{\pi}$ and $\hat{\gamma}$ gives the following results for each location:

$$L(\hat{\pi}) = L(\pi) + L(\pi_{GCM}) - L(\bar{\pi}) \quad (2.13)$$

$$Z(\hat{\gamma}) = Z(\gamma) + Z(\gamma_{GCM}) - Z(\bar{\gamma}) \quad (2.14)$$

The transformed values produced by equations 2.13 and 2.14 are then back transformed to produce the final perturbed values, $\hat{\pi}$ and $\hat{\gamma}$, used by the generator to simulate the future climate at the local scale and a daily timescale.

$$\hat{\pi} = \frac{\exp[L(\pi) + L(\pi_{GCM}) - L(\bar{\pi})]}{1 + \exp[L(\pi) + L(\pi_{GCM}) - L(\bar{\pi})]} \quad (2.15)$$

$$\hat{\gamma} = \frac{\exp\{2[Z(\gamma) + Z(\gamma_{GCM}) - Z(\bar{\gamma})]\} - 1}{\exp\{2[Z(\gamma) + Z(\gamma_{GCM}) - Z(\bar{\gamma})]\} + 1} \quad (2.16)$$

While the event parameters are perturbed based on the assumption of absolute change, the amount parameters are perturbed based on the assumption of relative change and are not transformed prior to being used in the scaling relationship. The result for each location for $\hat{\mu}$ and $\hat{\sigma}^2$ is as follows:

$$\hat{\mu} = \frac{\mu^* \mu_{GCM}}{\bar{\mu}} \quad (2.17)$$

$$\hat{\sigma}^2 = \frac{\sigma^2 * \sigma_{GCM}^2}{\bar{\sigma}^2} \quad (2.18)$$

These amount parameters are then used to determine the values of $\hat{\alpha}$ and $\hat{\beta}$ for each location, through the moment estimators of the gamma distribution.

$$\hat{\alpha} = \frac{\hat{\mu}^2}{\hat{\sigma}^2} \quad (2.19)$$

$$\hat{\beta} = \frac{\hat{\sigma}^2}{\hat{\mu}} \quad (2.20)$$

The aforementioned perturbation process describes the general concept behind downscaling with weather generators. As such there are numerous variations on this general concept, including the distinction of the different first order transitions in the amount parameters (Chen et al, 2006) and incorporating spatial regression or other types of ESD into the perturbation procedure (Semenov and Barrow, 1997). However, for the purposes of this study, I will focus primarily on the general concept and the potential improvements that variations to this concept may offer for future downscaling efforts.

2.3.2. Changes and Improvements to the General Approach

The prior section described the general approach that is taken to downscaling GCM information using weather generators. However, while this general approach has existed for

several years there are few studies (e.g. Hashmi et al, 2009) which use the daily or sub-daily output produced by several GCMs, such as the Climate Forecast System (CFS) model. This output data can be used to directly solve for the four daily parameters required by the weather generator for each grid point in the future climate represented by the GCM. In addition, a recent study by Wilks (2008) suggests that the parameters of a weather generator might be more accurately interpolated by a locally weighted regression (LWR) than by global regressions, Thiessen polygons (nearest neighbor) or by domain or area averaging. While Wilks shows promising results in the context of interpolating parameters to unknown locations for running simulations in present climate, this approach has not been attempted with regards to downscaling applications. Specifically, the LWR developed by Wilks could be applied to interpolate the present climate values to each grid point in the GCM, replacing the area averaging currently used in most approaches. The likely benefit from this change would be a more accurate representation of the variability of precipitation with respect to topography, since the LWR approach developed by Wilks accounts for the elevation of each predictor location, where area averaging does not. Finally, the use of daily output instead of monthly information allows for the possibility of using approaches other than perturbing the amount parameters and using the moment estimators to determine the values of $\hat{\alpha}$ and $\hat{\beta}$ for each location. As mentioned in section 2.1.3, the moment estimators can produce bad results for small values of α , as is often the case with precipitation. However, the maximum likelihood approximation proposed by Greenwood and Durand (1960; Wilks, 2006), may allow for more accurate representation of the parameters of the gamma distribution.

For this alternate approach first consider the approximation proposed by Greenwood and Durand through the use of the D statistic shown in equation 2.21:

$$D = \ln(\bar{x}) - \frac{1}{n} \sum_{i=1}^n \ln(x_i) \quad (2.21)$$

where \bar{x} represents the mean daily nonzero precipitation as determined from observations at each location or from daily output precipitation for each GCM grid point in the future time, n is the total number of nonzero precipitation events and x is the daily nonzero precipitation for each event i . In the representation proposed by Greenwood and Durand, the values of the shape and scale are:

$$\hat{\alpha} = \frac{0.5000876 + 0.1648852D - 0.0544274D^2}{D}, \quad 0 \leq D \leq 0.5772 \quad (2.22)$$

$$\hat{\alpha} = \frac{8.898919 + 9.059950D + 0.9775373D^2}{17.79728D + 11.968477D^2 + D^3}, \quad 0.5772 \leq D \leq 17.0 \quad (2.23)$$

$$\hat{\beta} = \frac{\bar{x}}{\hat{\alpha}} \quad (2.24)$$

Where $\hat{\alpha}$ and $\hat{\beta}$ are the approximate shape and scale parameters to represent the distribution of x . The values of D can be determined for every location of interest, for each GCM grid point in the future climate, and interpolated to each GCM grid point in the present climate. In addition, the mean daily nonzero precipitation is also already determined and interpolated as described previously. Therefore it becomes possible to determine the values of α and β at every location of interest, for each GCM grid point in the future, and interpolate to each GCM grid point for the present climate. In short, through the use of daily or sub-daily GCM output and the Greenwood and Durand approximation, it becomes possible to directly perturb the parameters of the gamma distribution to reflect the future climate, potentially with more accuracy than with the use of the moment estimators following the perturbation of μ and σ^2 . Therefore, following the same scaling relationship used for μ and σ^2 , the perturbation of α and β becomes the following:

$$\hat{\alpha} = \frac{\alpha^* \alpha_{GCM}}{\bar{\alpha}} \quad (2.25)$$

$$\hat{\beta} = \frac{\beta^* \beta_{GCM}}{\bar{\beta}} \quad (2.26)$$

Where $\hat{\alpha}$ and $\hat{\beta}$ are now the estimated values of the shape and scale of the approximate gamma distribution for each location of interest in the future time. This new approach, with the direct perturbation of the parameters of the gamma distribution, is tested along with the LWR approach in an experimental setup to evaluate the error associated with each combination in downscaling with weather generators and the potential uses for local seasonal forecasting

2.4. Experimental Design

Given that this analysis evaluates improvements to both weather generators and the downscaling approach with weather generators, this section describes experimental design for both parts. Chapters 3, 4 and 5 present and discuss the results from each experiment.

2.4.1. Design for Evaluating Weather Generator Improvements

In order to evaluate the improvements of various weather generators, GiST, GiSTR and WGEN are compared over the state of North Carolina and over the broader Southeast United States. For each domain considered in this analysis, station observations from 1971 through 2000 will be used to determine the required information to run each weather generator. The period of 1971-2000 is used in this analysis since it is a standard period used to define climatologically normal precipitation at the COOP stations. All the stations used are NWS Cooperative Observer network (COOP) stations. For the North Carolina domain,

91 stations are used which have at least 25 complete years of data. The North Carolina domain stations are shown in Figure 2.2. For the domain covering the Southeast United States, 319 stations are used which have at least 29 complete years of data and at least 200 days of the remaining incomplete year. The Southeast U.S. domain stations are shown in Figure 2.3. The size of this domain is chosen to cover all states in the Southeast U.S. and extends northward to cover portions of the mid-Atlantic. While the primary focus will be on the Southeast U.S. extending the domain into the mid-Atlantic will allow for an assessment of the potential value of each generator and downscaling variation for users in this area as well as in the Southeast U.S. In both domains, these stations are chosen in order to provide robust values of all the information required for each generator to function. The observed parameters required for each location are determined and input into appropriate weather generator. The output from each weather generator is used to determine the estimates for several observed parameters. Each weather generator estimated parameter is evaluated against their observed counterparts at each station using the root mean square error (RMSE). In this way, each weather generator can be evaluated for how accurately it reflects the observed climate through how well the generator replicates the observed parameters used to initialize the generator. For each generator, in each domain, the parameters evaluated are described in Table 2.1.

Following the methodology used by Baigorria and Jones (2010), the observed parameters are determined and evaluated on the monthly timescale for each station. A numerical evaluation includes computation of the root mean square error (RMSE) and the error (generated - observed) of the parameters generated by each parameter compared to the observed values of each parameter. In addition, two-sample T-tests are performed to determine if specific generators have significantly larger error than others. In order to insure that the assumptions of normality and independence are met, this testing is performed using the RMSE of each generator for each parameter. In addition to a numerical evaluation of each weather generator, each generator will also be evaluated through a visual analysis of multiple parameters. This includes a visual comparison of the output of each generator in warm and cold months, the uncertainty of each generator in space, the generated compared to

observed probability distribution functions of precipitation amounts, and the frequency of dry and wet spells of various lengths. These experiments focus on answering the questions associated with evaluating the accuracy of each weather generator, including the strengths and weaknesses of each weather generator and the relationship between the error of each generator and any physical phenomena related to climate, extreme events or topography.

2.4.2. Design for Evaluating Downscaling Variations

A similar procedure to the previous section is used to evaluate four variations in the weather generator downscaling approach. The first set of variations involves the interpolation of weather generator parameters to each GCM grid point by either area weighted averaging (AW) or locally weighted regression (LWR). The second set of variations involves the use of the moment estimators (MOM) versus the Greenwood and Durand approximation (DFIT) of the gamma distribution parameters. The resulting combinations of each set of variations are

- LWR-MOM – LWR for parameter interpolation and the MOM for the gamma distribution parameters
- AW-DFIT – AW for parameter interpolation and the DFIT for the gamma distribution parameters.
- LWR-DFIT – LWR for parameter interpolation and the DFIT for the gamma distribution parameters.
- AW-MOM – AW for parameter interpolation and the MOM for the gamma distribution parameters.

AW-MOM reflects the general approach taken for downscaling with weather generators, while the remaining three runs reflect the combinations that will possibly improve upon the general AW-MOM approach. For the LWR interpolation this project uses

the method proposed by Wilks (2008), while the AW interpolation follows the simple area average also described by Wilks (2008). In order to limit the number of variables involved in the experimental design, only one generator is used for comparing the different combinations described previously. While the error resulting from downscaling with each combination may change for different generators, the difference in error between combinations should be approximately the same for different generators. Therefore, while these combinations should be tested for all generators, the focus of this project will be on the difference in error between all combinations with a constant weather generator.

For this analysis, the Climate Forecast System Reanalysis (CFSR) output is the GCM used (Saha et al, 2010). CFSR precipitation information is available on a T62 Gaussian grid, or an approximate resolution of two degrees latitude and longitude. Hourly output of precipitation rate is aggregated to produce a time series of daily total precipitation for each GCM grid point. This daily precipitation data is used to provide the future information for each weather generator parameter at each grid point. The observed data for the present time will be provided by the same 319 stations used in the weather generator evaluation section across the Southeast United States (Figure 2.4). CFSR output is used for this analysis for several reasons. First, the CFS model is used by the Climate Prediction Center to provide operational seasonal forecasts for precipitation and temperature in the Southeast U.S, and may be used for future operational forecast downscaling. This CFS Reanalysis includes coupled atmosphere, ocean, sea-ice, and land surface models which are unavailable operationally in other GCMs. Finally, while the CFS Reforecast is also available for much of the time period, the CFSR data are used to provide consistent initial conditions for the Reforecast. Therefore, the CFSR data are used to eliminate any additional errors in the analysis that may be associated with the modeling processes in the CFS Reforecast data.

In this analysis, two time periods will be used for the downscaling time frame; 1979 to 2000 and 2001 to 2009. Two time periods will be considered in this analysis in order to determine if the error differences between combinations are consistent in time, given that the GCM and weather generator used are held constant. The base time period for this analysis is 1979-2000 time period, while the forecast time period is 2001-2009. The base time period in

this analysis is smaller than the 1971-2000 time period used for the weather generator evaluation since the CFSR data is only available for 1979 to the present. For the evaluation, the same parameters, numerical evaluation, significance testing, and visual analysis will be used as was described in the previous section. This experiment will focus on determining the strengths and weaknesses of each downscaling simulation along with the relationship between the error of each simulation and physical phenomena.

At this point there are a few important points to consider. First, no studies have made use of GiST or GiSTR for downscaling. Second, while GiST and GiSTR both make use of the spatial structure of precipitation in their respective generation processes, the correlation matrices and orthogonal probabilities themselves will not be perturbed in any way by CFSR. While GiST and GiSTR affords the possibility of altering the spatial structure of events and amounts to reflect the changes that may occur in time, the coarse resolution of CFSR and most GCMs means that the correlations between individual stations cannot be resolved. In contrast, most studies have considered perturbed the average correlation between stations and the influence on other parameters, rather than the individual correlations (Wilks, 1999b; Osborn, 1997). For this project, since the two time periods used are within the past 30 years, the spatial correlations used by GiST and GiSTR will not be altered in any way. That is, I will assume that the climate has not changed enough over the period to impact the local spatial structure of precipitation in the study domain. While this assumption is likely valid for recent time periods and for seasonal forecasting, this would likely not be valid for longer-term climate change studies. Therefore, further studies could be done to determine how to perturb the spatial correlations for individual stations in a changed climate, which might be particularly valuable with regards to changes in precipitation with the possible expansion of the Hadley cell. Regardless, for this study I will assume that the change in climate has not had a significant impact on the spatial structure of precipitation in the domain.

Finally, in addition to the comparison between downscaling combinations, a single downscale run of WGEN in each time period will be compared to the downscale run of GiSTR. For consistency, this comparison will only be for one combination in each time period. This comparison will indicate whether the results of the weather generator evaluation

are consistent when downscaling is applied and how much if any further comparison may be required to solidify the analysis with regards to the four combinations of downscaling approaches. In addition, it is important to note that there is a caveat to these particular studies. As part of the amount generation process of GiST for current time (such as for inputs for crop modeling) and near time simulations (such as for seasonal forecasting or predictions for the next year) precipitation is restricted to be less than 110% of the average maximum precipitation for each month for each location (Baigorria, 2010, personal communication). This is meant to prevent unrealistic values of precipitation from occurring, and while it can be altered for climate change downscaling, it was retained in both GiST and GiSTR since the focus of this study was primarily on seasonal precipitation estimates for current and near time.

Table 2.1. Description of parameters used to evaluate the each weather generator.

Parameter	Description
μ	Mean Daily Nonzero Precipitation
σ^2	Variance of the Daily Nonzero Precipitation
$E[S(T)]$	Average total precipitation
$Var[S(T)]$	Inter-annual variability
P_{01}, P_{11}	1 st Order Markov transition probabilities
π	Unconditional probability of rain
γ	Persistence (lag 1 autocorrelation)
ρ	Correlation matrix of precipitation
ρ_{ev}	Correlation matrix of precipitation events
ρ_{am}	Correlation matrix of precipitation amounts
ρ_{ex}	Correlation matrix of precipitation extreme events

Example of Theoretical PDFs

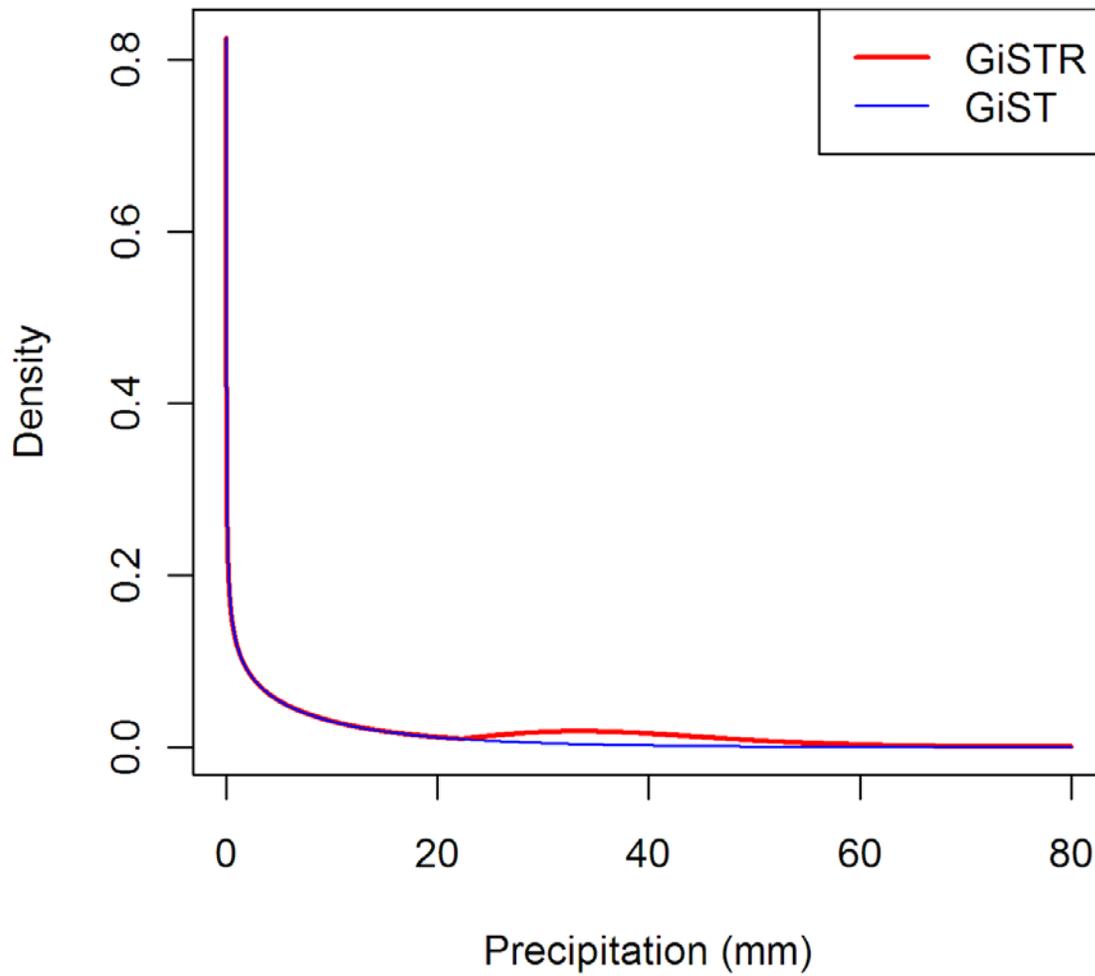


Figure 2.1. Example of Theoretical PDFs produced by GiST and GiSTR

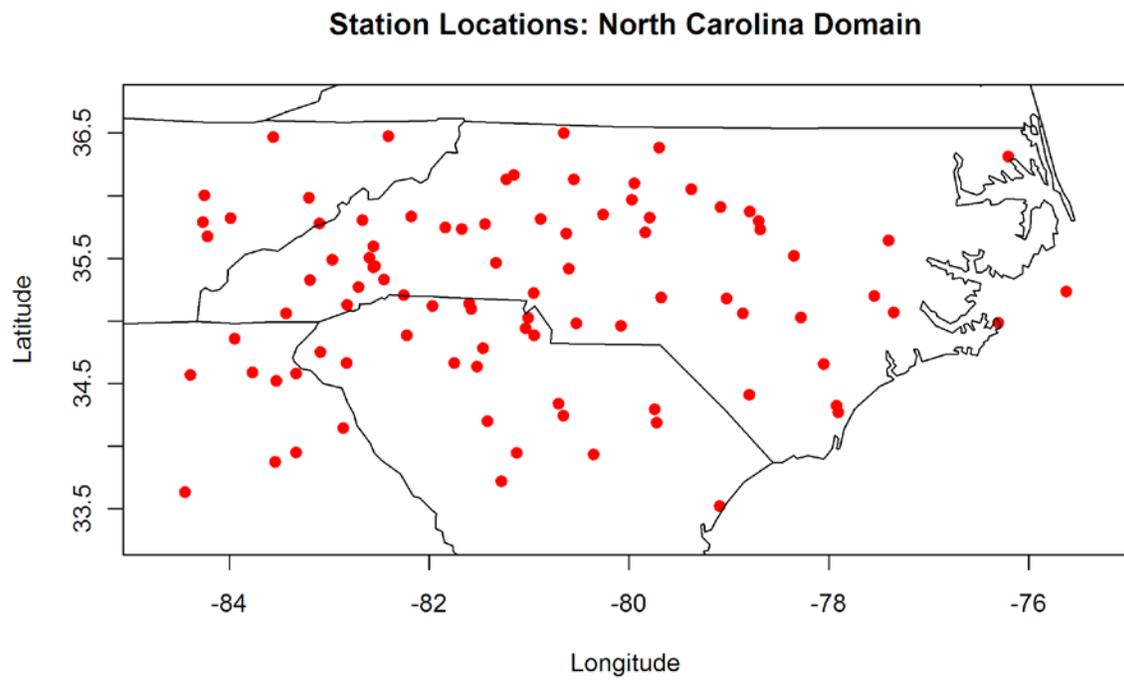


Figure 2.2. Locations of the 91 COOP stations in the North Carolina domain used in the weather generator evaluation

Station Locations: Southeast Domain

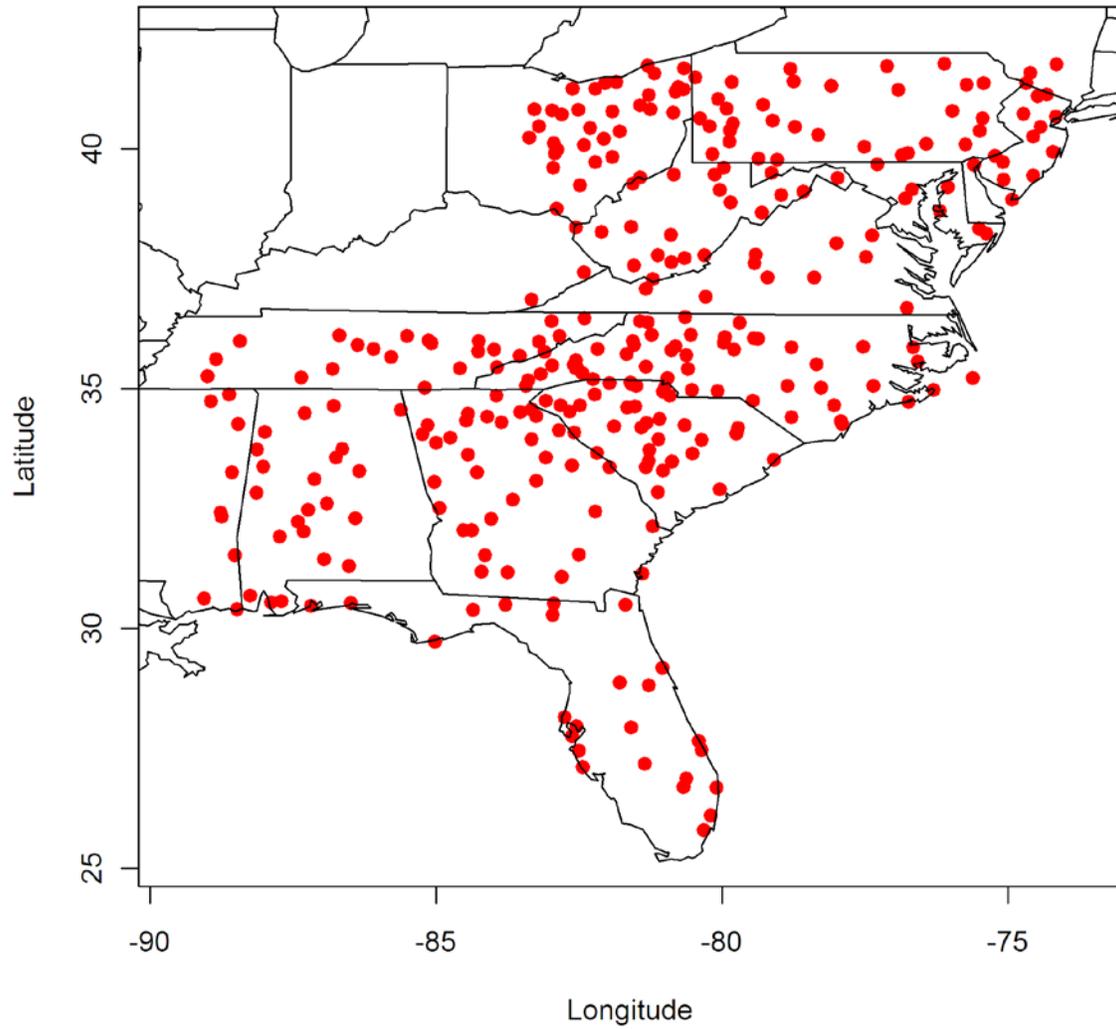


Figure 2.3. Locations of the 319 COOP stations in the Southeast United States domain used in the weather generator evaluation.

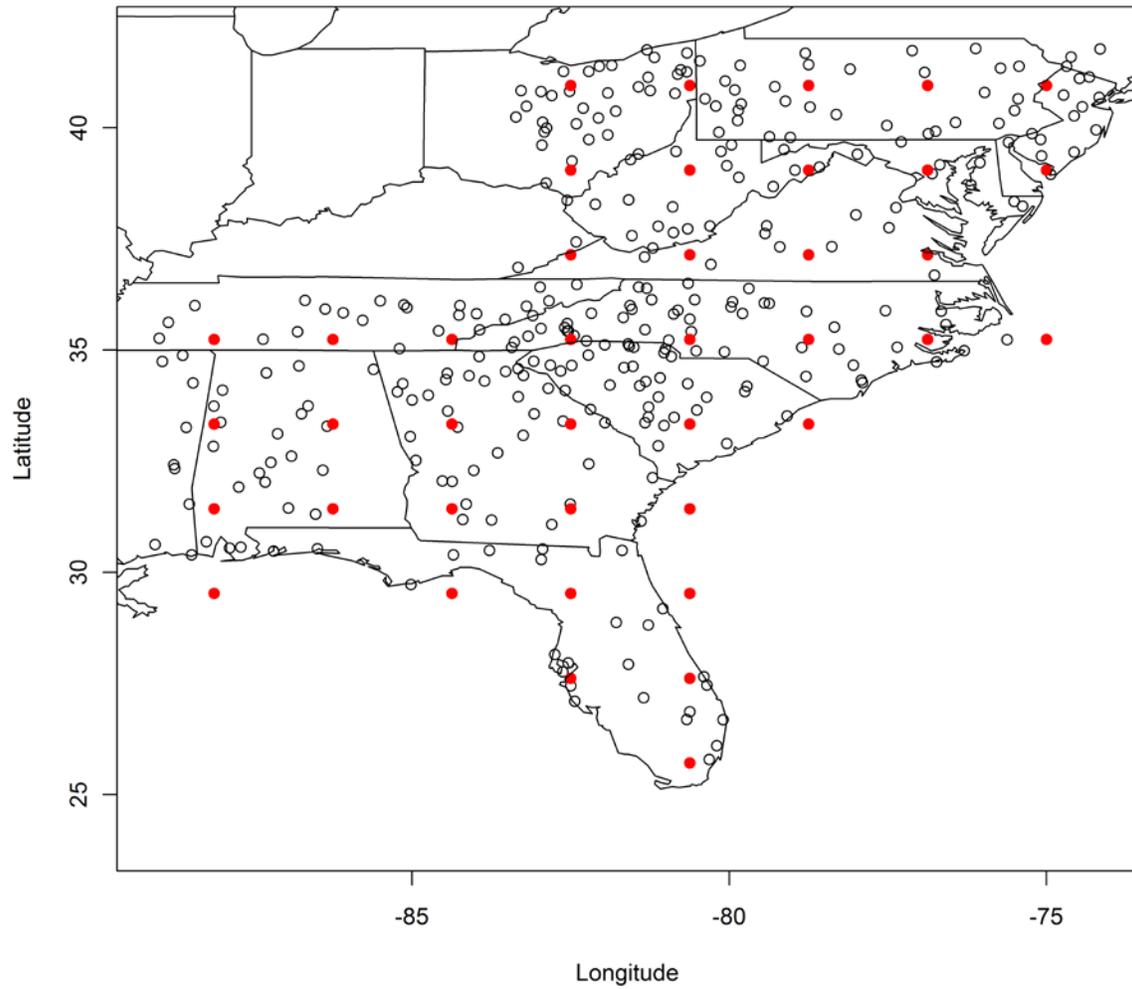


Figure 2.4. NWS COOP stations (black) circles and CFSR grid points (red dots) used in the downscaling evaluation.

Chapter 3. Weather Generator Evaluation

In this chapter the results of the evaluation of WGEN, GiST, and GiSTR is presented. Each generator was run over two separate domains producing an ensemble of thirty members using the observations from stations for each domain. The first domain includes 91 COOP stations which cover the state of North Carolina (Figure 2.2), while the second domain includes 319 COOP stations covering the entire Southeast United States (Figure 2.3). The main focus of this chapter is to evaluate the strengths and weaknesses of each weather generator with regards to replicating the observed climate in each domain. These two domains were chosen to analyze the change in error with increasing domain size in order to determine changes with regards to domain size for each generator. For this analysis, the observed climate is measured as observed parameters for each NWS Cooperative observer station. In addition, several parameters which are not used in the basic generator process are analyzed to help determine the error of each generator. A description of these parameters is provided in Table 2.1. In addition to comparing RMSE of each of these parameters, the significance of the difference in RMSE is tested using a two sample t-test assuming the variance of RMSE is not equal. The level of significance for each test is fixed at 0.05. Given that the process of each generator is strongly different, it is not appropriate to assume that the variance of RMSE for each generator for each parameter will be the same. Along with the numeric analysis there is a visual analysis to determine the accuracy of each weather generator with respect to the spatial patterns of error and the probability distribution functions produced.

3.1. North Carolina Domain

In this domain each generator has similar results with regards to the event parameters, but there are interesting results with regards to the amount parameters and spatial structure.

3.1.1. Temporal Parameters

The temporal parameters reflect the occurrence of precipitation events at each station in the domain. Therefore, since these parameters influence the number of days with precipitation and the lengths of dry and wet spells it is important that any weather generator accurately captures these parameters. Recall from Chapter 2 that these parameters include the 1st order Markov transition probabilities (P_{01} and P_{11}), the unconditional probability of precipitation (π), and persistence (γ). A comparison of the root mean square error (RMSE) of the 1st order Markov transition probabilities for each month, shown in Table 3.1, indicates that WGEN has a lower RMSE by 0.03 to 0.042 on average for every month. In addition, there is also no trend in time for the error of each generator for the 1st order probabilities. While there is no trend in time, testing the error of each generator with a two-sample T-test against every other generator shows a distinct result. The results of these two-sample T-tests for the 1st order transition probabilities are summarized in Table 3.2. These two sample T-test results indicate that the RMSE of the 1st order transition probabilities for WGEN is significantly smaller than for GiST and GiSTR. However, the test also shows that there is no statistically significant difference in RMSE between GiST and GiSTR for these probabilities. For this chapter and subsequent chapters, please note that significance refers to the statistical significance of various tests, unless otherwise noted. Each of these comparisons and tests indicate that the values of each parameter simulated by each generator at each station shows that WGEN is consistently more accurate than GiST and GiSTR for replicating the occurrence of precipitation events.

Given the relationship between P_{01} and P_{11} and π and γ (Equations 2.5 and 2.6) it is expected that the results for π and γ are similar to results for P_{01} and P_{11} . Table 3.3 shows a comparison of the error of π and γ between each generator for all months and on average. WGEN has an RMSE for each parameter less than GiST and GiSTR by 0.013 to 0.08 on average across all months. In addition, there is no apparent seasonal trend in the error of each generator for these parameters. Moreover, the two-sample t-tests for RMSE of each of these parameters also show similar results to the testing for the Markov probabilities. A

summary of the results of the two-sample T-tests for π and γ is shown in Table 3.4. The testing shows that for both π and γ , WGEN has a significantly smaller RMSE than both GiST and GiSTR, while there is no significant difference in RMSE of each parameter for between GiST and GiSTR.

RMSE for all of these parameters is significantly less for WGEN compared to GiST and GiSTR, which makes sense given that the generation processes in WGEN focuses specifically on temporal structure while both GiST and GiSTR emphasize both the spatial and temporal structure of precipitation. Since these four parameters ($P_{01}, P_{11}, \pi, \gamma$) represent the temporal structure of precipitation events, it is likely that increased RMSE of these parameters for GiST and GiSTR is partially the result of dual focus represented in the event generation processes of GiST and GiSTR. In addition, while the T-tests indicate that WGEN has significantly smaller RMSE for each parameter, the difference in the average RMSE is less than 0.08. This suggests that the error for replicating the occurrence of precipitation is practically insignificant between generators, or that each generator will produce roughly the same number of rainy days.

3.1.2. Spatial Structure Parameters

While the analysis for the temporal parameters shows that WGEN is the most accurate for replicating these parameters, the analysis of the spatial structure shows that WGEN has the largest error for replicating the spatial structure in the domain. Figure 3.1 shows the trend in RMSE of the correlation matrix of precipitation (a, ρ), the correlation matrix of precipitation events (b, ρ_{ev}), the correlation matrix of precipitation amounts (c, ρ_{am}), and the correlation matrix of precipitation extreme events (d, ρ_{ex}) for each generator across all months. Regardless of the spatial structure parameter considered, the error of WGEN is larger than both GiST and GiSTR by up to 0.5 in winter months, and up to 0.25 in summer months. In addition, there is almost no difference in the error of the spatial structure between GiST and GiSTR regardless of the parameter considered. The two-sample T-test was also applied to RMSE of the spatial structure parameters, and these results are summarized in

Table 3.5. The results of this testing indicates that the error of WGEN for these parameters is significantly larger than GiST and GiSTR. However, the testing also shows RMSE of each correlation matrix is not significantly different between GiST and GiSTR. While the RMSE of each generator for each correlation matrix indicates the error associated with the spatial structure, it does not indicate the ability of each generator to replicate the decay in correlation with distance. In order to evaluate the ability of each generator to capture the decay in correlation with distance, the average decay function correlogram for January for precipitation (Figure 3.2a), precipitation events (Figure 3.2b), precipitation amounts (Figure 3.2c), precipitation extreme events (Figure 3.2d) for the North Carolina domain. The average decay function correlogram represents the average correlation between stations in 5km distance intervals both observed and generated. For instance, the observed value of the average decay function correlogram at 15km is the observed average correlation between stations which are 10 to 15km apart. Figure 3.2 shows that regardless of the distinction between events, amounts, or extreme events, WGEN fails to reproduce the observed decay in the correlation between locations with distance. Figure 3.2 also shows that GiST and GiSTR both more accurately reproduce the observed average decay function correlogram regardless of the distinction between events, amount, or extreme events. However, GiST and GiSTR most accurately reproduce the average decay function correlogram for precipitation events (Figure 3.2b) compared to precipitation, precipitation amounts, and precipitation extreme events. These patterns are consistent with the RMSE of each correlation matrix, but they are also consistent for other months. Figure 3.3 shows the observed average decay function correlogram compared to those generated by each generator for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d) for July. For this month, WGEN does not reproduce the observed average decay function correlogram, while GiST and GiSTR more accurately reproduce the observed correlogram. However, unlike the results for the January, the correlograms produced for July by GiST and GiSTR accurately replicates the observed correlogram for all different parameters. This is consistent with the decrease in RMSE from January to July for precipitation, precipitation amounts, and precipitation extreme events. In addition, because the observed average decay

function correlogram decreases rapidly in July for all four parameters there are distances (particularly for precipitation amounts, Figure 3.3c, and precipitation extreme events, Figure 3.3d) where WGEN replicates the observed correlogram. The rapid decay in the observed correlogram in July is indicative of the isolated convection which is a dominant structure of precipitation in summer. Therefore the rapid decay in the observed correlogram associated with the convective precipitation in summer allows WGEN to be more accurate to reproduce spatial structure than in other months. However, given that the observed decay function correlogram is not zero and that GiST and GiSTR accurately replicate the observed correlogram, it indicates that while isolated convection is a dominant form of precipitation in summer, it is not the only structure which influences precipitation in summer. As such, GiST and GiSTR both more accurately capture the spatial structure of precipitation than WGEN in all months. This result is also consistent with the results shown by the RMSE comparison for each correlation matrix.

The results of this first group of analyses of the temporal parameters and the spatial structure indicates that although GiST and GiSTR have a higher error when replicating the Markov probabilities, π and γ , both generators replicate the spatial structure more accurately than WGEN. This makes sense given that GiST and GiSTR consider both spatial and temporal structure while WGEN only considers the temporal structure of precipitation. Moreover, the analysis confirms that GiST and GiSTR produce nearly identical results. However, while the testing shows that the error of the temporal parameters for GiST and GiSTR is significantly larger than WGEN, the difference in error is less than 0.08 in all cases. This suggests that while there is significant difference, there may be little practical difference when applied in the context of seasonal forecasting.

3.1.3. Precipitation Amount Parameters

While there are clear differences between weather generators for the temporal parameters and the spatial structure, there are several interesting differences between each of the generators for the precipitation amount parameters.

3.1.3.1. Mean Daily Nonzero Precipitation (μ)

The mean daily nonzero precipitation (μ) simulated by each generator indicates that GiST has the lowest error for replicating this parameter. On average, the RMSE for WGEN, GiST and GiSTR is 1.04, 0.45 and 4.83 mm respectively. A comparison of the RMSE of this parameter between generators for each month is shown in Figure 3.4. The RMSE of this parameter is largest for GiSTR in every month, and the RMSE for each generator is also largest for August through October and smallest in the cold season. The large difference in RMSE of μ for GiSTR compared to GiST and WGEN is likely the cause of the re-sampling of extreme events built into GiSTR. This re-sampling approach likely produces too many heavy precipitation events, resulting in an over-estimation of μ compared to observations. For example, consider the month of August, where RMSE of WGEN, GiST and GiSTR is 1.32, 0.37 and 9.08 mm respectively. Figure 3.5 shows the difference between observed and generated values of μ for WGEN (a), GiST (b), and GiSTR (c) for August. Both WGEN and GiST do not show any general spatial patterning overall. However, GiSTR overestimates the values of μ for most stations in the domain. For several stations in coastal areas, GiSTR overestimates by more than 7 mm which forces RMSE for this month to be larger than the annual average RMSE for GiSTR. In addition, T-test results, summarized in Table 3.6, indicate there are significant differences in RMSE of μ between all three weather generators. Both WGEN and GiST have significantly smaller RMSE than GiSTR, while RMSE of GiST is significantly smaller than WGEN. This indicates that GiST has the lowest error for mean daily nonzero precipitation in this domain.

3.1.3.2. Variance of Daily Nonzero Precipitation (σ^2)

In addition to the differences in RMSE of μ between generators, there are also differences in RMSE of the variance of daily nonzero precipitation (σ^2) between generators. Comparing the average RMSE of σ^2 for each generator shows that GiSTR has the highest accuracy for replicating this parameter. The average RMSE for WGEN, GiST and GiSTR

for σ^2 is 80.68, 122.49 and 65.15 mm² respectively. Figure 3.6 shows a comparison of the RMSE of σ^2 between generators across every month. In every month, GiSTR has the smallest RMSE for this parameter while GiST has the largest error. Regardless of the generator used the error for this parameter also peaks in September and October. The two-sample T-test results for the RMSE between generators is summarized in Table 3.7. The RMSE of σ^2 decreases from GiST to WGEN (by 41.81 mm²) to GiSTR (by another 14.83 mm²), but only the difference in RMSE between GiST and GiSTR is significant. As already mentioned, the RMSE of σ^2 peaks during September and October. These months are the peak months for precipitation resulting from tropical cyclones in the state of North Carolina. While tropical cyclones generally produce intense precipitation over a period of days, these events are relatively rare occurrences. As such, the occurrence of tropical precipitation results in larger observed values of σ^2 . Therefore, the result of the high RMSE in all three generators during these peak months is because σ^2 is underestimated by each generator for coastal stations. For example, in September, RMSE of σ^2 for WGEN, GiST and GiSTR is 197.65, 275.66, and 142.96 mm² respectively. The difference between the observed and generated values of σ^2 for all stations WGEN (a), GiST (b) and GiSTR (c) for September for all stations is shown in Figure 3.7. The primary source of error for each weather generator for this parameter is the underestimation of observed values of σ^2 for coastal stations in the domain. In addition, the magnitude of the error in this area is smallest for GiSTR. The results in this section suggest that while GiSTR overestimates values of μ , it more accurately replicating the values of σ^2 or the variability throughout the year, but also better captures the variability associated with tropical precipitation.

3.1.3.3. Average Total Precipitation ($E[S(T)]$)

While the mean daily nonzero precipitation (μ) and the variance of the daily nonzero precipitation (σ^2) are directly related to the average total precipitation ($E[S(T)]$) and the inter-annual variability ($Var[S(T)]$), RMSE patterns of μ for each generator are reflected in RMSE of $E[S(T)]$, while RMSE patterns of σ^2 are not. For replicating values of $E[S(T)]$, GiST

shows the lowest error while GiSTR has the largest error compared to the other weather generators. The average RMSE for WGEN, GiST and GiSTR is 12.24, 8.44, and 63.51 mm respectively. Figure 3.8 shows a comparison of the RMSE of $E[S(T)]$ between generators for each month. This comparison indicates that GiST has the smallest error for all months, with the peak error for each generator occurring in August through October, similarly to the RMSE of μ . Given that GiSTR overestimates μ because of the re-sampling of heavy precipitation and that $E[S(T)]$ is the average total precipitation per month, it is likely that GiSTR is also overestimating the values of $E[S(T)]$ for most stations. The consistent over-sampling of heavy precipitation by GiSTR on the daily timescale of each month results in error that aggregates to the monthly timescale. The result is a large overestimation of the values of $E[S(T)]$ by GiSTR for all months. Figure 3.9 shows boxplots of the difference between the generated and observed values of $E[S(T)]$ by month for WGEN (a), GiST (b), and GiSTR (c). For most months and most stations WGEN (Figure 3.9a) and GiST (Figure 3.9b) closely estimate $E[S(T)]$, while GiSTR (Figure 3.9c) overestimates $E[S(T)]$ for all months and stations. In addition, the T-tests on the RMSE of $E[S(T)]$, summarized in Table 3.8, indicate similar results to the T-tests on the RMSE of μ . The test results indicate that RMSE of GiSTR is significantly larger than both GiST and WGEN, while RMSE of GiST is significantly smaller than WGEN.

3.1.3.4. Inter-annual Variability ($Var[S(T)]$)

In contrast to the RMSE of the variance of the daily nonzero precipitation (σ^2) for each generator, RMSE of the inter-annual variability ($Var[S(T)]$) for GiST and GiSTR are larger than WGEN in this domain. The average RMSE for WGEN, GiST and GiSTR is 923.39, 1370.96, and 1610.76 mm^2 respectively. Figure 3.10 shows a comparison of the RMSE of $Var[S(T)]$ between generators for all months. The RMSE for this parameter is smallest for WGEN in all months except September and October. The peak RMSE for each generator is also in August and September. The two-sample T-tests results of the RMSE between each generator, summarized in Table 3.9, indicates that WGEN has a significantly

smaller RMSE than GiST and GiSTR. However, the tests also show that the RMSE for $Var[S(T)]$ is not significantly different between GiST and GiSTR.

One of the interesting results from evaluating how each generator replicates the inter-annual variability ($Var[S(T)]$) is that the RMSE for this parameter is not significantly different between GiST and GiSTR. Recall that GiST has a smaller RMSE than GiSTR when replicating values of the mean daily nonzero precipitation (μ), while GiSTR has a smaller RMSE than GiST when replicating values of σ^2 . Both μ and σ^2 , which are daily parameters, have a direct relationship with $Var[S(T)]$ (Equation 2.8). Therefore, while GiST better replicates μ and GiSTR better replicates σ^2 , the larger RMSE of the opposite parameter (σ^2 for GiST, μ for GiSTR) causes the RMSE of the inter-annual variability for both generators to become approximately the same. Both GiST and GiSTR are less accurate than WGEN in replicating $Var[S(T)]$ for opposite reasons. These opposite errors contribute to both generators having increased error compared to WGEN when simulating $Var[S(T)]$.

3.1.4. Probability Distribution Function Analysis

In addition to the numeric evaluation presented above, the probability distribution functions (PDF) of nonzero precipitation amounts for each station were analyzed for each generator and compared to the observed PDF for each month and station. For the winter months for all three generators, the PDFs of each station are well matched, such as the example for station 315838 in Morganton, NC comparing the observed PDF to those generated by WGEN (a), GiST (b), and GiSTR (c) shown in Figure 3.11. However, with regards to summer months, there are several stations in the coastal areas of the Carolinas where the PDF of GiSTR does not capture the observed PDF. For example, consider the PDFs generated by WGEN (a), GiST (b), and GiSTR (c) compared to the observed PDF for station 319457 in Wilmington, NC for September (shown in Figure 3.12), GiST and WGEN accurately capture the observed PDF. However, GiSTR captures most of the PDF for this station but shows an unrealistic tendency to produce precipitation amounts around 75 mm.

In particular for most stations GiSTR seems to be capped at particular values for each station and is most apparent for summer months. While there are anomalies for particular stations, the PDF of all available nonzero precipitation produced by GiSTR are similar to the theoretical PDFs presented in Chapter 2 for most months. The observed and GiSTR generated PDF for all stations combined for January (a) and September (b) are shown in Figure 3.13. While this shows that GiSTR generated PDF is similar to the theoretical PDF shown in Figure 2.2, the broad overestimation of the frequency of some amounts also indicates the tendency of GiSTR to overestimate values of μ regardless of month.

The capping feature present in some of these PDFs is the artifact of programming code retained from the original implementation of GiST. GiST contains a condition which prevents radically extreme values from appearing. That is, any precipitation amounts which are greater than 110% of the average maximum amount of precipitation are reset to being 110% of the average maximum precipitation amount for that month (Baigorria, 2010, personal communication). While this cap is not visible in the PDFs of GiST and the winter PDFs of GiSTR, this cap is hit more frequently in the summer because much larger amounts of precipitation generally occur in summer and GiSTR re-samples heavy precipitation amounts. This re-sampling causes a spike in the PDF produced by GiSTR. While this cap is designed to assure that a generator will reproduce realistic precipitation amounts in GiST, this cap may accentuate RMSE of μ and σ^2 since increased frequency of amounts at the cap likely causes values of μ to be overestimated and values of σ^2 to be underestimated. Future work might include re-evaluating both GiST and GiSTR with the cap removed.

3.1.5. Dry and Wet Spell Frequency Analysis

The frequency of dry and wet spells of various lengths in days were also examined to determine the accuracy of each generator to replicate dry and wet spells. While there is a seasonal trend in the observed frequencies of dry and wet spells, there is also very little difference in how well each generator replicates these frequencies. Figure 3.14 shows boxplots of the observed frequency of wet spells of various lengths in January (a), compared

to those generated by WGEN (b), GiST (c), and GiSTR (d) for all stations in domain. The wet spells produced by GiST, GiSTR and WGEN generally match the observed wet spells. However, in all three generators, there is a slight overestimation of the frequency of wet spells which are only one day long. In addition, there is a slight underestimation of the frequency of wet spells which are two days long, and accurately replicate the frequency of wet spells three days and longer. Figure 3.15 shows boxplots of the observed dry spells of various lengths in January (a), compared to those generated by WGEN (b), GiST (c), and GiSTR (d) for all stations in the domain. For dry spells in January, each generator overestimates the frequency of one and two day dry spells, but also underestimates the frequency of three day dry spells. In addition, each generator accurately replicates the frequency of dry spells longer than four days.

Figure 3.16 shows boxplots of the observed frequency of wet spells in July (a) compared to those generated by WGEN (b), GiST(c), and GiSTR(d) for all stations in the domain. In July, the wet spells produced by all three generators seem to replicate the observed wet spell frequencies. However, GiST (Figure 3.16c) and GiSTR (Figure 3.16d) both provide more accurate replications of wet spells with lengths of one to two days. WGEN (Figure 3.16b) appears to slightly underestimate the frequency of one day wet spells and overestimate the frequency of two day wet spells. Each generator also adequately captures the frequency of wet spells longer than three days. Figure 3.17 shows boxplots of the observed frequency of dry spells in July (a) compared to boxplots of the generated frequency of dry spells in July for WGEN (b), GiST (c), and GiSTR (d) for all stations in the domain. GiST and GiSTR more accurately replicate the majority of the frequency of dry spells compared to WGEN. WGEN (Figure 3.17b) seems to underestimate the frequency of one day dry spells in the domain, while both GiST and GiSTR underestimate the frequency of two day dry spells. In addition, each generator accurately replicates the frequency of dry spells longer than three days.

Given that GiST and GiSTR use the same process to generate precipitation events and that RMSE of the Markov transition probabilities between these two generators are approximately the same it is understandable that GiST and GiSTR would share similar

patterns for the dry and wet spell frequencies regardless of season. However, GiST and GiSTR provide slightly better replications of the dry and wet spell frequencies compared to WGEN, while WGEN has a significantly lower RMSE for the Markov transition probabilities. Although WGEN does better in replicating the Markov transition probabilities, these probabilities are part of the 1st order chain. That is, probabilities are only conditioned on the previous day. As such, the generator has no “memory” for what has happened two or more days prior to the current day, which has been shown to cause errors when replicating the appropriate dry and wet spell frequencies. This suggests that the orthogonal Markov chain used by GiST and GiSTR provides an additional level of “memory” which allows GiST and GiSTR to better replicate the observed dry and wet spell frequencies than WGEN. This also suggests that although each GiST and GiSTR have a significantly larger error for replicating the temporal parameters than WGEN, the orthogonal Markov chain provides a sufficient means for each generator to replicate both the spatial and temporal structure of precipitation in the domain, since the dry and wet spell frequencies produced by each generator show little difference to the observed frequencies in the domain.

While there are several preliminary conclusions that can be discussed and drawn from the weather generator evaluation in the North Carolina domain, the question remains whether these conclusions will hold over the broader region of the Southeast United States. The next section of this chapter discusses the results of the weather generator evaluation in the Southeast United States, while the final section discusses of the results of the entire weather generator evaluation across both domains.

3.2. Southeast United States Evaluation

Many of the relationships discussed above in the analysis of the various parameters for each generator in the North Carolina domain are also reflected for each generator across the Southeast United States. This section will summarize the results of the weather generator evaluation across the broader Southeast United States, while the final section of the chapter discusses the implications of the results.

3.2.1. Temporal Parameters

The error of the 1st order Markov probabilities (P_{01} and P_{11}) for each generator in this domain shows a similar result to those presented in Section 3.1.1. A comparison of the RMSE of each parameter between each generator for all months, shown in Table 3.10, indicates that there is no seasonal trend in the error for either parameter in this domain. The average RMSE for WGEN is less than GiST and GiSTR by up to 0.06 for values of P_{01} , and by up to 0.37 for values of P_{11} . Regardless of the month considered, the error for each parameter for WGEN is 0.03 to 0.05 less than GiST and GiSTR. In addition the two-sample T-tests for the RMSE of each parameter between generators, summarized in Table 3.11, show similar results to the previous domain. The testing indicates that the error for each parameter simulated by WGEN is significantly less than GiST and GiSTR. However, the two-sample T-test also indicates that GiST has a significantly larger RMSE for values of P_{01} in the Southeast domain than GiSTR. Up to this point, all the T-tests used have been focused primarily on determining if the error for one generator is significantly larger or smaller than another, that is, all the tests have been one-tailed T-tests. In order to determine if any differences were significant, the T-test for the error of GiST versus GiSTR was repeated as two tailed test, which is commonly used for determining significant difference. This additional analysis indicates that while RMSE of P_{01} for GiST for this parameter is significantly larger it is not significantly different (p-value of 0.069 with a level of significance of 0.05) from RMSE of GiSTR for the same parameter.

Similarly to the analysis in Section 3.1.1, the error of the related parameters, the unconditional probability of rain (π) and persistence (γ), for each generator reflects the patterns shown for the Markov transition probabilities in the Southeast U.S. domain. The RMSE of WGEN for each parameter is smaller than GiST and GiSTR by 0.02 to 0.09 on average. A comparison of the RMSE of π and γ between generators, shown in Table 3.12, indicates that WGEN also has the smallest RMSE for each parameter regardless of month. In addition, the comparison also indicates that there is no seasonal trend in the RMSE of either parameter for any generator. The results of the two-sample T-test comparing the RMSE of

these parameters between generators are summarized in Table 3.13. These tests indicate that WGEN has a significantly smaller RMSE for both π and γ than GiST and GiSTR.

While these results also match the results from the smaller domain, the testing also indicates that GiST has a significantly larger RMSE for γ than GiSTR. Following the line of analysis presented for P_{01} , a two-tailed test was also performed to determine significant difference. The follow up analysis for significant difference also shows that while GiST has a significantly larger RMSE, it does not have a significantly different RMSE compared to GiSTR (p-value of 0.061, level of significance of 0.05). As indicated by the results in the North Carolina domain, GiST and GiSTR experience a slightly higher error when replicating the temporal parameters of precipitation events compared to WGEN. As mentioned in Section 1.1 this is possibly the result of the dual focus on the spatial and temporal structure of precipitation events.

3.2.2. Spatial Structure Parameters

Following the approach for the previous domain, the ability of each generator to replicate the spatial structure of precipitation is evaluated. Similarly to the results for the previous domain, WGEN has a larger RMSE, by 0.04 to 0.24 on average, than GiST and GiSTR for every spatial structure parameter. Figure 3.18 shows the RMSE of each generator for the correlation matrix of precipitation (a, ρ), the correlation matrix of precipitation events (b, ρ_{ev}), the correlation matrix of precipitation amounts (c, ρ_{am}), and the correlation matrix of precipitation extreme events (d, ρ_{ex}) for all months for the Southeast U.S. domain. This comparison indicates that the difference between WGEN and GiST or GiSTR is largest for winter months and smallest for summer months for every parameter. These results are consistent with the results presented for the smaller domain. Regardless of month the RMSE for WGEN is also larger than the other generators. Following the earlier analysis, the observed average decay function correlogram is compared to those generated by each generator to determine if each generator reproduces the observed decay in the spatial structure with distance between stations. Figure 3.19 shows the observed average decay for

correlogram compared to those created by each generator for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d) for January in the Southeast U.S. domain. Figure 3.19 shows that WGEN fails to replicate the observed average decay function correlogram for precipitation, precipitation events, precipitation amounts, and precipitation extreme events. GiST and GiSTR more accurately replicate the observed correlogram in all four instances, but best replicates the observed correlogram for precipitation events (Figure 3.19b). For precipitation (Figure 3.19a), precipitation amounts (Figure 3.19c), and precipitation extreme events (Figure 3.19d), GiST and GiSTR show a similar pattern compared to the observed correlogram, but tend to underestimate the observed correlogram. This indicates that like the smaller domain, there is a tendency to underestimate the correlation between stations at most distances for precipitation, precipitation amounts, and precipitation extreme events. This tendency is much less evident during summer months. Figure 3.20 shows the observed average decay function correlogram compared to those created by each generator for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d) for July. Similarly to the previous domain the observed correlogram decays rapidly with distance in response to isolated convection. This also allows WGEN to replicate the observed correlogram for long distances. However, GiST and GiSTR show a similar pattern to the observed correlogram, but show a tendency to underestimate the correlogram. This underestimation is smallest for precipitation amounts (Figure 3.20c) and precipitation extreme events (Figure 3.20d). The tendency to underestimate the observed correlogram shows the tendency of each generator to underestimate the correlation of precipitation and therefore the spatial structure in all four cases. However, GiST and GiSTR are more accurate in warm months than in cold months and are also more accurate than WGEN in representing the decay in spatial structure with distance. These results are consistent with the results presented for the smaller domain. Table 3.14 summarizes the results of the two-sample T-test for the RMSE of each spatial structure parameter between generators. In each T-test, the p-values indicate that the error of WGEN is significantly larger for all four spatial structure parameters. In addition, the T-test also

shows that there is no significant difference in the RMSE between GiST and GiSTR for three of the four spatial structure parameters.

However, while most T-tests for GiST versus GiSTR show no significant difference, the test does show that GiST has a significantly larger RMSE than GiSTR when replicating the correlation matrix of precipitation amounts (ρ_{am}). Given that the primary difference between GiST and GiSTR is the re-sampling of heavy precipitation amounts, this is a possible cause for this difference. In addition, while the T-tests show that RMSE of ρ_{ex} for WGEN is significantly larger than for GiST and GiSTR, a T-test for significant difference (i.e. two tailed) shows that RMSE of the parameter for WGEN is not significantly different from GiST (p-value of 0.062) or GiSTR (p-value of 0.052). These results are contradictory to the results in the North Carolina domain for these parameters. Despite these contradictory results, the general conclusions for the spatial structure and temporal parameters of the North Carolina domain hold in the Southeast U.S. domain:

- WGEN better replicates the temporal parameters (π , γ and the Markov transition probabilities) than GiST and GiSTR, but
- The difference in the RMSE for these parameters between generators is only 0.03 and 0.085 on average.
- GiST and GiSTR better replicate the spatial structure of precipitation in all aspects analyzed.

3.2.3. Precipitation Amount Parameters

While the evaluation of the event parameters for each weather generator follows the same results for both domains, the evaluation for the amount parameters (μ , σ^2 , $E[S(T)]$ and $Var[S(T)]$) show several interesting results in Southeast U.S. domain which are not present in the North Carolina domain.

3.2.3.1. Mean Daily Nonzero Precipitation (μ)

Following the results of the North Carolina domain, GiST has the smallest error for values of μ while GiSTR has the largest error. On average, the RMSE for WGEN, GiST and GiSTR is 1.04, 0.075, and 2.49 mm respectively. A comparison of the RMSE of this parameter between generators for all months is shown in Figure 3.21. GiST also has the smallest error for most months in this domain, while GiSTR always has the largest error, with the peak RMSE in August through October for each generator. The two-sample T-tests for significance, summarized in Table 3.15, show that the GiST has a significantly smaller RMSE than both WGEN and GiSTR. The T-test results also indicate that WGEN has a significantly smaller RMSE than GiSTR. Both of these results are consistent with the results for the North Carolina domain.

While the same overestimation problem is presented by GiSTR in areas of the domain compared to WGEN and GiST, RMSE of GiSTR for this parameter for each month is less in the Southeast U.S. domain than in the smaller North Carolina domain. This suggests that while there are still some stations that are overestimated by large amounts, there are many stations where the value of μ is replicated accurately by GiSTR in the larger domain. As an example, consider the month of August, where the RMSE for WGEN, GiST, GiSTR is 1.05, 1.07, and 3.6 mm respectively. Figure 3.22 shows the difference between observed and generated values of μ for WGEN (a), GiST (b), and GiSTR (c). GiST (Figure 3.22b) and WGEN (Figure 3.22a) do not show a tendency to overestimate or underestimate throughout the domain. In contrast, GiSTR (Figure 3.22c) overestimates for most stations by more than 2 mm and in some cases more than 7 mm, but also has more stations where precipitation is overestimated or underestimated by less than 2 mm. Another pattern shown in Figure 3.22c indicates that GiSTR overestimates values of μ in Florida by larger values than in the remainder of the domain. This pattern persists for June through September, indicating that heavy precipitation amounts are oversampled more in the summer in Florida compared to other areas of the domain. The higher amount of oversampling may influence values of σ^2 in summer. In addition, note that the error of GiST for this parameter is larger in July and

August than WGEN in this domain. Though this difference is less than 0.03, this may be related to a tendency for GiST to underestimate values of μ for stations in central South Carolina and Ohio. Figure 3.22b shows that in August values of μ in these small areas are underestimated by more than 4 mm. Given that there are several stations in Ohio where the underestimation is by more than 4 mm, and that there are underestimations by more than 2 mm along the northern edge of the domain, it is possible that the error is the result of an edge effect and the use of station information outside the domain may improve the representation of the spatial structure in these areas. This in turn could improve the values of μ on the edges of the domain in summer months. While there is not an apparent increase in error at all edges of the domain or during all months, the possible influence of an edge effect suggests that when using GiST or GiSTR in downscaling for seasonal forecasting or as an input for crop models the number of locations used should cover an area larger than the area of interest to reduce the error associated with an edge effect where possible.

The error for stations in central South Carolina is also common only to summer months. There is a possible physical reason that GiST underestimates μ in this region where GiSTR and WGEN do not. Stations in this area are located in the Sandhills region of the Carolinas. The contrasting soil types have been suggested by several researchers to cause a sea-breeze like circulation which contributes to the total precipitation in this region (Wootten et al, 2010; Raman et al. 2005). Given that in the larger domain the structure of precipitation in these months is primarily unstructured convection, this circulation, which occurs in a small region and is not as strong as an actual sea-breeze circulation (and events resulting from this circulation occur more rarely than isolated events), it is possible that GiST does not capture the spatial structure associated with this circulation and therefore underestimates values of μ for these stations. The lack of spatial structure in WGEN may allow it the flexibility to account for this phenomenon, and GiSTR compensates for this spatial structure through the re-sampling of heavy precipitation amounts, which are also associated with these events. However, while the use of GiST in this domain may not capture this small area structure, it is possible, given the results for the spatial structure in the North Carolina domain, that using GiST just in this small area may allow it to capture the spatial structure caused by this

Sandhills circulation. In addition, GiST does not show the same tendency to underestimate for other stations in or near the Sandhills, which suggests that further research is required to determine an exact physical cause for the error of μ in central South Carolina. This also suggests that for applications in regions of a sharp soil or vegetation contrast it would be appropriate to apply GiST and GiSTR in small domain surrounding this region compared to a larger domain.

3.2.3.2. Variance of Daily Nonzero Precipitation (σ^2)

In southeast U.S. domain, each generator shows similar error to the results presented for the North Carolina domain in Section 3.1.3.2. The error for values of σ^2 is largest for GiST and smallest for GiSTR. On average, the RMSE for WGEN, GiST and GiSTR is 85.28, 124.08, and 81.18 mm² respectively. The RMSE of this parameter between generators for each month is compared in Figure 3.23. This comparison shows that GiSTR has the smallest RMSE for most months while GiST has the largest RMSE for all months. In addition, the peak RMSE for each generator is in September and October. In this domain, the T-tests for difference in the RMSE of this parameter between generators, summarized in Table 3.16, also indicate similar results to the smaller domain. The error of GiSTR is significantly less than for GiST. In addition, the RMSE for σ^2 is not significantly different between GiSTR and WGEN. However, contrary to the results in the North Carolina domain, the tests in the Southeast domain also show that RMSE of σ^2 for GiST is significantly larger than WGEN.

The results from this set of tests and analyses show similar results to the North Carolina domain; as such it is reasonable to assume that the larger RMSE values in September and October also reflects an underestimation of the values of σ^2 across the domain. However, while there are some underestimates along the East Coast, the largest underestimates are along the Northern Gulf Coast of Alabama and Florida. Using September as an example, the RMSE for σ^2 for WGEN, GiST and GiSTR is 133.66, 188.67, and 141.75 mm² respectively. The difference between observed and generated values of σ^2 for each

station for WGEN (a), GiST (b), and GiSTR (c) for September is shown in Figure 3.24. For all three generators there is a general underestimation of σ^2 across the domain, with the largest errors along the East Coast (up to 500 mm²) and the Northern Gulf Coast (over 700 mm²). WGEN (Figure 3.24a) and GiST (Figure 3.24b) do not show any other spatial patterns across the domain. In addition, GiST underestimates σ^2 by up to 500 mm² for more stations than both WGEN and GiSTR during September. Figure 3.24c indicates that GiSTR underestimates σ^2 by 0 to 100 mm² in the northern portions of the domain, underestimates by approximately 100 to 300 mm² in the central portion of the domain, and *overestimates* by 0 to 100 mm² in most of Florida. The pattern of greater underestimation of σ^2 is present throughout the year, but is most strongly observed in June through September. In addition, the overestimation of σ^2 in Florida is only observed during June through September. Neither of these patterns is present in the analysis of σ^2 for WGEN or GiST. The overestimation of σ^2 by GiSTR in Florida in the summer months also corresponds to an overestimate of μ by GiSTR in the same location and time period. In the case of Florida the main reason for this double overestimate during the summer by GiSTR involves the re-sampling of heavy precipitation. In Florida the oversampling of heavy precipitation causes the generated values of μ and σ^2 to be much larger than in other portions of the domain. This suggests that GiSTR re-samples the higher precipitation amounts more frequently than in other areas of the domain. This results in a large overestimate of μ and a slight overestimate σ^2 in Florida.

3.2.3.3. Average Total Precipitation ($E[S(T)]$)

In contrast to the North Carolina domain, the error the average total precipitation ($E[S(T)]$) and the inter-annual variability ($Var[S(T)]$) for each generator better reflects RMSE of their daily counterparts, μ and σ^2 . The average RMSE of $E[S(T)]$ for WGEN, GiST, and GiSTR is 11.64, 12.16 and 39.71 mm respectively. As in the North Carolina domain, GiSTR has the largest error for $E[S(T)]$ while GiST and WGEN have similar error. Figure 3.25 shows a comparison of the RMSE of this parameter between generators for every month. GiSTR has the largest RMSE for every month, and the peak RMSE for each generator is in

August through October. The two-sample T-test results for the difference in the RMSE of this parameter between generators are summarized in Table 3.17. As with the North Carolina domain, the T-tests show that GiSTR has a significantly larger RMSE than both WGEN and GiST. However, unlike the North Carolina domain there is no significant difference for RMSE for $E[S(T)]$ between GiST and WGEN.

Similarly to RMSE of μ for GiSTR in this domain, RMSE of $E[S(T)]$ for GiSTR is smaller than in the Southeast domain than in the North Carolina domain (average RMSE for North Carolina domain is 63.51 mm, average RMSE for Southeast domain is 39.71 mm). Since there is a direct relationship between μ and $E[S(T)]$, the same reason is the likely cause of this improvement. That is, there are more stations where GiSTR replicates $E[S(T)]$ with smaller error, which results in an RMSE that is less than in the North Carolina domain. In contrast to the North Carolina domain, the error of $E[S(T)]$ is slightly larger for GiST than WGEN in the Southeast domain. GiST has an average RMSE of 8.43 mm in North Carolina domain and an average RMSE of 12.16 mm in the Southeast domain. This is also in contrast to the difference in the error of μ between GiST (0.075 mm) and WGEN (1.04 mm) in the Southeast U.S. domain. Given the direct relationship between μ and $E[S(T)]$ (Equation 2.7), and the previous discussion regarding the summer underestimations of μ , it is possible the error for μ in these areas contributes to the increased error of $E[S(T)]$ by aggregation to the monthly timescale. However, given that the T-test results show that there is no significant difference in the error between GiST and WGEN, both are capable of replicating the average total precipitation ($E[S(T)]$) for stations across the region.

3.2.3.4. Inter-annual Variability ($Var[S(T)]$)

The error pattern of the inter-annual variability ($Var[S(T)]$) for each generator does show some distinct differences from the results for the North Carolina domain. In the previous domain, the error of GiST and GiSTR are shown to be larger than WGEN. On average, the RMSE of $Var[S(T)]$ for WGEN, GiST, and GiSTR is 873.97, 1288.31, and 836.89 mm² respectively in this domain. Therefore, in this domain GiSTR has the smallest

error for replicating $Var[S(T)]$, while GiST has the largest error. Figure 3.26 shows a comparison of the RMSE of $Var[S(T)]$ between generators for each month. The RMSE of $Var[S(T)]$ is smallest for GiSTR for most months and largest for GiST in all months with the peak RMSE for each generator in September. In addition, the two-sample T-test results, summarized in Table 3.18, indicate that WGEN has a significantly lower RMSE than GiST, which is consistent with results from the North Carolina domain. However, GiSTR also has a significantly lower RMSE than GiST, which is not consistent with results from the North Carolina domain. Given that $Var[S(T)]$ is directly related to both μ and σ^2 (Equation 2.8) it becomes apparent why RMSE of $Var[S(T)]$ for GiSTR improved on average between domains. While RMSE of σ^2 increased by approximately 23.2% (65.85 mm² to 81.18 mm²) with the increase in the number of stations used, RMSE of μ decreased by 48.4% (4.83 mm to 2.49mm). This sharp decrease caused RMSE of $Var[S(T)]$ for GiSTR to decrease between the North Carolina domain and the Southeast domain by 48% (from 1610.76 mm² to 836.89 mm²). However, as mentioned previously the error of GiSTR for both μ and σ^2 is larger in Florida than in the rest of the domain in summer as compared to WGEN and GiST. Therefore, the RMSE of GiSTR for $Var[S(T)]$ is also larger in Florida in summer than in the rest of the Southeast domain. Given this information, it is fair to assume that the re-sampling approach of GiSTR has lower RMSE when replicating $Var[S(T)]$ in the majority of the Southeast domain, but suffers higher error in all amount parameters during summer in Florida.

3.2.4. Probability Distribution Function Analysis

Following the analysis from the North Carolina domain, the PDFs produced by each weather generator vary significantly between seasons, but also between stations in the domain. In most cases all three weather generators replicate much of the observed PDFs for stations throughout the domain for winter months. For example, consider the PDFs produced by each generator compared to the observations for several stations in January. A comparison between in the observed and generated PDFs for WGEN (a), GiST (b), and

GiSTR (c) for station 310301 in Asheville, NC is shown in Figure 3.27. For this station, each generator accurately replicates much of the PDF of nonzero precipitation in January. However, both WGEN (Figure 3.27a) and GiST (Figure 3.23b) show a tendency to underestimate the frequency of amounts between 10 and 20 mm. GiSTR (Figure 3.27c) shows a tendency to overestimate the frequency of amounts between 20 and 30 mm, which also shows a cap around 30 mm. In addition, consider a comparison between the observed and generated PDFs for WGEN (a), GiST (b), and GiSTR (c) for station 460102 in Alderson, WV shown in Figure 3.28. The PDFs for each generator also replicate much of the observed PDF for this station in January. However, GiSTR (Figure 3.28c) again shows a tendency to hit the capping value around 25 mm. Finally, consider the comparison between the observed and generated PDFs for WGEN (a), GiST (b), and GiSTR (c) for station 089176 in Venice, FL in January shown in Figure 3.29. This comparison again indicates that each generator replicates most of the observed PDF for this station. However, it is again apparent that both GiST (Figure 3.29b) and GiSTR (Figure 3.29c) hit the capping value for this station. Despite the ability of each generator to replicate much of observed PDF for each station, it is evident in this comparison that GiST and GiSTR reach their capping values regardless of location. In addition it is also apparent that while both GiST and GiSTR reach their capping values even in January, the re-sampling of heavy precipitation causes GiSTR to reach the capping values more frequently. This indicates that values of μ can be overestimated during any season by GiSTR, though with larger overestimations in warm months.

As an example for warm months consider the PDFs produced for the same stations used previously in September. Figure 3.30 shows a comparison between the observed and generated PDFs for WGEN (a), GiST (b), and GiSTR (c) for station 310301 in September for Asheville, NC. In Asheville, each weather generator produces similar results and GiSTR already shows a tendency to reach the capping value in cold months. In September, each generator replicates the majority of the PDF accurately. However, GiST (Figure 3.30b) and GiSTR (Figure 3.30c) are frequently hitting the capping value for this station. In addition, the fact that the value of the PDF is larger at the capping value for GiSTR than GiST indicates that GiSTR will more frequently produce values above this value. Figure 3.31

shows a comparison between the observed and generated PDFs for WGEN (a), GiST (b), and GiSTR (c) for station 460102 in September for Alderson, WV. For the Alderson, WGEN (Figure 3.31a) and GiST (Figure 3.31b) accurately reproduce the PDF for September. However, it is also apparent that while GiSTR (Figure 3.31c) capture most of the PDF, it also frequently hits the capping value in September. Figure 3.32 shows a comparison between the observed and generated PDFs for WGEN (a), GiST (b), and GiSTR (c) for station 089176 in September for Venice, FL. For this station, the PDFs produced by each weather generator capture most of the observed PDF. However, there are also clear instances where the PDF of each generator failed to replicate the observed PDF in Florida. For instance, GiST (Figure 3.32b) in September seems to provide the best replication of the PDF, while WGEN (Figure 3.32a) has some distinct overestimations and underestimations of the PDF for some values less than 20 mm. Finally, GiSTR (Figure 3.32c) replicates most of the PDF for stations in Florida, but also meets its cap more often than for other stations in the domain. This tendency is common for multiple stations in summer and is most strongly visible in Florida. The more frequent occurrence of values at the capping value mentioned previously is indicative of the overestimation of μ and σ^2 in summer in Florida.

3.2.5. Dry and Wet Spell Frequency Analysis

Dry and Wet spell frequencies generated by each generator are also compared against observations in the Southeast U.S. domain, and show similar trends to the dry and wet spell frequencies produced by each generator in the North Carolina domain. Boxplots of the observed frequency of wet spells of various lengths in January (a), compared to those generated by WGEN (b), GiST (c), and GiSTR (d) for all stations in the domain are shown in Figure 3.33. In January each generator underestimates the frequency of one day wet spells while overestimating the frequency of two day wet spells. Each generator also accurately replicates the frequency of wet spells three days and longer. Figure 3.34 show boxplots of the observed frequency of dry spells of various lengths in January (a) compared to those generated by WGEN (b), GiST (c), and GiSTR (d) for all stations in the domain. For these

dry spells, each generator overestimates the frequency of two-day dry spells, but replicates the frequency of the lengths of other dry spells in the domain. These patterns are common among all cold months, suggesting that each generator accurately replicates most wet and dry spell frequencies during the cold months at each station.

Boxplots of the observed frequency of wet spells of various lengths in July (a) compared to those generated by WGEN (b), GiST (c), and GiSTR (d) for all stations in the domain are shown in Figure 3.35. For July wet spells, WGEN (Figure 3.35b) underestimates the frequency of one day wet spells while overestimating two day wet spells. This pattern is present in GiST (Figure 3.35b) and GiSTR (Figure 3.35c), but with much less magnitude in the Southeast domain than WGEN. Each generator also accurately captures wet spells of three days and longer. Finally, Figure 3.36 show boxplots of the observed frequency of dry spells of various lengths in July (a) compared to those generated by WGEN (b), GiST (c), and GiSTR (d) for all stations in the domain. For these dry spells in July, WGEN underestimates the frequency of one-day dry spells, while both GiST and GiSTR overestimate the frequency of one-day dry spells across the domain. These patterns are also similar for all warm months. In addition each generator replicates the frequency of the longest dry and wet spells at each station accurately during both the warm and cold months. In section 3.2.1 it is shown that the RMSE for values of the temporal parameters is significantly larger for GiST and GiSTR than WGEN. However, the increase in error is also shown to be less than 0.09 on average. The results in this section indicate that there is little difference between generators for accurately replicating the frequency of dry and wet spells of different lengths. These results, consistent with those presented in Section 3.1, suggests that although the error for the temporal parameters is significantly larger for GiST and GiSTR, there is little difference in the dry and wet spell frequencies or the number of precipitation events produced. As such, GiST and GiSTR are capable of replicating the spatial and temporal structure of precipitation, which might have value in downscaling seasonal precipitation forecasts.

3.3. Summary Discussion

For each of these domains there are distinct results and conclusions that can be drawn about the effectiveness of each generator. There are also several avenues where future research may improve upon each generator. This section summarizes the results of the weather generator evaluation in each domain and their implications. A listing of which generator exhibits the lowest error for every parameter in both domains is shown in Table 3.20.

The evaluation shows that for temporal parameters (P_{01} , P_{11} , π , γ) in both domains, WGEN has the lowest error on average compared to other generators. However, for the spatial structure of precipitation including the raw spatial structure, structure of precipitation events, amounts, and extreme events, GiST and GiSTR have lower error than WGEN for replicating the observed spatial structure in both domains. For replicating those parameters which reflect daily nonzero precipitation amounts, μ and σ^2 , the generator with the lowest error on average is GiST and GiSTR respectively in both domains. However, for the monthly counterparts of these daily parameters, $E[S(T)]$ and $Var[S(T)]$, there are several discrepancies between domains. For the North Carolina domain, the generator with the lowest error on average for $E[S(T)]$ is GiST. For the Southeast U.S. domain, WGEN has the lowest error on average for $E[S(T)]$. However, the T-tests for this parameter indicated that there is no significant difference in the error of this parameter between GiST and WGEN. Finally, for the North Carolina domain the generator with the lowest error when replicating $Var[S(T)]$ (which characterizes the inter-annual variability) is WGEN, while the generator with the lowest error for this parameter in the Southeast domain is GiSTR.

For those parameters reflecting the temporal structure, it is likely that the reason WGEN has the lowest error for these parameters is that the entire focus for generating precipitation events is on the temporal structure of precipitation. GiST and GiSTR focus on both the spatial and temporal structure of precipitation amounts. GiST and GiSTR use an orthogonal Markov chain to generate precipitation events, and as a result the spatial and temporal structures are used in tandem during the generation process. This interactive

approach affects the error of the generator for replicating both the spatial structure parameters and the temporal parameters. The result of this interaction is a large improvement in error with regards to the spatial structure and a slight increase in error for temporal parameters of precipitation events. In contrast, WGEN only uses the 1st order Markov chain for generating precipitation events. As a result of the lack of interaction between the spatial structure and temporal structure in the generator process, the error for the temporal parameters is slightly smaller for WGEN. However, because WGEN does not consider spatial structure in the event generation process, the error of WGEN is much larger for reproducing the spatial structure of precipitation. Therefore, while a dual focus creates slightly larger error with regards to the temporal parameters, GiST and GiSTR compensate for this error by improving the spatial structure.

In addition, though the error for the temporal parameters for GiST and GiSTR is significantly more than WGEN, the difference in RMSE is less than 0.09 on average and the frequency of dry and wet spells are well replicated by all three generators despite the larger RMSE produced for the four temporal parameters by GiST and GiSTR. As mentioned previously, it is possible that this is because of the issue of “memory”. The 1st order Markov Chains are only conditioned upon the events occurring at the same location the previous day. As such, there have been a few arguments suggesting that WGEN and other 1st order Markov chain generators have a lack of “memory” for those events occurring several days to several weeks prior to the current day. This is often the source of one of the main disadvantages for Markov chain generators. Given that GiST and GiSTR also replicate dry and wet spell frequencies while RMSE for the temporal parameters is larger suggests that the orthogonal Markov chain used by GiST and GiSTR provides another level of “memory” to allow for multiple cases where GiST and GiSTR better replicates observed dry and wet spell frequencies particularly for wet spells in summer. While it is possible that GiST and GiSTR provide better representations of dry and wet spell frequencies for this reason, the analysis in this study shows that there is very little difference in the dry and wet spell frequencies produced by each generator. Given this information, each generator is capable of reproducing the occurrence of precipitation events. However, for applications to forecasting

and to crop modeling, it is recommended that GiST and GiSTR be used, given their ability to replicate both the spatial and temporal structure of precipitation.

GiST has been shown to have lower error on average when replicating μ than WGEN and GiSTR. The only difference between GiST and WGEN is the consideration of the spatial structure during the event and amount generation process. Therefore, the addition of the spatial structure during the generation process also improves the ability of a weather generator to replicate values of μ . However, this does not indicate that the PDFs or temporal variance of precipitation are replicated accurately. Given that GiST shows the largest error in each domain for σ^2 , this suggests that while GiST replicates low precipitation amounts accurately it does not replicate large precipitation amounts as often as WGEN and GiSTR. Since nonzero precipitation amounts follow a gamma distribution (small precipitation amounts occur much more frequently than large amounts), it is possible that GiST accurately captures the frequency of small amounts while not capturing the frequency of large amounts. The result of this combination yields the accurate estimation of μ and the overall underestimation of σ^2 common to all three weather generators. Given that the difference between GiST and WGEN is only the spatial structure this combination suggests that the spatial structure built into the GiST generation process causes GiST to provide a better estimate the mean daily nonzero precipitation (μ) but a worse estimate of the variance of the daily nonzero precipitation (σ^2). It is also shown that GiSTR has lower error when replicating σ^2 and higher error when replicating μ .

The results of the analysis suggest that the re-sampling of heavy precipitation amounts done by GiSTR produces opposite results to the evaluation of GiST. While the GiSTR better replicates the frequency of larger precipitation amounts, these values are over-sampled by the generator. This causes μ to be overestimated and σ^2 to be underestimated by a smaller magnitude for most stations and overestimated for some stations. However, as mentioned previously, GiST and GiSTR contain a capping measure to prevent values that are unrealistic from occurring. However, since GiSTR produces heavy precipitation more frequently than GiST, this cap is reached more frequently, increasing the frequency of precipitation amounts at this cap. As a result, the PDFs produced by GiSTR become more

unrealistic, despite the ability to replicate the values of σ^2 . In addition, as described in Chapter 2, GiSTR uses an additional gamma distribution to model precipitation extremes. While this additional gamma distribution does improve values of σ^2 it also overlaps with the gamma distribution that models smaller precipitation amounts. Therefore, precipitation amounts where the two overlap are oversampled by GiSTR. The result of the capping value and the oversampling by GiSTR causes an overestimate of μ in both domains. The error of μ , when aggregated to the monthly timescale causes GiSTR to overestimate $E[S(T)]$ across both domains. However, the error of both μ and $E[S(T)]$ produced by GiSTR decreases from the North Carolina domain to the Southeast U.S. domain. There are several possible reasons for this result. First, while several stations from the North Carolina domain are included in the Southeast U.S. domain, the more stringent quality control requirements used for the Southeast U.S. domain may have caused some stations used in the smaller domain to be dropped from the Southeast domain. While not immediately evident from this analysis, some of the stations that may have been removed likely are locations with the largest error in the North Carolina domain, resulting in GiSTR producing lower error for both μ and $E[S(T)]$. However, considering those sixty-nine stations which are common to both domains does not reduce the error of μ and $E[S(T)]$ for GiSTR in the North Carolina domain. Second, the larger number of stations used and the stations used in the northern portions of the domain may allow that more stations have both μ and $E[S(T)]$ replicated accurately causing the average error to be less across the domain. Third, the larger number of stations also increased the information used by the spatial structure component of events and amounts, allowing for better replication of the observed distribution of precipitation amounts. This possibility is unlikely since the increase in station caused the error between domains for the same parameters for GiSTR to increase. Finally, for $Var[S(T)]$, the error of GiSTR also improves on average from the smaller domain to the larger domain. Given that $Var[S(T)]$ is directly related to μ and σ^2 , the 48% improvement in $Var[S(T)]$ between domains matches a 48.4% improvement in μ between domains.

For GiSTR and WGEN there are very few visible spatial patterns in the error with regards to μ , σ^2 , $E[S(T)]$. For most months, μ and $E[S(T)]$ are either underestimated or

overestimated by small amounts (1 to 2mm and 0 to 15 mm respectively). In addition, for most months GiST and WGEN underestimate values of σ^2 across the domain, with the largest error on the northern Gulf Coast in summer. Regardless of the generator used, there is a slight increase in error for values of σ^2 for stations along the coast during the peak of the Atlantic hurricane season. This suggests that regardless of the generator used, all generators produce a greater underestimate of σ^2 for coastal stations related to a lack of tropical precipitation. While this underestimation is reduced by GiSTR, this generator shows a clear pattern in summer to underestimate σ^2 slightly in the northern portions of the Southeast domain, overestimate σ^2 in Florida, and underestimate σ^2 by larger values in the middle of the Southeast domain. This difference might be related to the differences in climatology between areas in the domain. Florida and the northern Gulf Coast have a more tropical climate than the rest of this domain and receive the largest annual precipitation. The central portion of the domain is temperate to sub-tropical, with no dry season and mild winters. However, the northern portion of the domain has lower temperatures and experience harsher winters with less unstructured and tropical convection. These subtle differences may result in the difference in the observed values of σ^2 ; the northern portion of the domain in summer has the smallest observed values of σ^2 , the largest values are in the central portion of the domain, and the values of σ^2 are smaller in Florida in general than in the rest of the domain. This suggests that the more frequent occurrence of precipitation in this region causes the values of σ^2 to decrease from the central portions of the domain. The less frequent occurrence of precipitation in the central portion of the domain causes observed values σ^2 in the central portion of the domain to be larger than other areas.

Finally, while precipitation likely occurs with equal frequency in the northern portions of the domain it is possible that the heavy precipitation amounts which occur in the central and southern portions of the domain occur less frequently, which causes observed values σ^2 to decrease in the northern portion of the domain compared to other portions of the domain (such as the example for September shown in Figure 3.37). However, while there are these subtle differences in the frequency of precipitation and precipitation extremes, what is defined as extreme precipitation changes between the northern and southern portions of the

domain. Recall, that GiSTR uses an additional gamma distribution to model the top 10% of precipitation amounts. Given that the value of 90th percentile is larger in Florida during summer (in September values of 39.37 mm in Florida versus values of 17.93 mm for some North Carolina stations), it is also possible that GiSTR over-samples precipitation at these much higher values more frequently than observed in Florida in summer causing an overestimation of σ^2 in Florida. This over-sampling of values, which are in most cases greater than 25 mm, may also cause GiSTR to overestimate values of μ in summer in Florida by values of 10mm or more.

Both of these are possibilities since precipitation occurs more frequently in summer. The amounts remain small, which keeps observed values of μ similar to other values in the domain. The more frequent occurrence of small amounts causes the observed values of σ^2 in Florida to decrease compared to other stations in the Southeast domain. The lower values of the 90th percentile in North Carolina indicate that GiSTR will re-sample lower values than for stations in Florida, which causes the underestimation of values of σ^2 in the central portion of the North Carolina domain. Finally, the values of the 90th percentile in the northern portion of the domain are generally between 19 and 25 mm, which may allow GiSTR to better estimate observed values of σ^2 in these areas. However, this argument is only speculation and future work should include further analysis of this particular feature in GiSTR.

While GiST suffers higher error for the variance of daily nonzero precipitation and slightly higher error for the 1st order Markov transition probabilities, it captures mean daily precipitation amounts, average monthly total precipitation, dry and wet spell frequencies, and the spatial structure of precipitation more accurately than GiSTR and WGEN in the North Carolina domain and the Southeast U.S. domain. While WGEN matches GiST and GiSTR for the dry and wet spell frequencies and 1st order transition probabilities, GiST and GiSTR are both capable of reproducing the spatial structure of precipitation in both domains. However, GiST fails to capture much of variance of daily nonzero precipitation compared to GiSTR. The high error in mean daily nonzero precipitation and average monthly total precipitation shows that while GiSTR captures the variance and precipitation extremes, it does not provide an ideal solution for modeling the entire distribution of precipitation

amounts across the Southeast U.S. Therefore, future work in this area should include the application of different approaches to capturing extreme precipitation. Possible methods include a hybrid distribution approach (Furrer and Katz, 2008), a mixed distribution approach (Wilks and Wilby, 1999), and the use of different distributions.

Given that it accurately reproduces the temporal and spatial structure of precipitation in multiple domains and provides the best replication of the mean daily nonzero precipitation and the average total precipitation, GiST is recommended for seasonal forecasting and crop modeling in near term simulations. However, since GiST has the highest error for replicating the variance of daily nonzero precipitation and the inter-annual variability, it should be noted that GiST would likely have high error for rare or cyclic events, such as the effect of an active Atlantic hurricane season. Therefore, it is recommended that for situations where high variability may be a concern, GiSTR should be used for forecasting and crop modeling. However, it is important to note that the capping value present in both GiST and GiSTR does have effect on the simulated precipitation produced by GiSTR. For future research, it is recommended that this analysis be repeated with the capping value increased or removed, which may improve the simulation of precipitation produced by both GiST and GiSTR.

Table 3.1. RMSE comparison of generated Markov Transition probabilities (P_{0I} and P_{1I}) values for the NC domain.

month	P_{0I}			P_{1I}		
	WGEN	GiST	GiSTR	WGEN	GiST	GiSTR
1	0.017	0.037	0.043	0.027	0.055	0.040
2	0.017	0.046	0.049	0.032	0.043	0.046
3	0.017	0.056	0.046	0.030	0.062	0.067
4	0.017	0.063	0.068	0.034	0.073	0.062
5	0.016	0.065	0.066	0.025	0.072	0.055
6	0.020	0.038	0.050	0.029	0.058	0.064
7	0.018	0.057	0.058	0.031	0.062	0.060
8	0.018	0.057	0.063	0.034	0.077	0.065
9	0.017	0.056	0.071	0.036	0.079	0.071
10	0.014	0.066	0.069	0.035	0.071	0.084
11	0.020	0.053	0.057	0.026	0.054	0.061
12	0.015	0.071	0.070	0.026	0.064	0.061
average	0.017	0.055	0.059	0.030	0.064	0.061

Table 3.2. Summary of Results from the two-sample T-tests with the 1st order Markov Transition Probabilities (P_{0I} and P_{1I}) for the NC domain. P-values in red indicate that the difference in RMSE between the two generators compared is significantly less than zero. The Difference in average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.

Generators compared		P_{0I}		P_{1I}	
		Difference in Average RMSE	P-value	Difference in Average RMSE	P-value
WGEN	GiST	-0.038	4.44E-08	-0.034	2.92E-08
WGEN	GiSTR	-0.042	3.17E-09	-0.031	1.52E-07
GiST	GiSTR	-0.0038	0.19	0.0028	0.27

Table 3.3. RMSE comparison of generated unconditional probability of rain (π) and persistence (γ) values for the NC domain.

Month	π			γ		
	WGEN	GiST	GiSTR	WGEN	GiST	GiSTR
1	0.018	0.016	0.022	0.032	0.087	0.078
2	0.020	0.027	0.030	0.036	0.081	0.087
3	0.020	0.029	0.032	0.033	0.11	0.099
4	0.020	0.041	0.043	0.037	0.12	0.11
5	0.017	0.034	0.035	0.029	0.13	0.11
6	0.020	0.028	0.025	0.035	0.084	0.11
7	0.022	0.020	0.028	0.032	0.12	0.11
8	0.022	0.023	0.032	0.037	0.13	0.12
9	0.022	0.029	0.049	0.041	0.13	0.13
10	0.017	0.045	0.045	0.040	0.13	0.15
11	0.020	0.036	0.032	0.034	0.094	0.11
12	0.017	0.057	0.051	0.030	0.11	0.11
average	0.020	0.032	0.036	0.035	0.11	0.11

Table 3.4. Summary of Results from the two-sample T-tests with unconditional probability of rain (π) and persistence (γ) for the NC domain. P-values in red indicate that the difference in RMSE between the two generators compared is significantly less than zero. The Difference in average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.

Generators compared		π		γ	
		Difference in Average RMSE	P-value	Difference in Average RMSE	P-value
WGEN	GiST	-0.012	0.0014	-0.075	6.36E-09
WGEN	GiSTR	-0.016	4.64E-05	-0.076	3.48E-09
GiST	GiSTR	-0.0035	0.21	-0.00041	0.48

Table 3.5. Summary of results of two-sample T-tests for each correlation matrix (p-values only) for the NC domain. P-values in red indicate that RMSE of the first generator is significantly larger than the second generator.

Generators Compared		ρ	ρ_{ev}	ρ_{am}	ρ_{ex}
WGEN	GiST	2.64E-06	7.42E-09	6.45E-06	0.00046
WGEN	GiSTR	2.49E-06	7.35E-09	5.23E-06	0.00013
GiST	GiSTR	0.35	0.43	0.23	0.15

Table 3.6. Summary of results of the two-sample T-test on the generated values of the mean daily nonzero precipitation (μ) for the North Carolina domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.

Generators Compared		Difference in Average RMSE (mm)	P-value
WGEN	GiST	0.59	7.49E-08
WGEN	GiSTR	-3.79	2.1E-05
GiST	GiSTR	-4.38	5.47E-06

Table 3.7. Summary of results of the two-sample T-test on the generated values of the variance of daily nonzero precipitation (σ^2) for the NC domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.

Generators Compared		Difference in Average RMSE (mm ²)	P-value
WGEN	GiST	-41.81	0.055
WGEN	GiSTR	14.83	0.24
GiST	GiSTR	56.64	0.016

Table 3.8. Summary of results of the two-sample T-test on the generated values of the average total precipitation ($E[S(T)]$) for the NC domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.

Generators Compared		Difference in Average RMSE (mm)	P-value
WGEN	GiST	3.80	0.00017
WGEN	GiSTR	-51.27	5.46E-06
GiST	GiSTR	-55.08	2.87E-06

Table 3.9. Summary of results of the two-sample T-test on the generated values of the inter-annual variability ($Var[S(T)]$) for the NC domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.

Generators Compared		Difference in Average RMSE (mm^2)	P-value
WGEN	GiST	-447.57	0.028
WGEN	GiSTR	-687.37	0.017
GiST	GiSTR	-239.80	0.23

Table 3.10. RMSE comparison of generated Markov Transition probabilities (P_{0I} and P_{1I}) values for the Southeast U.S.

month	P_{0I}			P_{1I}		
	WGEN	GiST	GiSTR	WGEN	GiST	GiSTR
1	0.018	0.068	0.068	0.027	0.056	0.055
2	0.019	0.067	0.069	0.031	0.070	0.055
3	0.019	0.071	0.061	0.030	0.057	0.058
4	0.019	0.075	0.069	0.030	0.076	0.066
5	0.019	0.078	0.059	0.029	0.059	0.064
6	0.019	0.071	0.059	0.027	0.066	0.060
7	0.019	0.059	0.053	0.028	0.073	0.064
8	0.018	0.058	0.051	0.029	0.076	0.063
9	0.018	0.088	0.069	0.034	0.075	0.065
10	0.015	0.10	0.071	0.035	0.064	0.078
11	0.018	0.061	0.065	0.030	0.057	0.056
12	0.018	0.083	0.083	0.028	0.069	0.061
average	0.018	0.073	0.065	0.030	0.067	0.062

Table 3.11. Summary of Results from the two-sample T-tests with the Markov Transition Probabilities for the Southeast U.S. domain. P-values in red indicate that the difference in RMSE between the two generators compared is significantly less than zero.

Generators Compared		P_{0I}		P_{1I}	
		Average RMSE	P-value	Average RMSE	P-value
WGEN	GiST	-0.055	7.43E-09	-0.037	4.83E-10
WGEN	GiSTR	-0.047	6.52E-10	-0.032	2.84E-11
GiST	GiSTR	0.0087	0.035	0.0045	0.067

Table 3.12. RMSE comparison of generated unconditional probability of rain (π) and persistence (γ) values for the Southeast U.S. domain.

Month	π			γ		
	WGEN	GiST	GiSTR	WGEN	GiST	GiSTR
1	0.017	0.038	0.038	0.033	0.11	0.11
2	0.021	0.043	0.043	0.035	0.12	0.11
3	0.018	0.044	0.036	0.037	0.11	0.11
4	0.019	0.051	0.040	0.036	0.13	0.12
5	0.019	0.047	0.030	0.035	0.12	0.11
6	0.018	0.035	0.025	0.034	0.13	0.11
7	0.019	0.030	0.025	0.034	0.12	0.11
8	0.018	0.034	0.022	0.036	0.12	0.11
9	0.020	0.058	0.041	0.040	0.15	0.13
10	0.018	0.076	0.048	0.039	0.15	0.14
11	0.020	0.037	0.048	0.033	0.11	0.099
12	0.018	0.054	0.057	0.033	0.13	0.12
average	0.019	0.046	0.038	0.035	0.12	0.11

Table 3.13. Summary of Results from the two-sample T-tests with unconditional probability of rain (π) and persistence (γ) for the Southeast US domain. P-values in red indicate that the difference in RMSE between the two generators compared is smaller or larger than zero, depending on the Difference in the Average RMSE.

Generators Compared		π		γ	
		Difference in Average RMSE	P-value	Difference in Average RMSE	P-value
WGEN	GiST	-0.02687	7.43E-09	-0.08855	8.84E-12
WGEN	GiSTR	-0.01899	3.92E-05	-0.07921	3.03E-12
GiST	GiSTR	0.007881	0.058	0.009348	0.030

Table 3.14. Summary of results of one tailed two-sample T-tests for each correlation matrix (p-values only) for the Southeast US domain. P-values in red indicate that RMSE of the first generator is significantly larger than the second generator.

Generators Compared		ρ	ρ_{ev}	ρ_{am}	ρ_{ex}
WGEN	GiST	0.00016	2.13E-08	0.00013	0.031
WGEN	GiSTR	7.91E-05	2.77E-08	1.31E-05	0.027
GiST	GiSTR	0.31	0.14	0.0068	0.28

Table 3.15. Summary of results of the two-sample T-test on the generated values of the mean daily nonzero precipitation (μ) for the Southeast U.S. domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.

Generators Compared		Difference in Average RMSE (mm)	P-values
WGEN	GiST	0.30	0.00053
WGEN	GiSTR	-1.44	2.26E-06
GiST	GiSTR	-1.75	1.50E-07

Table 3.16. Summary of results of the two-sample T-test on the generated values of the variance of daily nonzero precipitation (σ^2) for the Southeast U.S. domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE is the average RMSE of the second generator subtracted from the average RMSE of the first generator.

Generators Compared		Difference in Average RMSE (mm ²)	P-values
WGEN	GiST	-38.79	0.0061
WGEN	GiSTR	4.11	0.39
GiST	GiSTR	42.90	0.0093

Table 3.17. Summary of results of the two-sample T-test on the generated values of the average total precipitation ($E[S(T)]$) for the Southeast U.S. domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE is the average RMSE of the second generator subtracted from the Average RMSE of the first generator

Generators Compared		Difference in Average RMSE (mm)	P-values
WGEN	GiST	-0.52	0.30
WGEN	GiSTR	-28.07	9.97E-10
GiST	GiSTR	-27.55	4.57E-11

Table 3.18. Summary of results of the two sample T-test on the generated values of the inter-annual variability ($Var[S(T)]$) for the Southeast U.S. domain. P-values in red indicate RMSE of one generator is larger than the other. The Difference in Average RMSE (mm^2) is the average RMSE of the second generator subtracted from the first generator.

Generators Compared		Difference in Average RMSE (mm^2)	P-values
WGEN	GiST	-414.34	0.00026
WGEN	GiSTR	37.08	0.37
GiST	GiSTR	451.42	0.00073

Table 3.19 .Summary results of which generator has the lowest error on average for each parameter tested.

Parameter	North Carolina Domain	Southeast U.S. Domain
1st Order Markov Transition Probabilities (P_{01}, P_{11})	WGEN	WGEN
Unconditional Probability of Rain (π)	WGEN	WGEN
Persistence (γ)	WGEN	WGEN
Spatial Structure ($\rho, \rho_{ev}, \rho_{am}, \rho_{ex}$)	GiST/GiSTR	GiST/GiSTR
Mean Daily Nonzero Precipitation (μ)	GiST	GiST
Variance of Daily Nonzero Precipitation (σ^2)	GiSTR	GiSTR
Average Monthly Total Precipitation ($E[S(T)]$)	GiST	GiST/WGEN
Variance of the Monthly Total Precipitation / Inter-annual Variability ($Var[S(T)]$)	WGEN	GiSTR

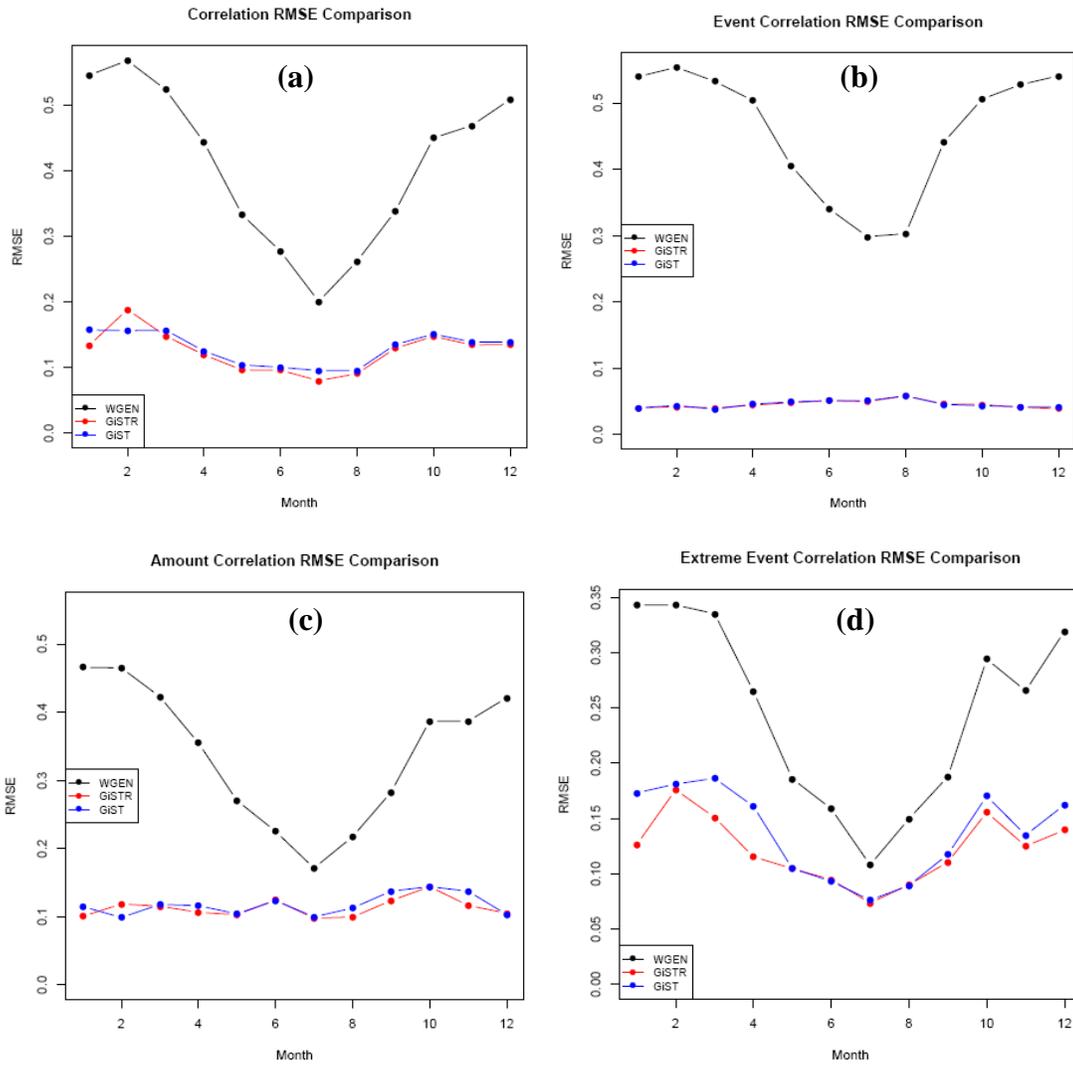


Figure 3.1. RMSE of each weather generator by month for the correlation matrix of precipitation (ρ) (a), the correlation matrix of precipitation events (ρ_{ev})(b), the correlation matrix of precipitation amounts (ρ_{am})(c), and the correlation matrix of precipitation extreme events (ρ_{ex})(d) for North Carolina domain.

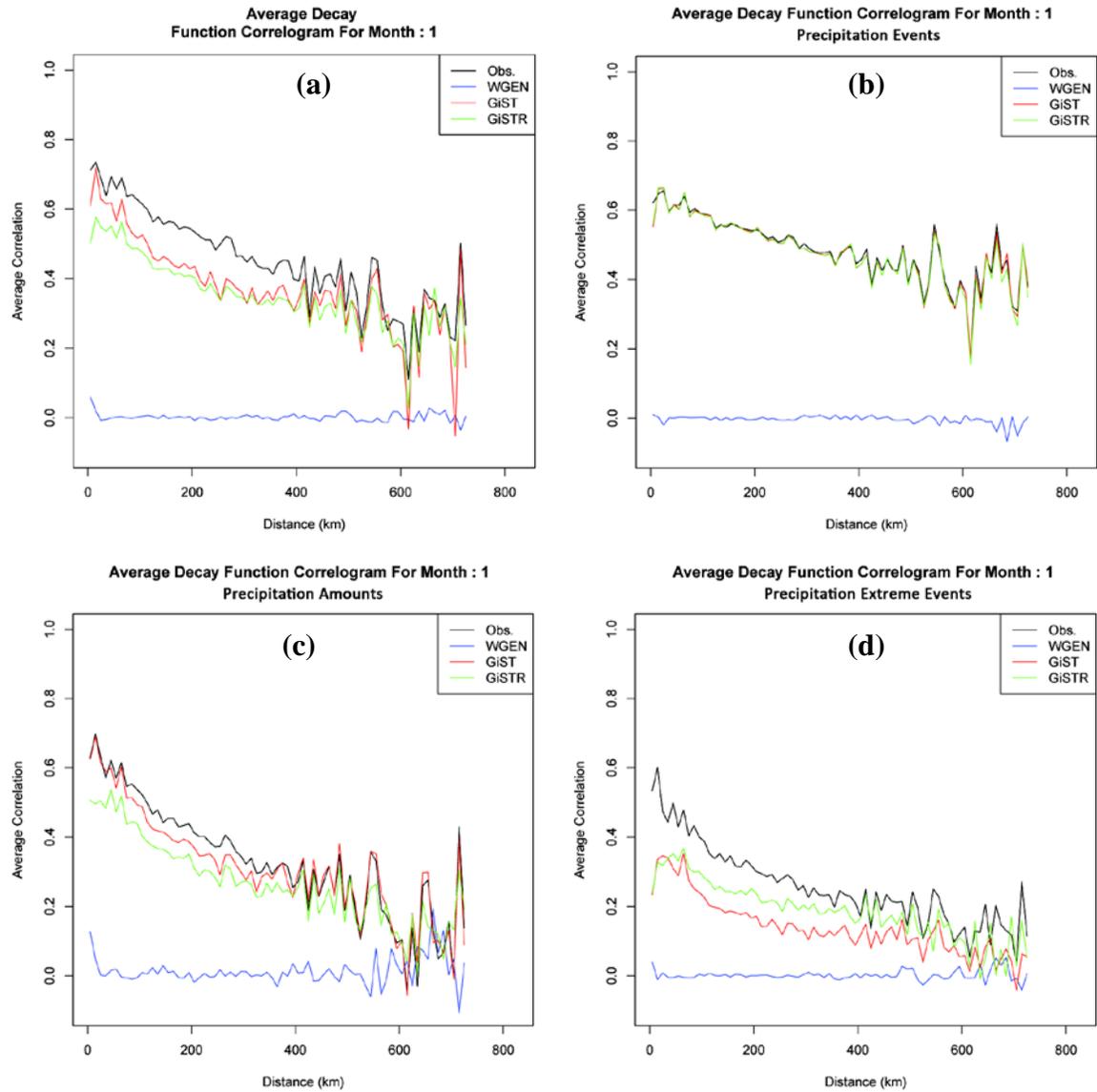


Figure 3.2. Average Decay Function Correlogram for each generator and observations for January for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). North Carolina domain.

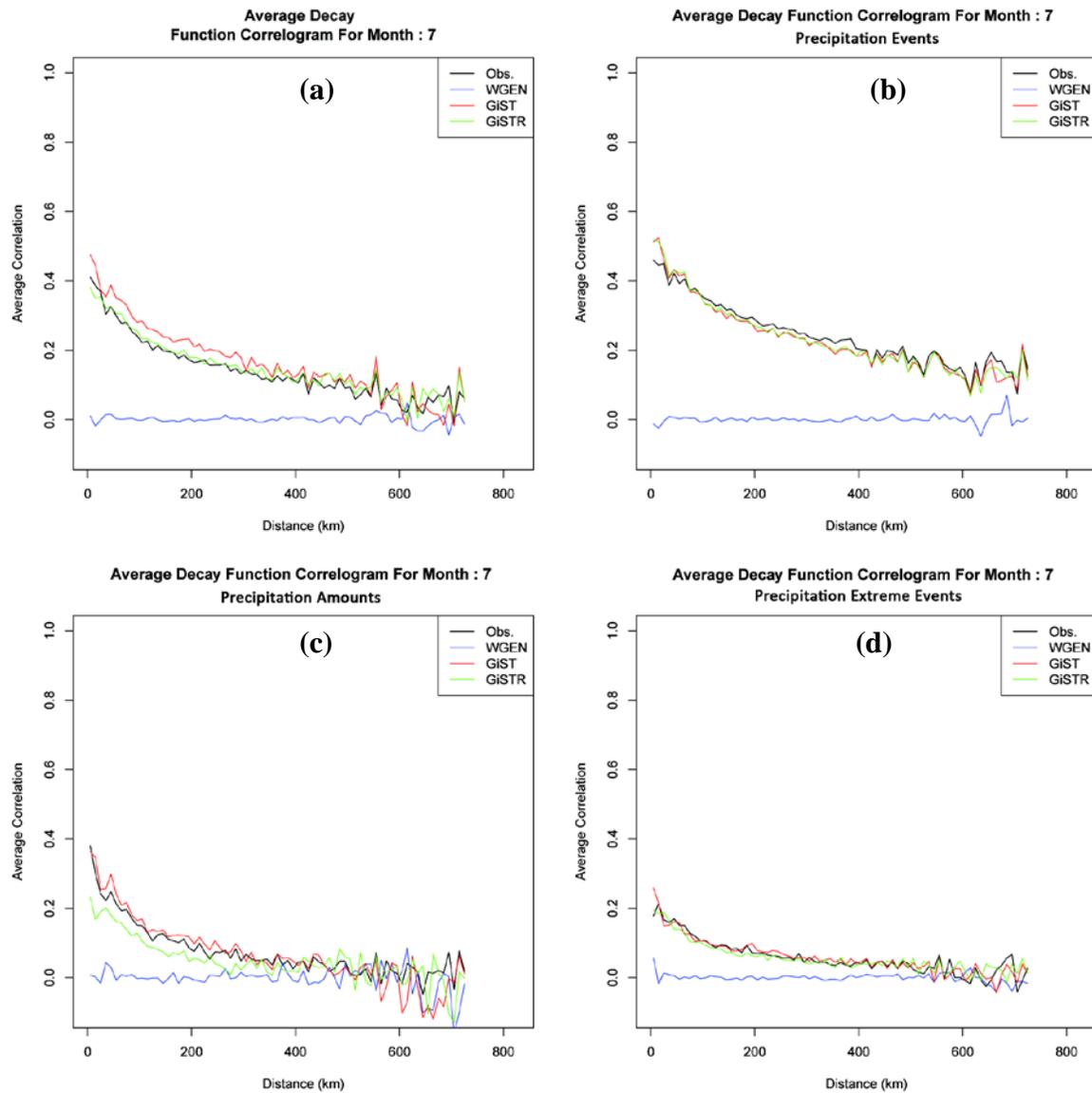


Figure 3.3. Average Decay Function Correlogram for each generator and observations for July for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). North Carolina domain.

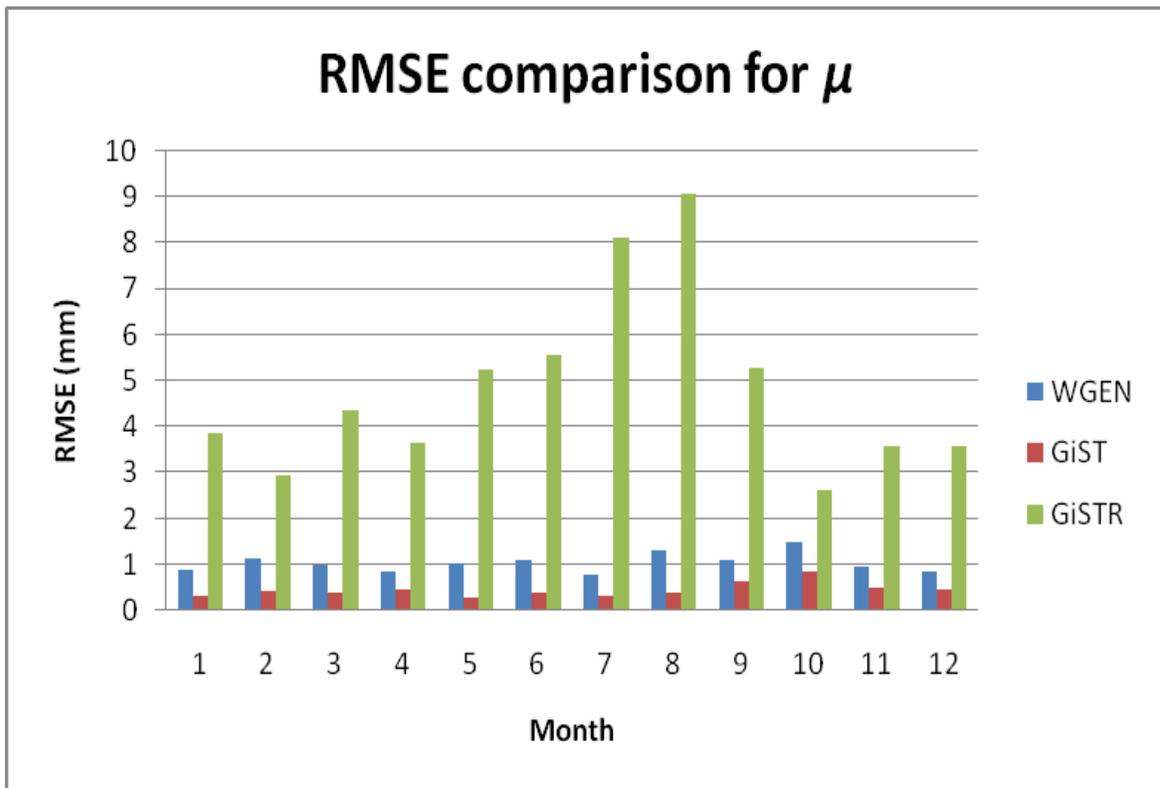
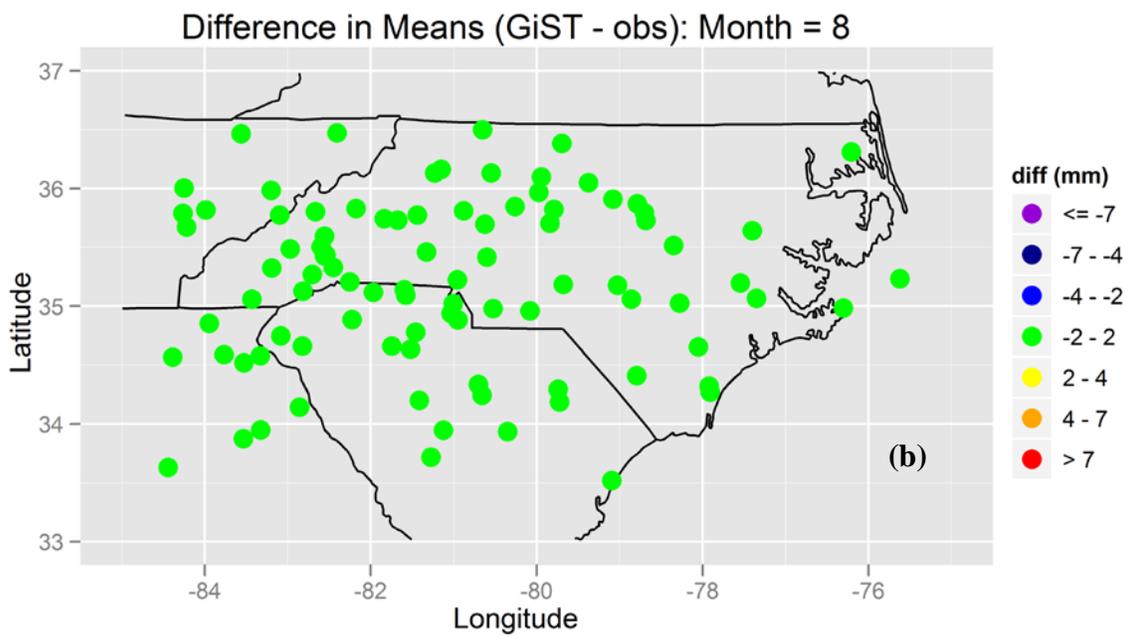
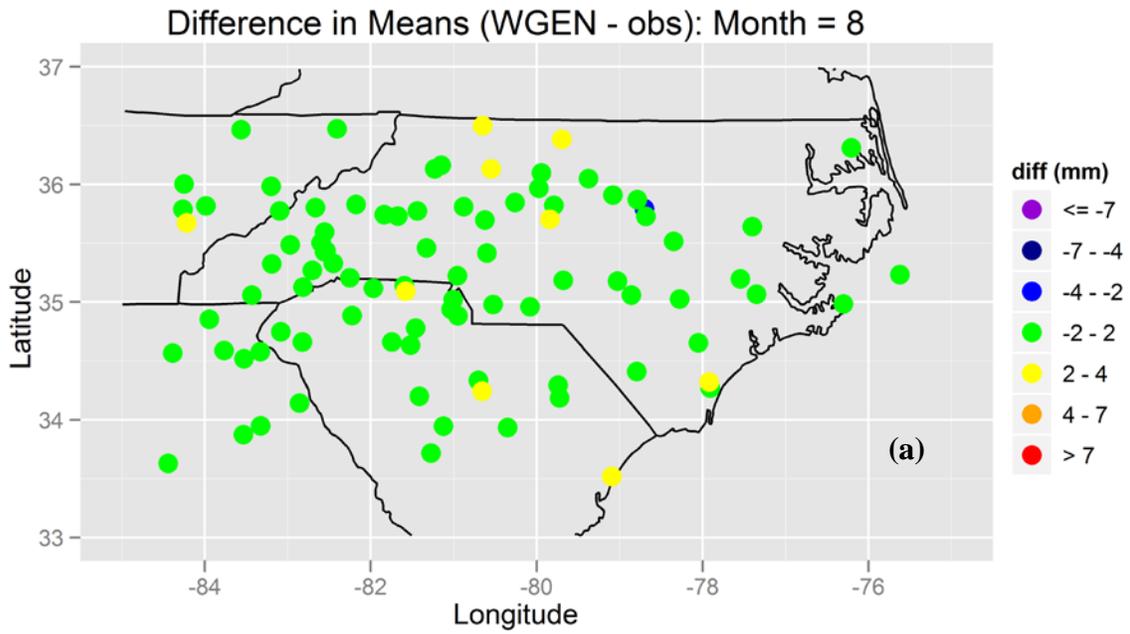
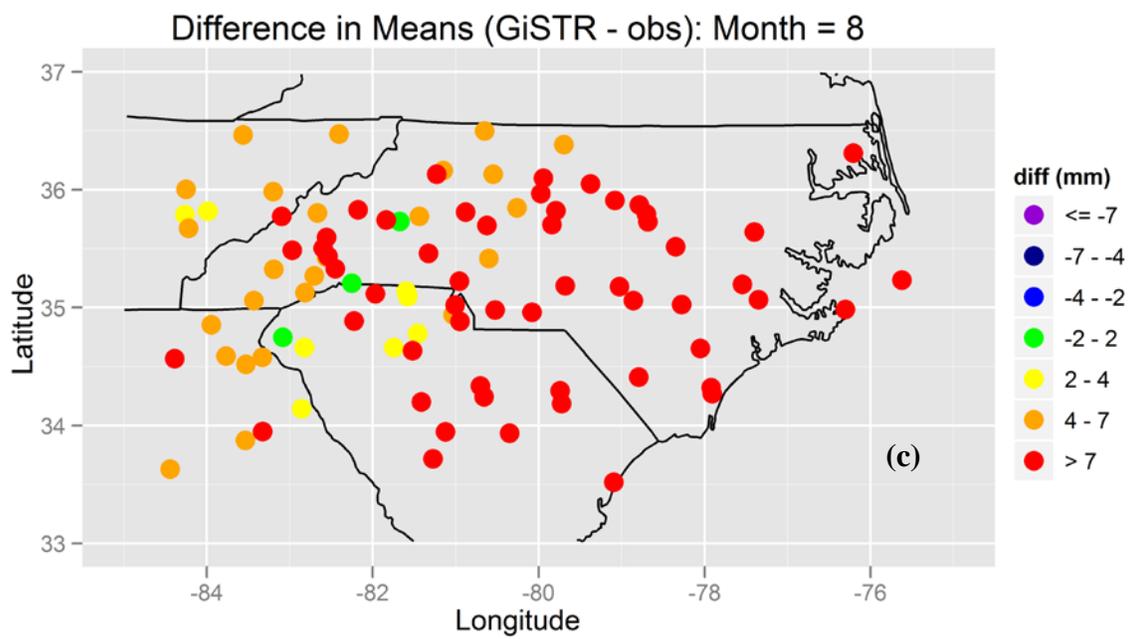


Figure 3.4. RMSE (mm) comparison for the generated values of mean daily nonzero precipitation (μ) for North Carolina domain.

Figure 3.5. Difference between generated and observed values of mean daily nonzero precipitation (μ) across the North Carolina domain for the month of August for WGEN(a), GiST(b), and GiSTR (c)





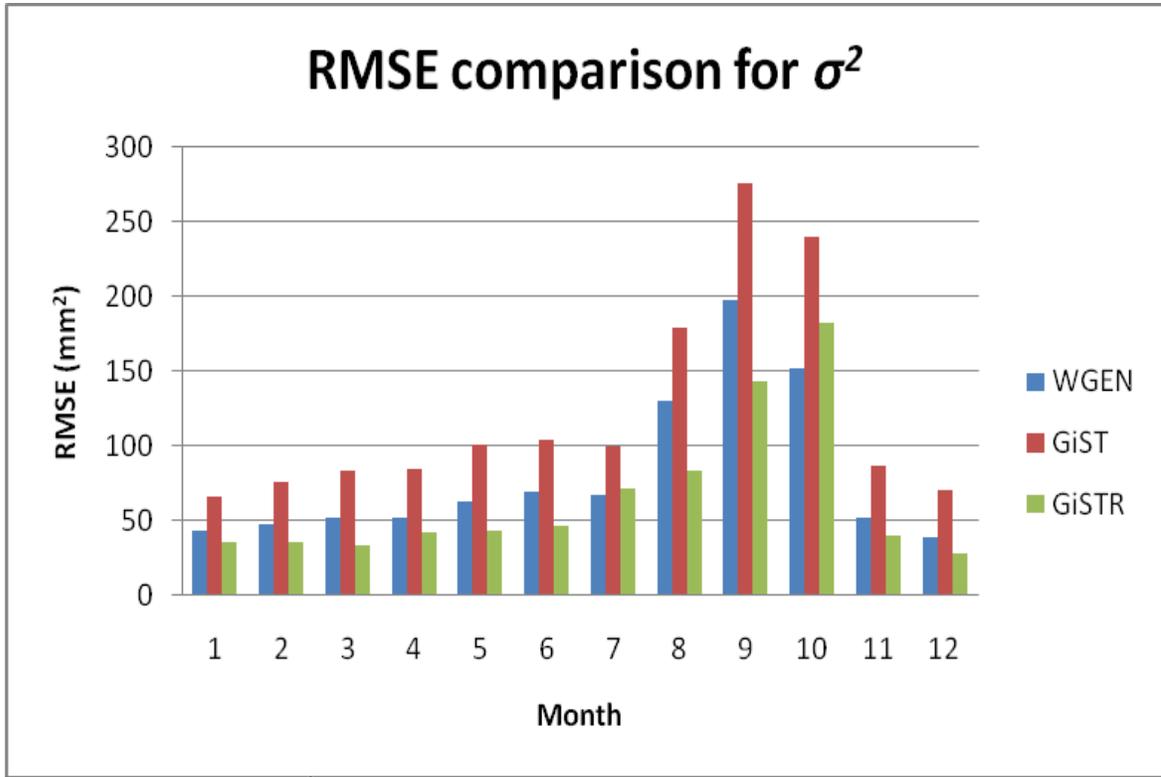
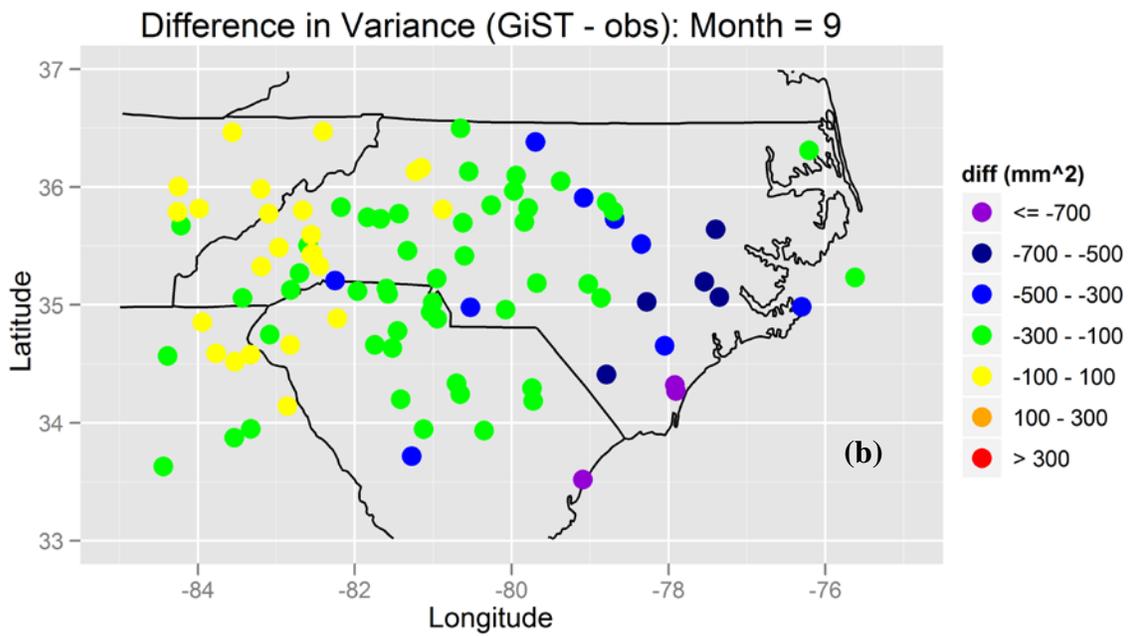
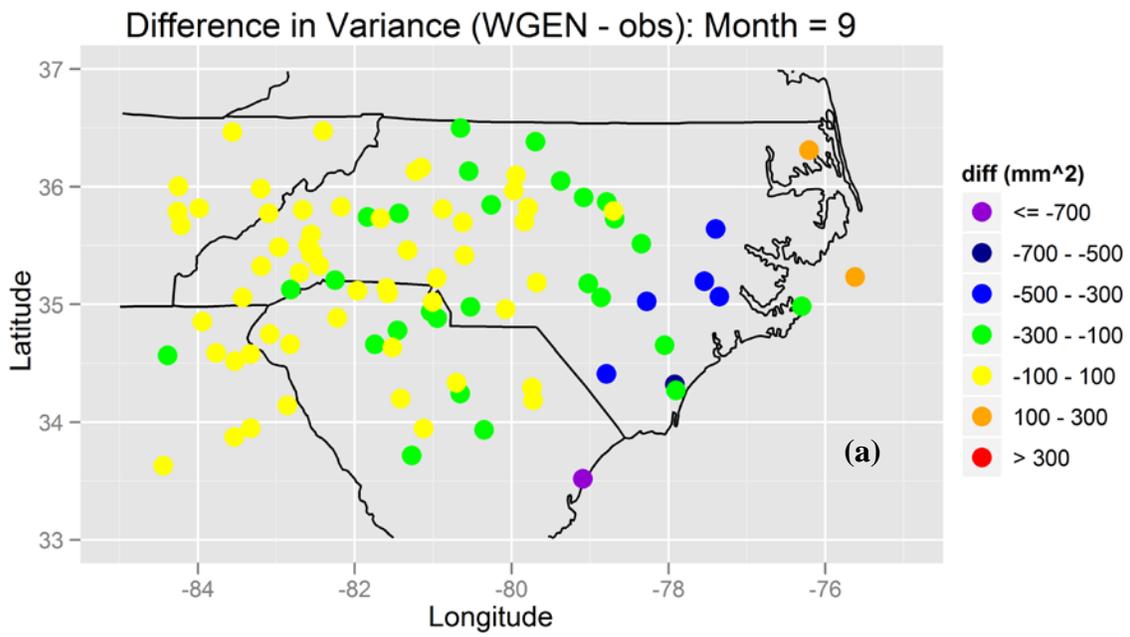
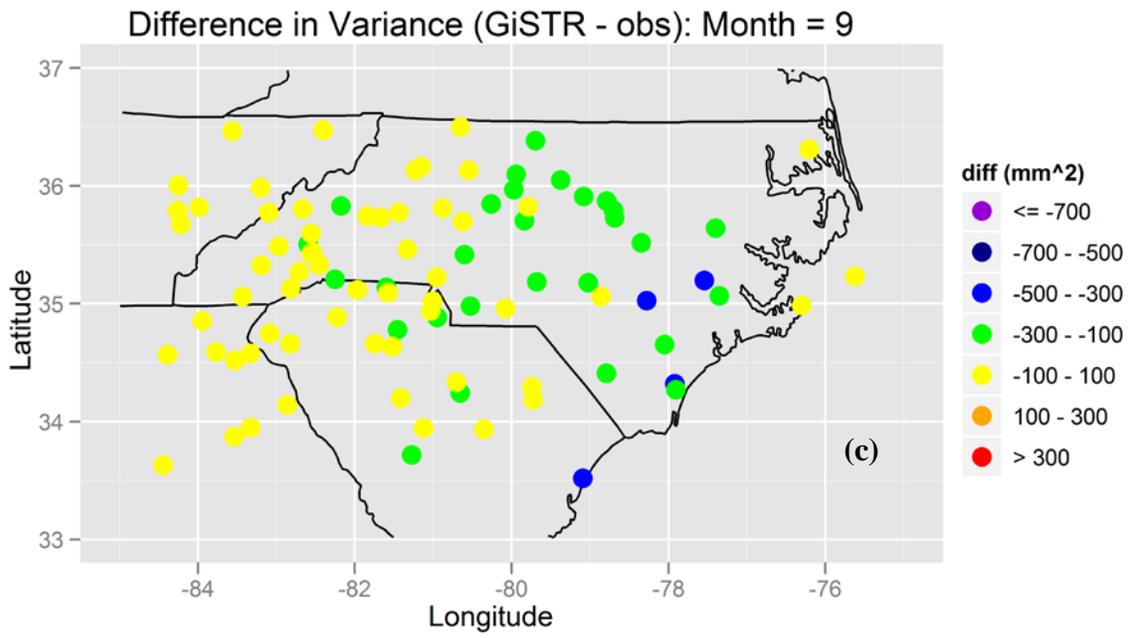


Figure 3.6. RMSE (mm²) comparison for the generated values of the variance of the daily nonzero precipitation (σ^2) for North Carolina domain.

Figure 3.7. Difference between generated and observed values of the variance of daily nonzero precipitation (σ^2) across the North Carolina domain for the month of September for WGEN(a), GiST(b), and GiSTR (c)





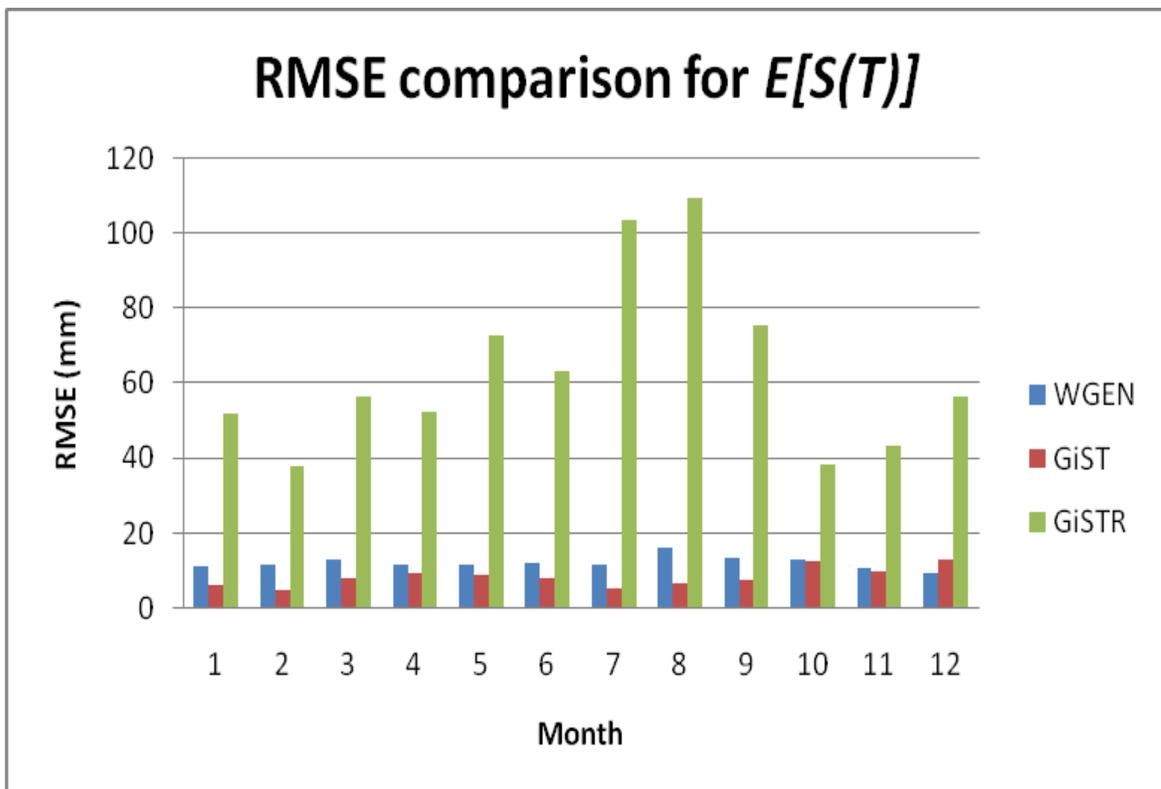


Figure 3.8. RMSE (mm) comparison for generated values of the average total precipitation ($E[S(T)]$) for North Carolina domain.

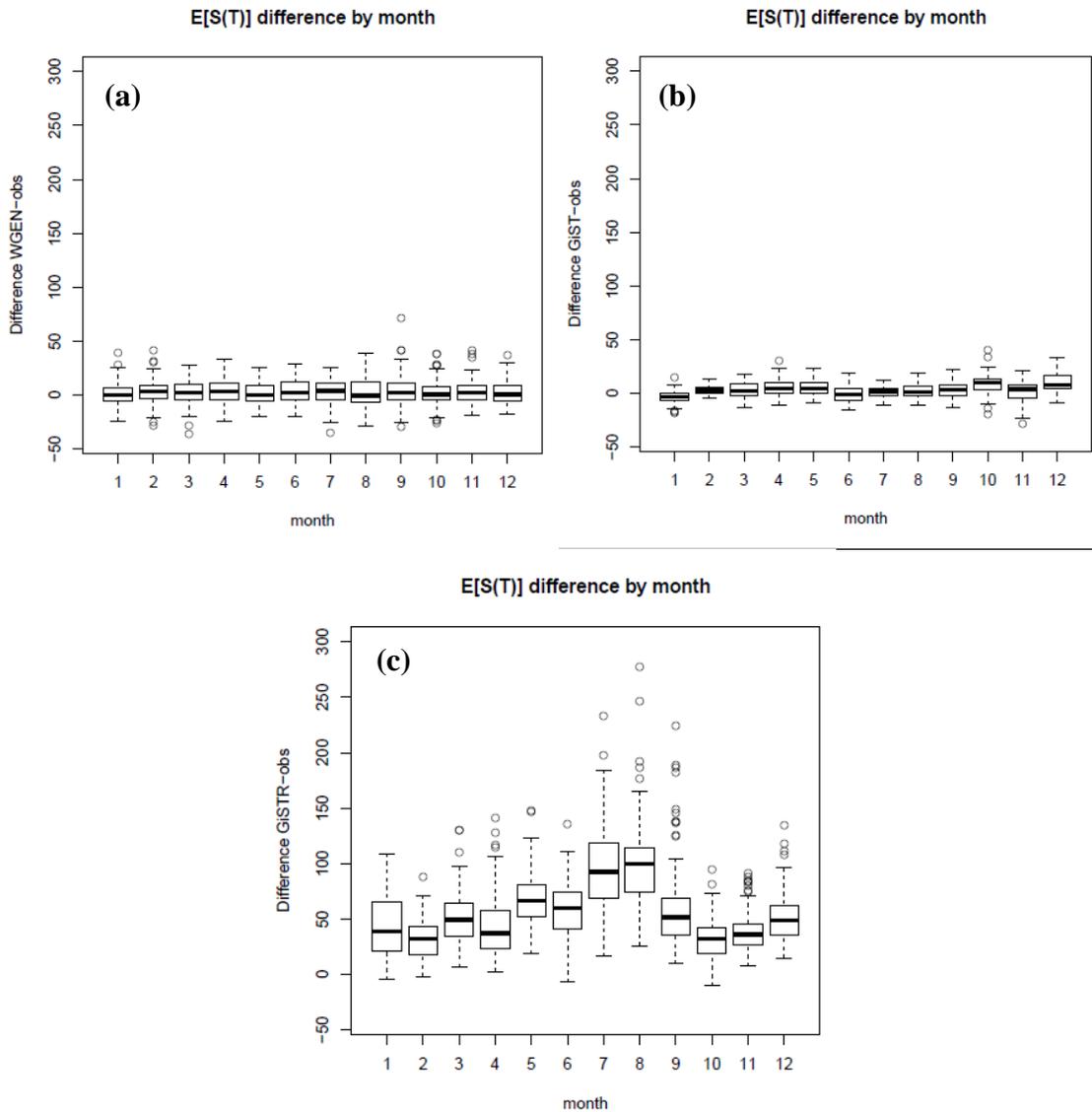


Figure 3.9. Boxplots of the difference between generated and observed values of the average total precipitation ($E[S(T)]$) for WGEN (a), GiST (b), and GiSTR(c) for North Carolina domain.

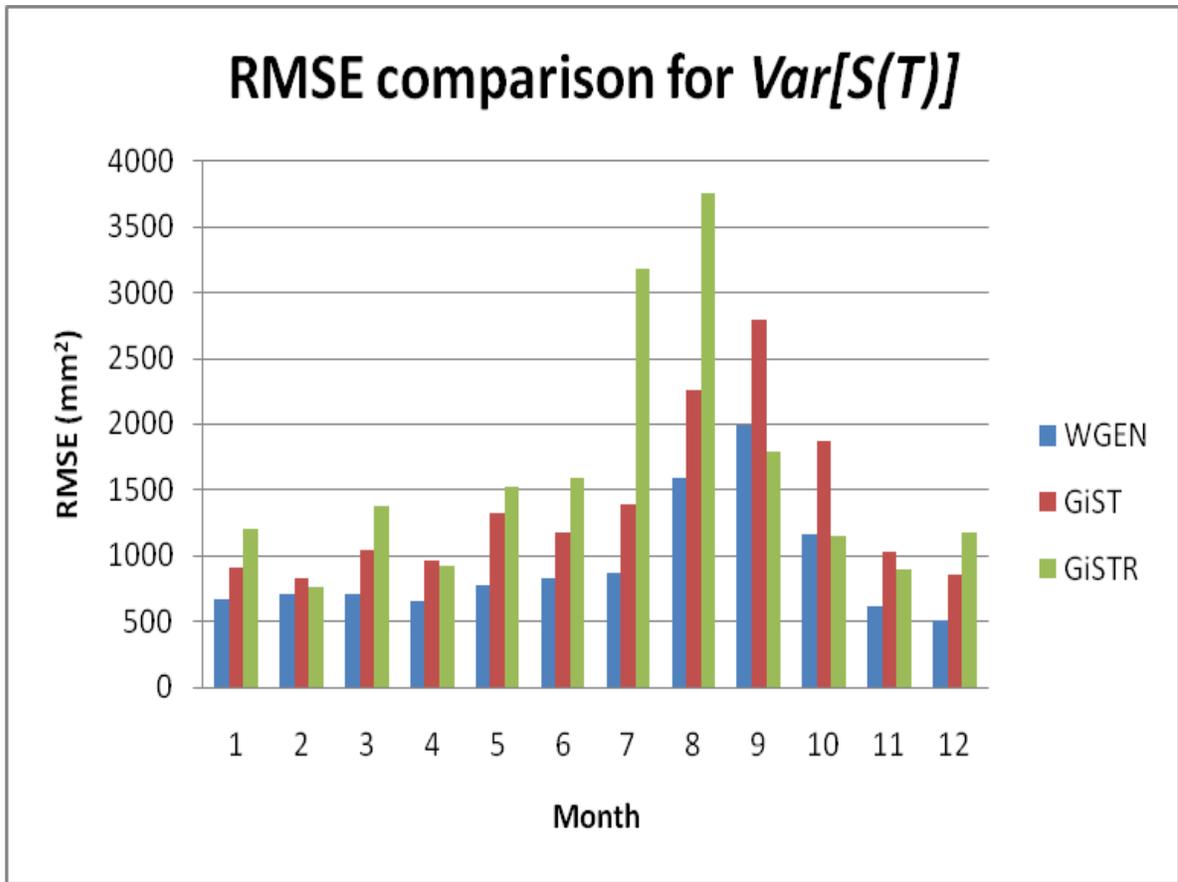
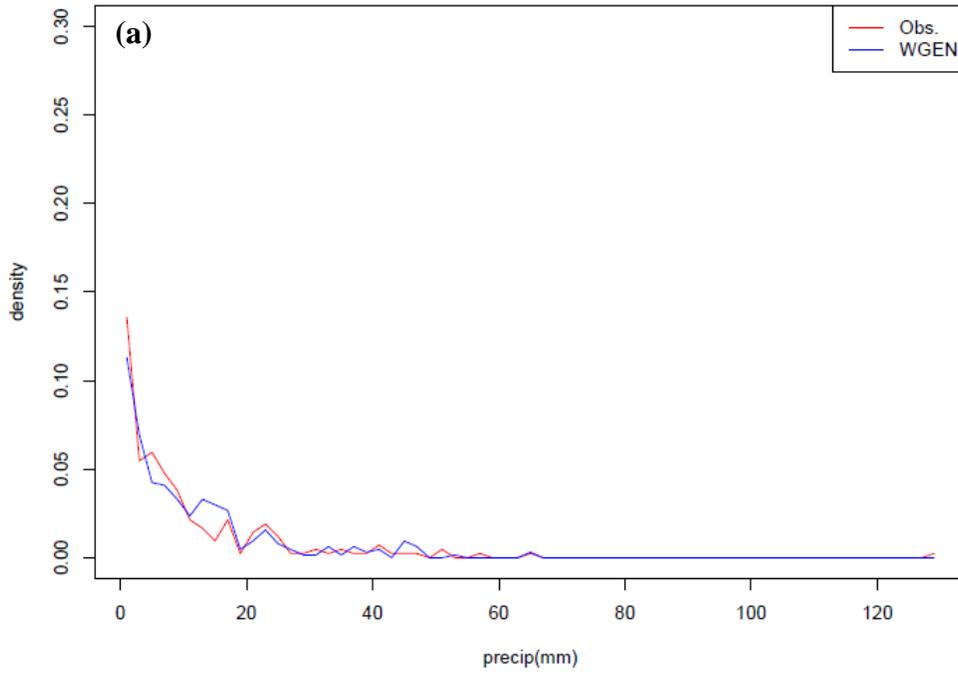


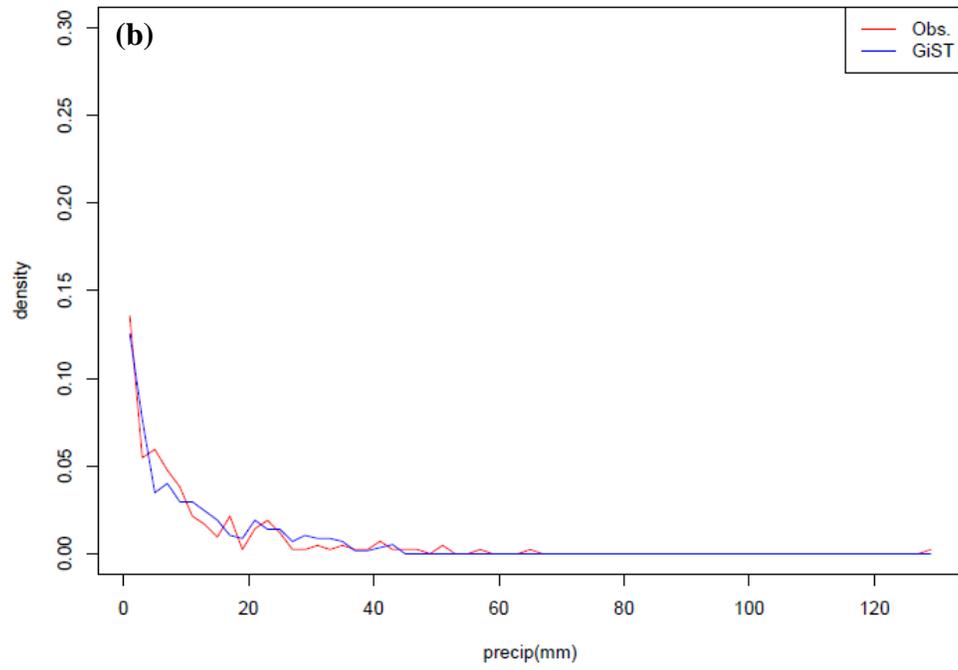
Figure 3.10. RMSE (mm²) comparison for generated values of the inter-annual variability ($Var[S(T)]$) for North Carolina domain.

Figure 3.11. Generated vs. Observed PDF examples for January for station 315838 in Morganton, NC for WGEN (a), GiST (b), and GiSTR (c).

Comparison of Nonzero Rainfall PDF
Station: 315838 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 315838 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 315838 Month: 1

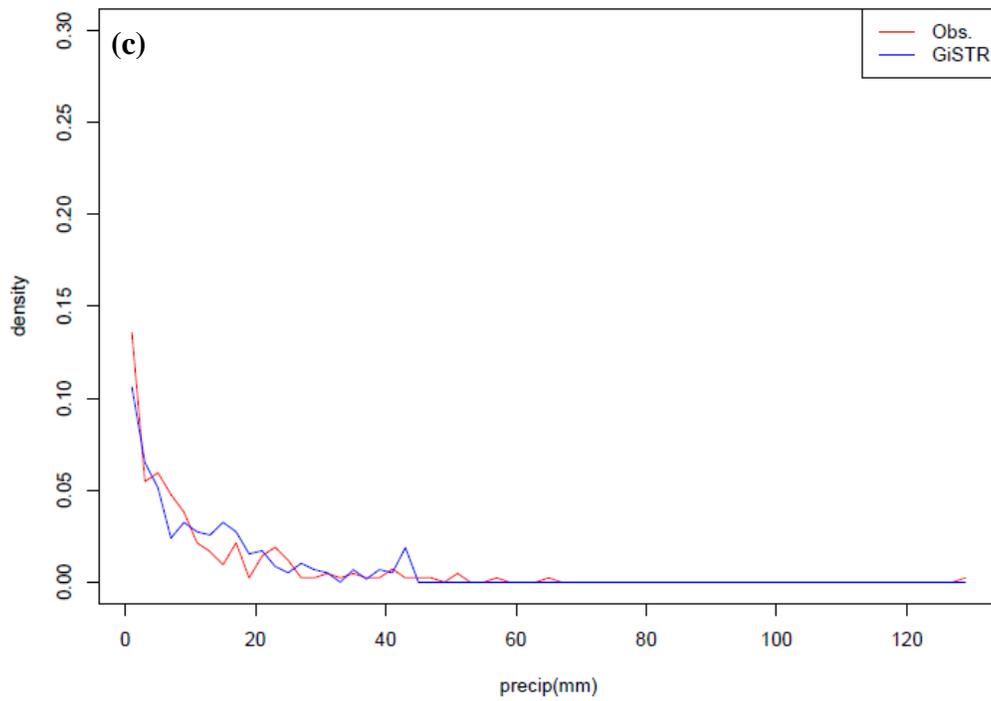
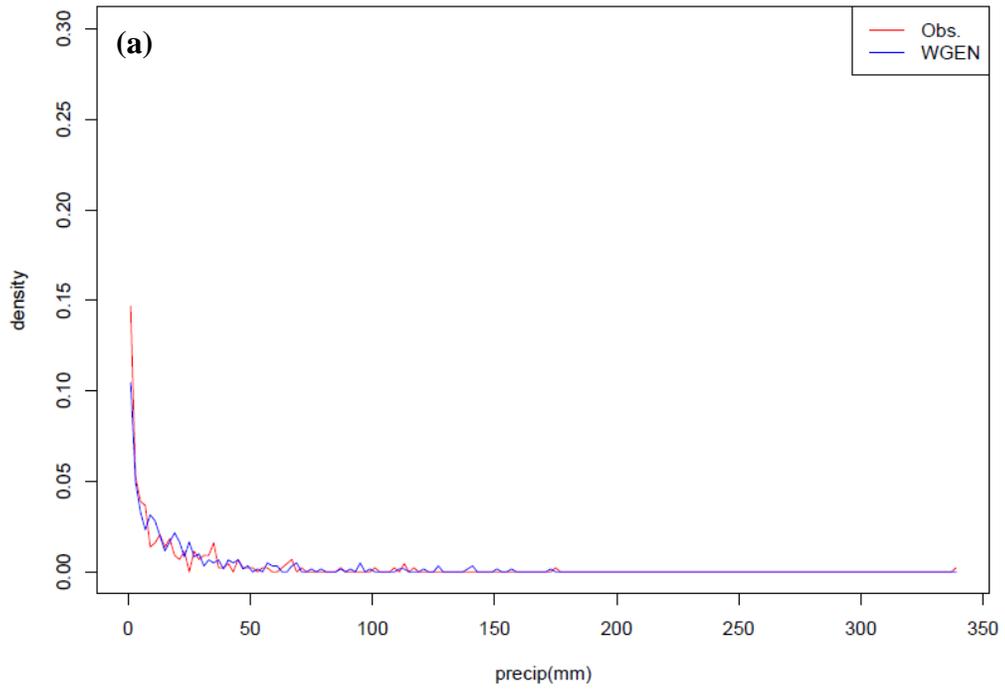
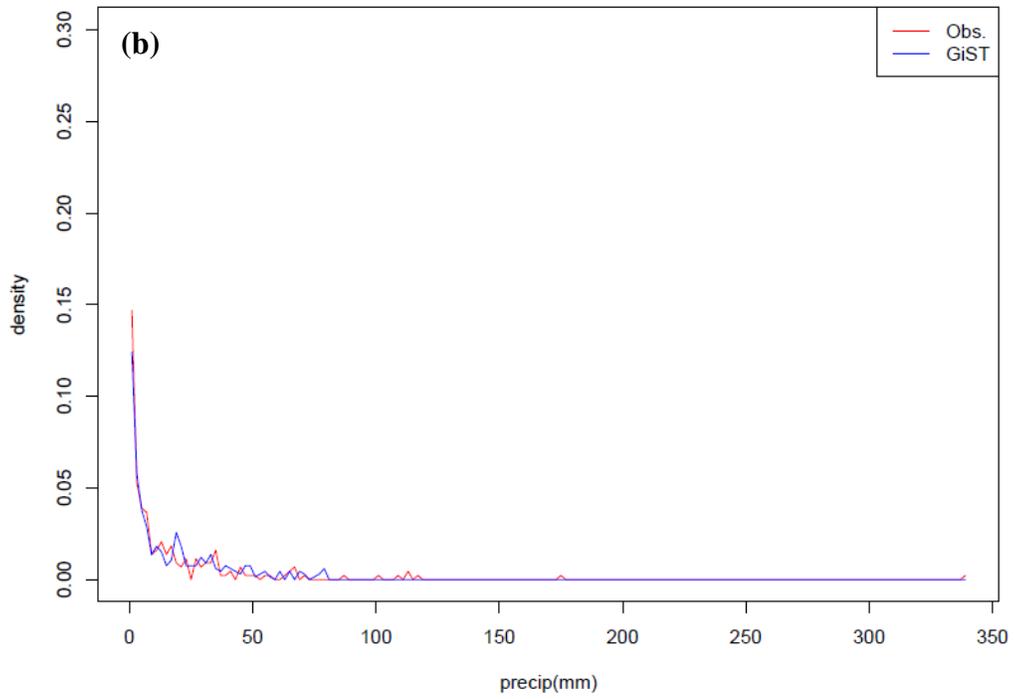


Figure 3.12. Generated vs. Observed PDF examples for September for station 319457 in Wilmington, NC for WGEN (a), GiST (b), and GiSTR (c).

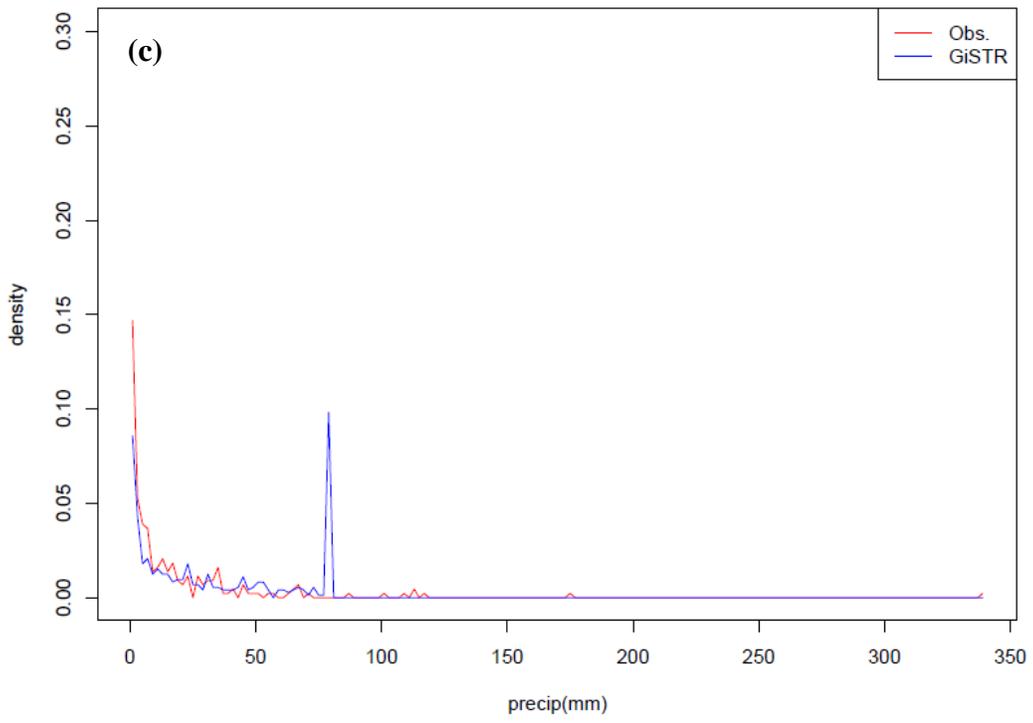
Comparison of Nonzero Rainfall PDF
Station: 319457 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 319457 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 319457 Month: 9



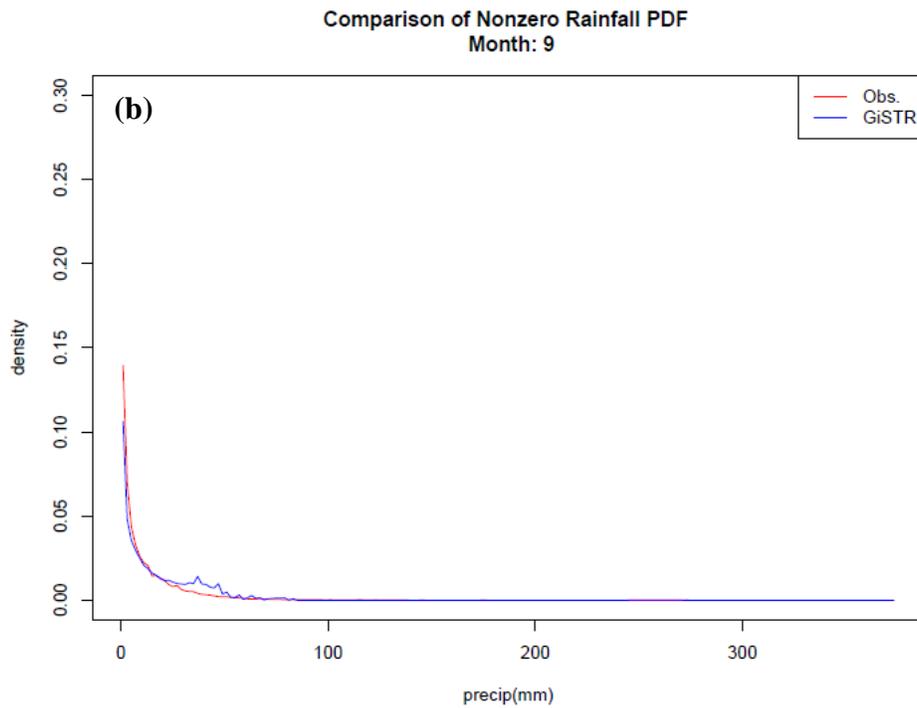
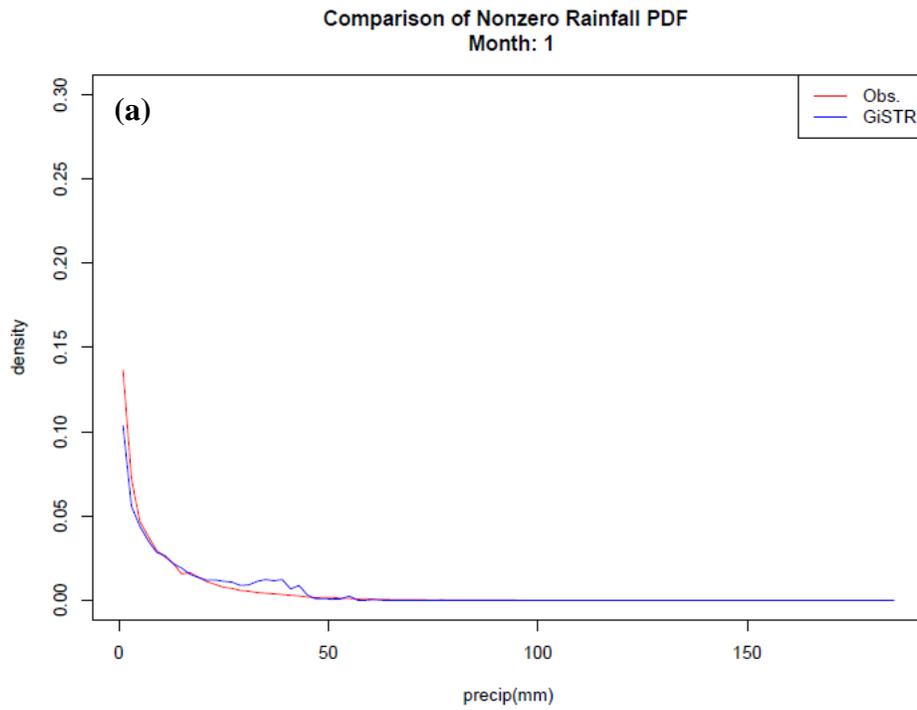


Figure 3.13. Monthly PDF examples, generated vs. observed for GiSTR for January (a) and September (b) for all stations in the domain.

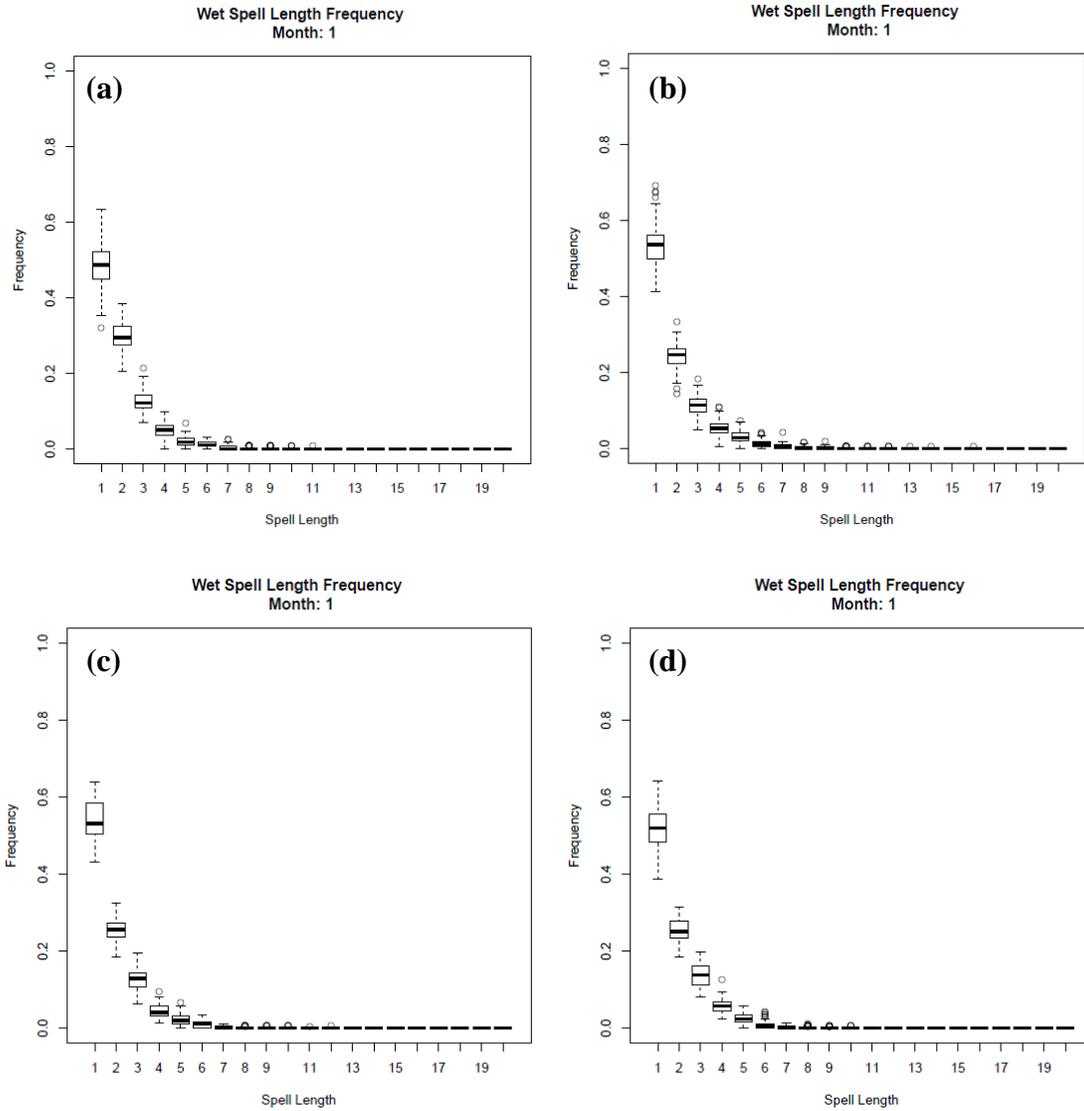


Figure 3.14. Boxplots of the frequency of January Wet Spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for North Carolina domain.

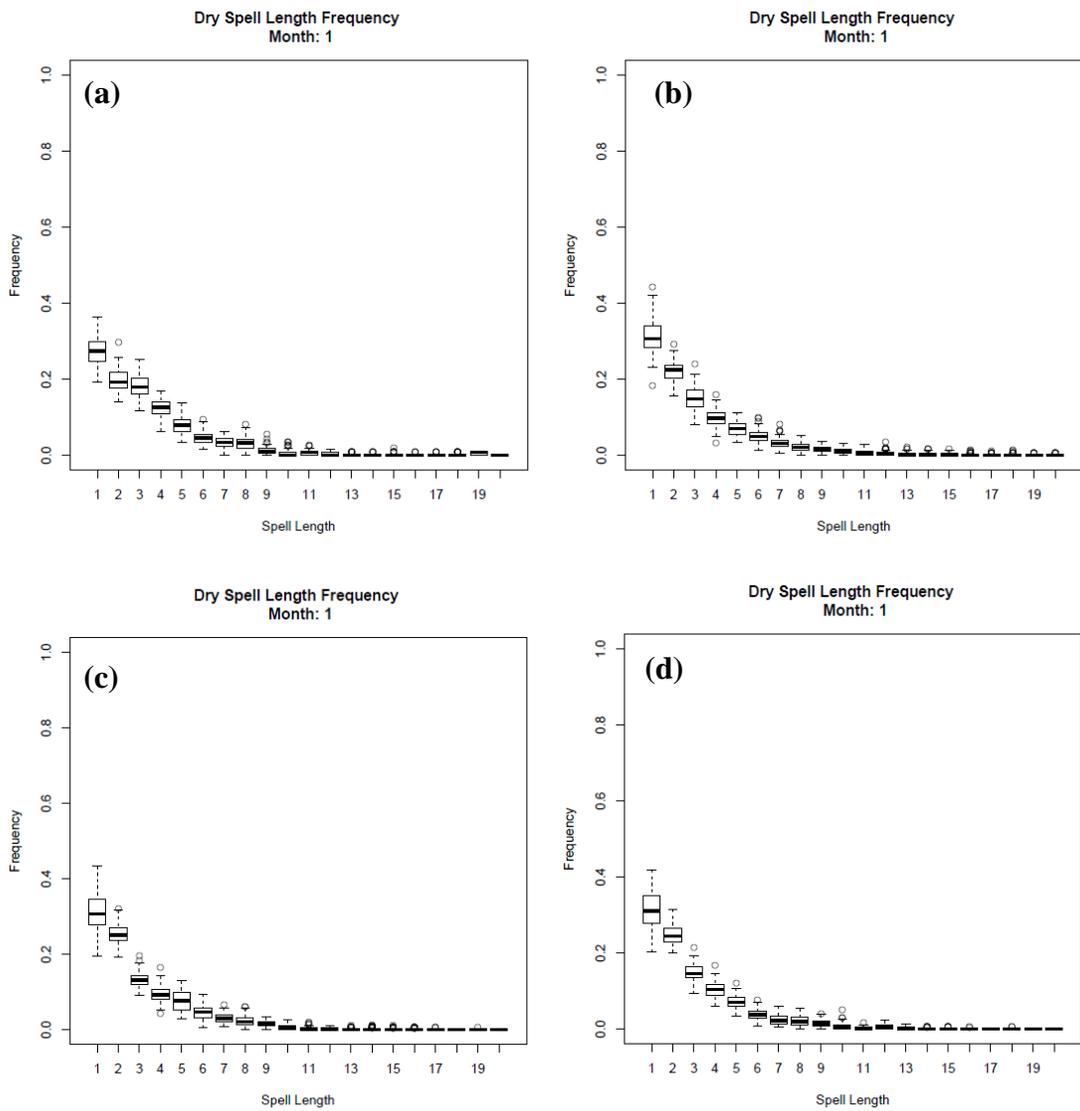


Figure 3.15. Boxplots of the frequency of January dry spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for North Carolina domain.

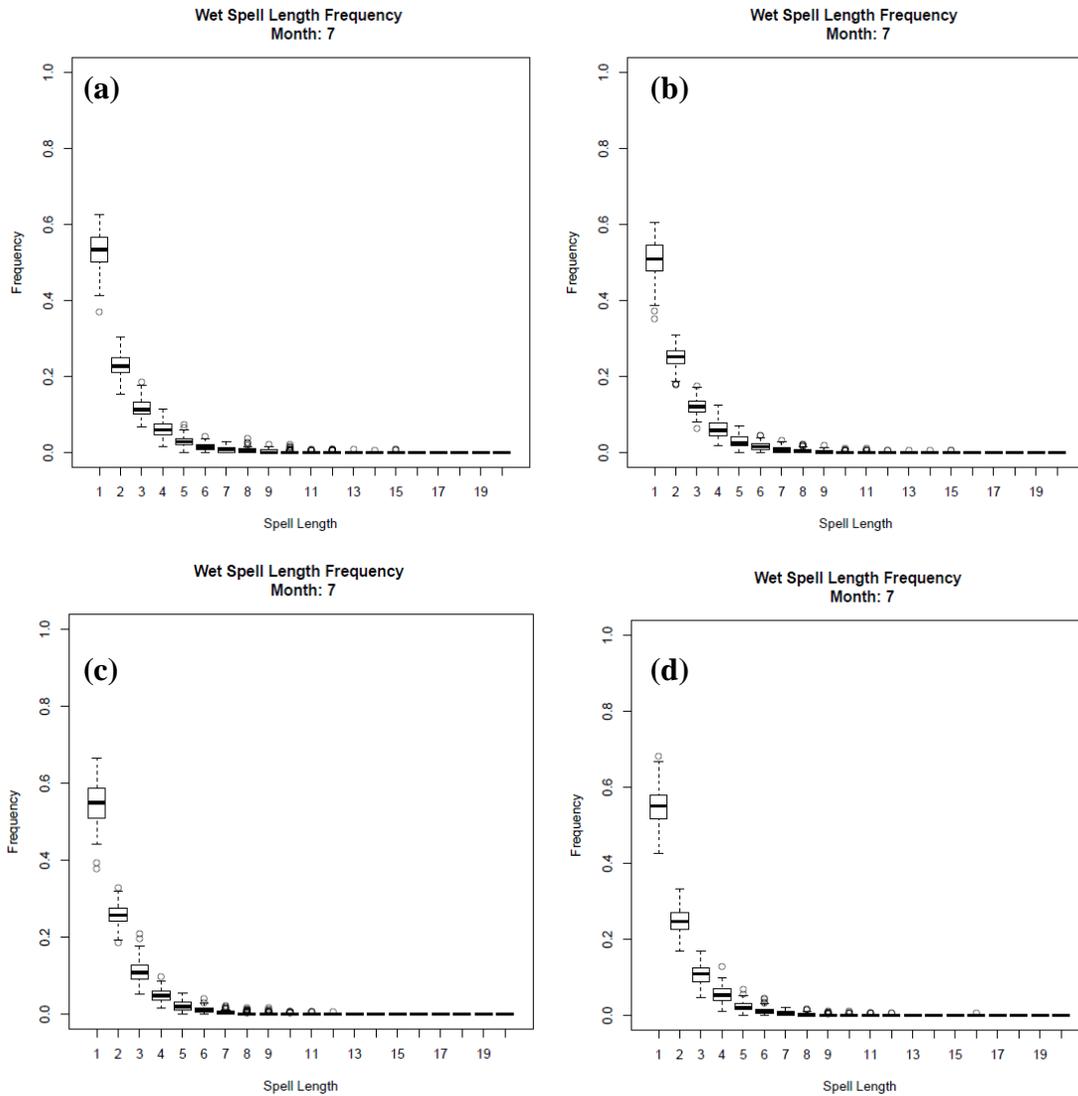


Figure 3.16. Boxplots of the frequency of July Wet Spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for North Carolina domain.

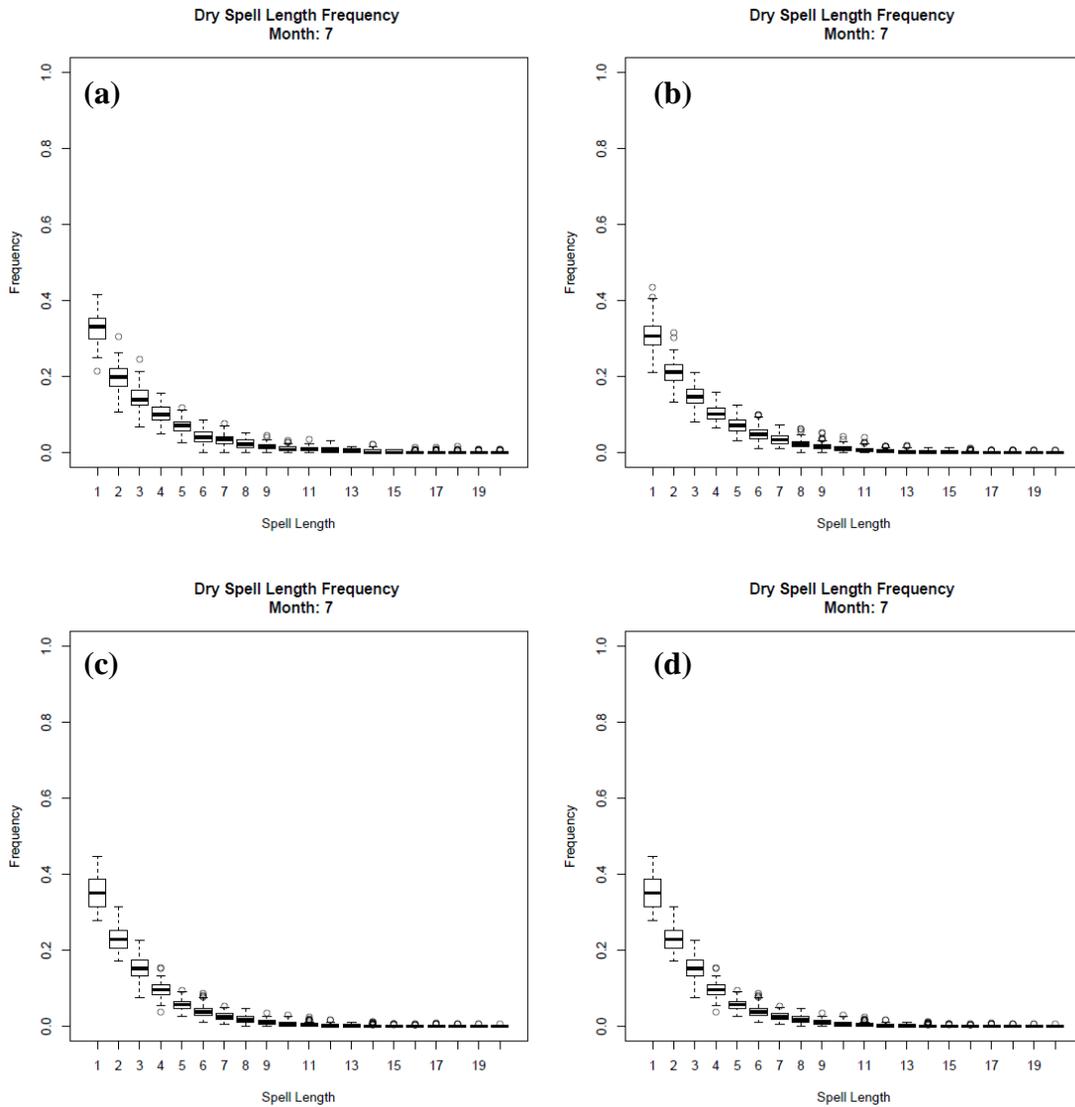


Figure 3.17. Boxplots of the frequency of July dry spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for North Carolina domain.

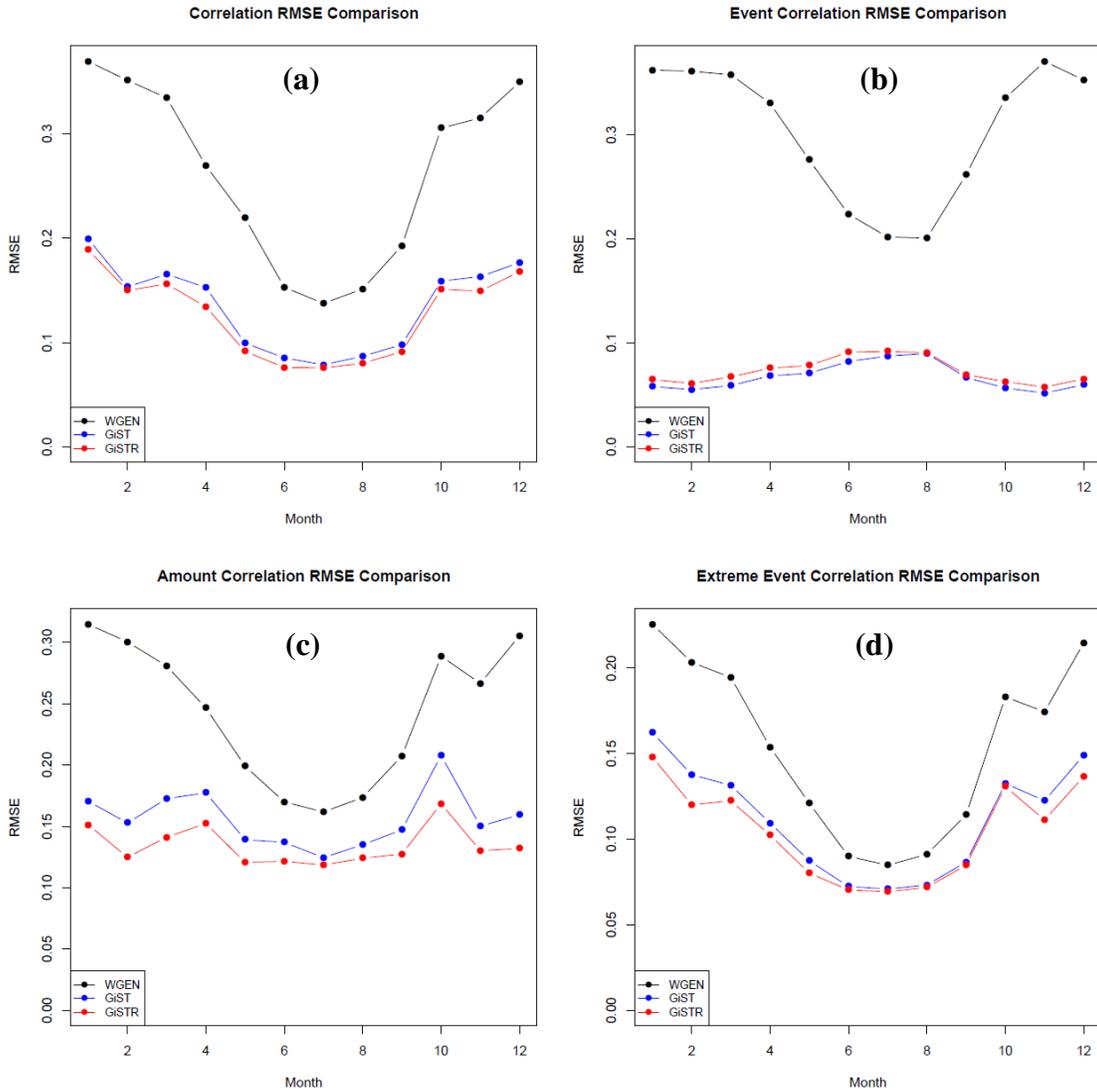


Figure 3.18. RMSE of each weather generator by month for the correlation matrix of precipitation (ρ) (a), the correlation matrix of precipitation events (ρ_{ev})(b), the correlation matrix of precipitation amounts (ρ_{am})(c), and the correlation matrix of precipitation extreme events (ρ_{ex})(d) for the Southeast US domain.

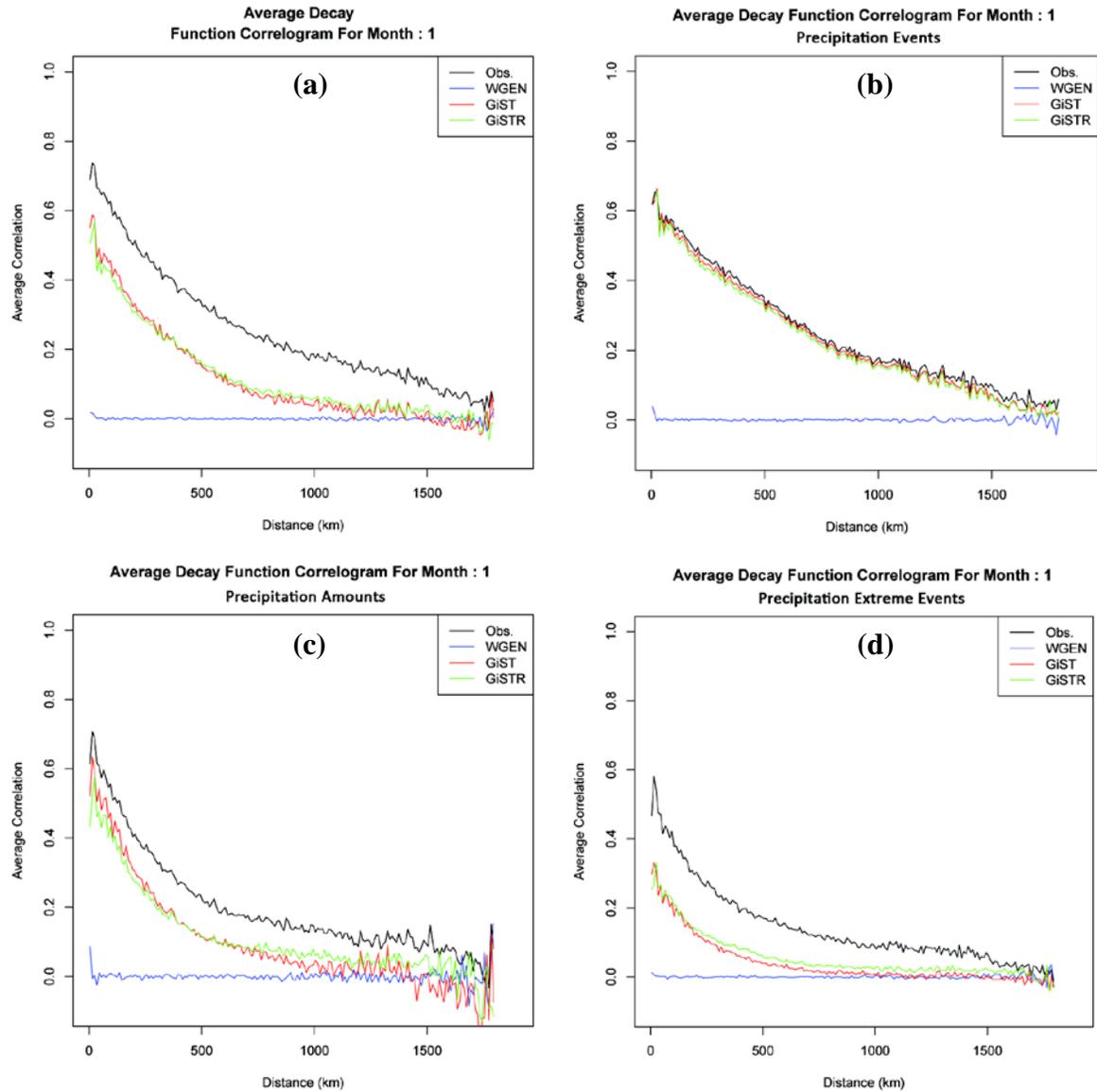


Figure 3.19. Average Decay Function Correlogram for each generator and observations for January for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). Southeast U.S. domain.

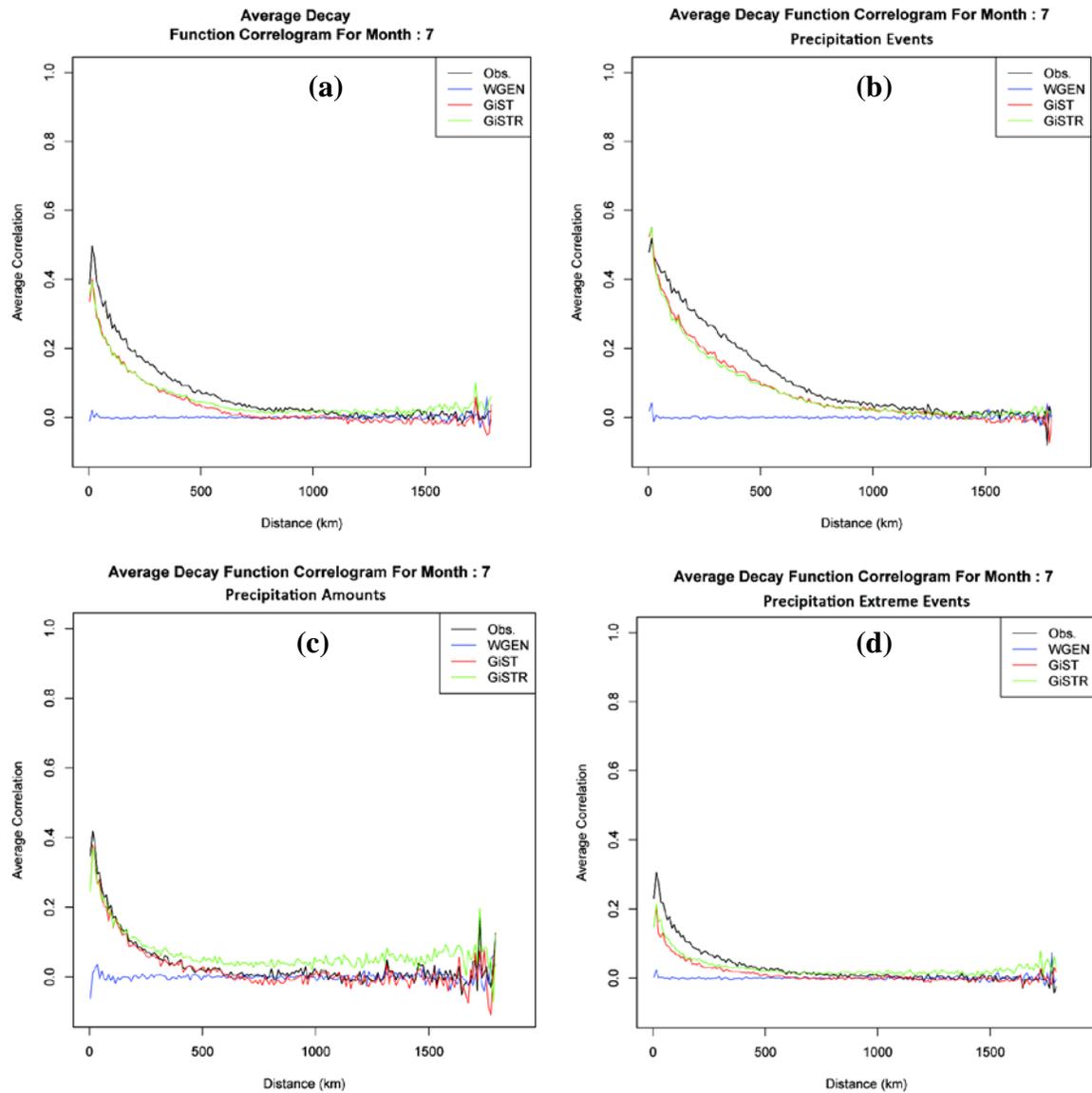


Figure 3.20. Average Decay Function Correlogram for each generator and observations for July for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). Southeast U.S. domain.

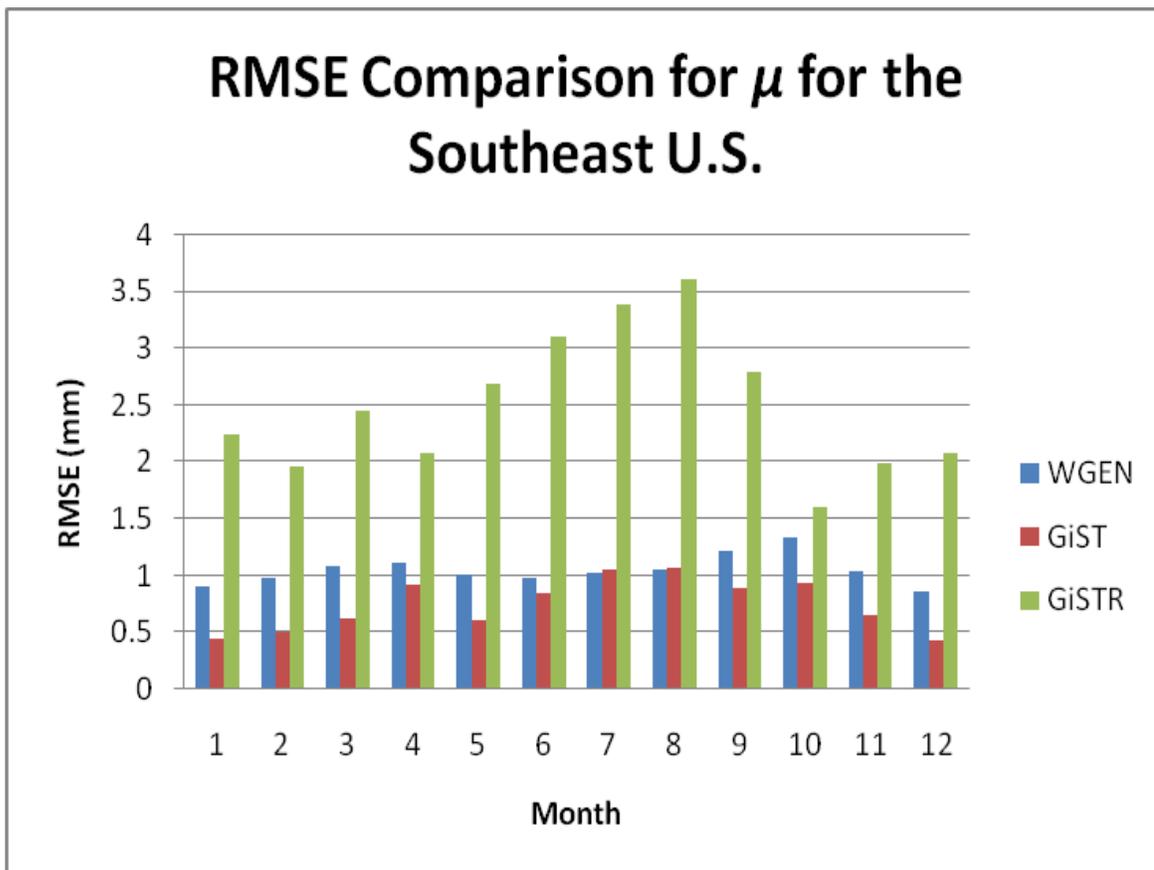
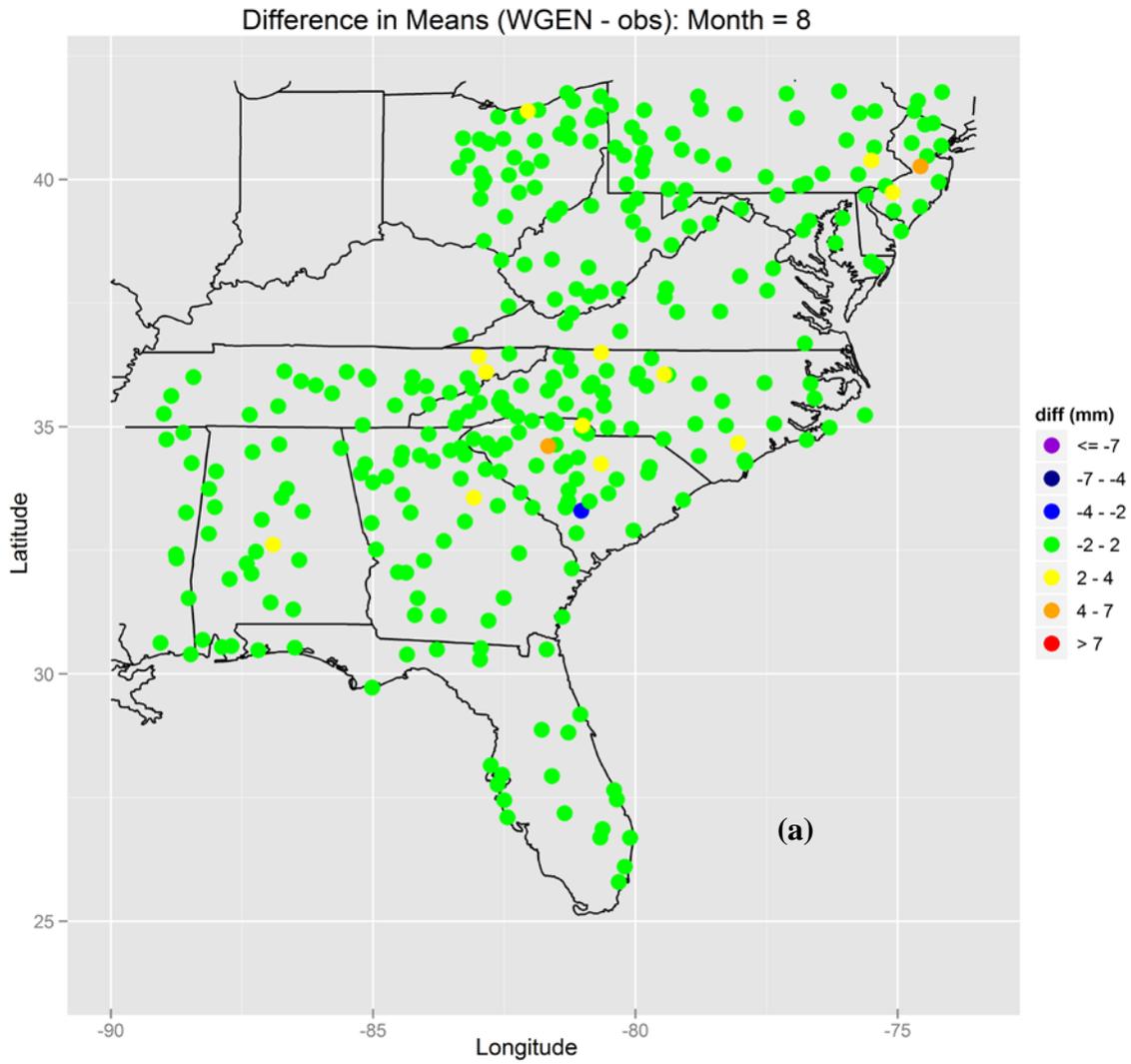
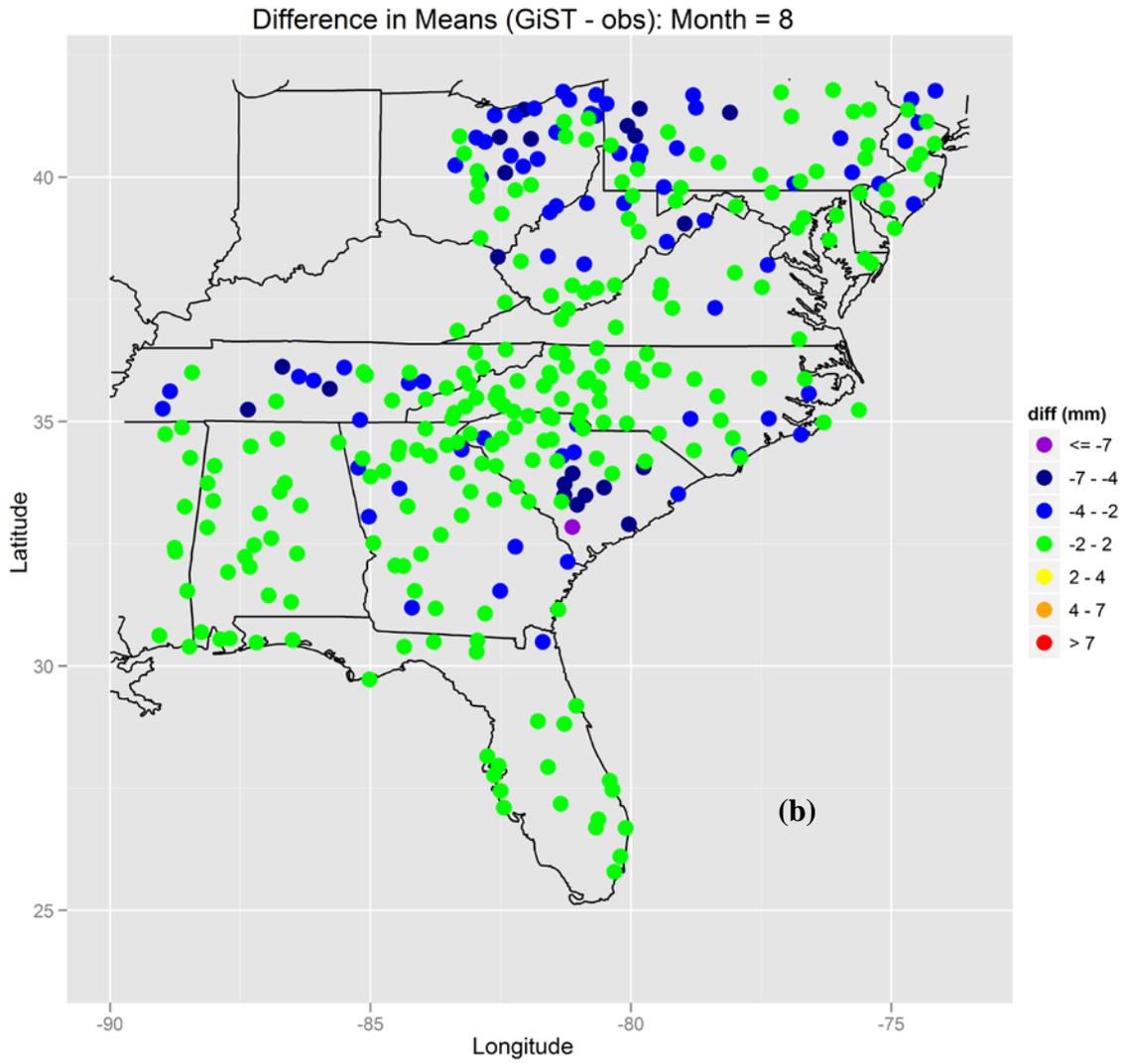
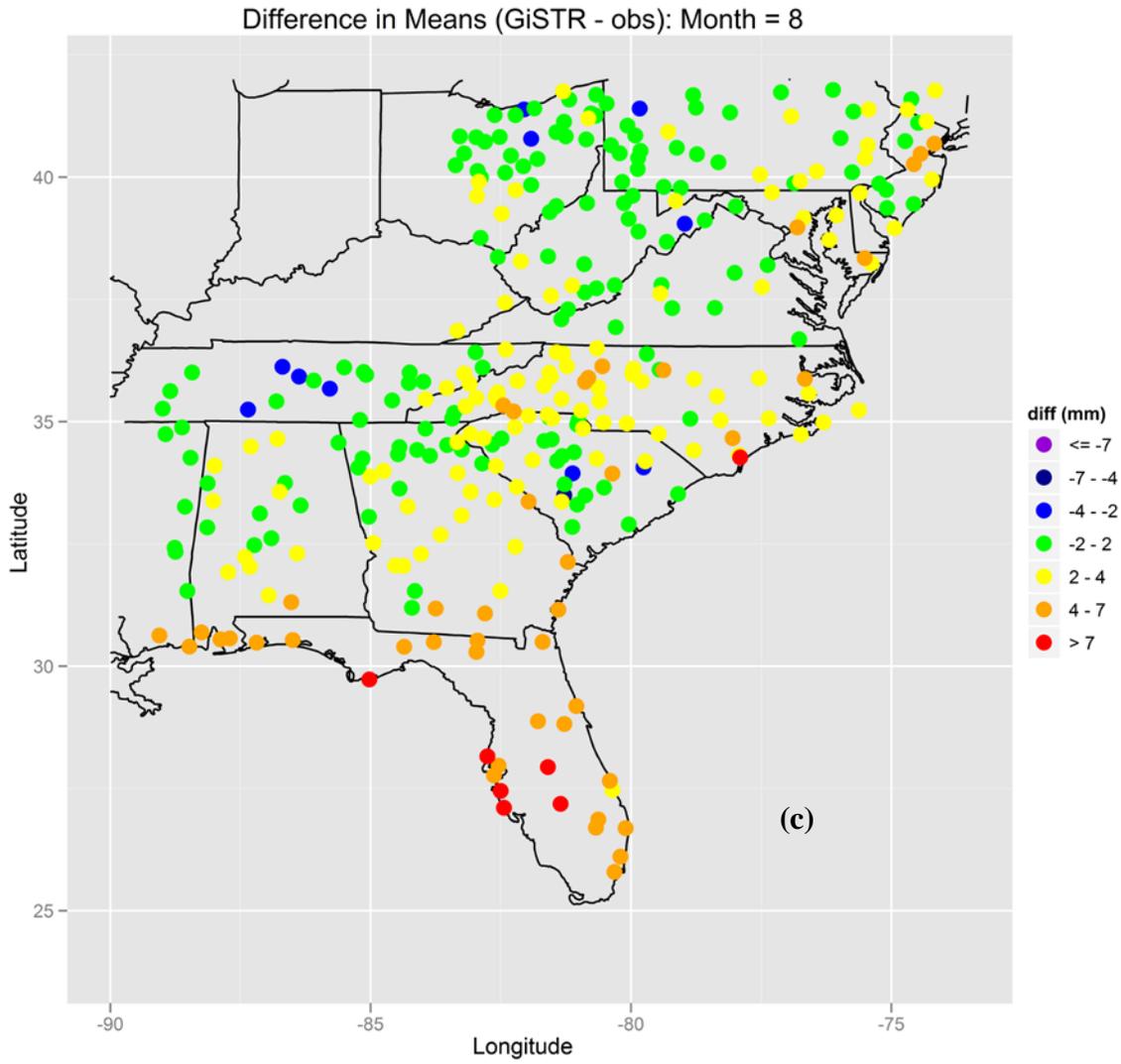


Figure 3.21. RMSE (mm) comparison for the generated values of mean daily nonzero precipitation (μ) for the Southeast US domain.

Figure 3.22. Difference between generated and observed values of mean daily nonzero precipitation (μ) (mm) across the Southeast U.S. domain for the month of August for WGEN(a), GiST(b), and GiSTR (c).







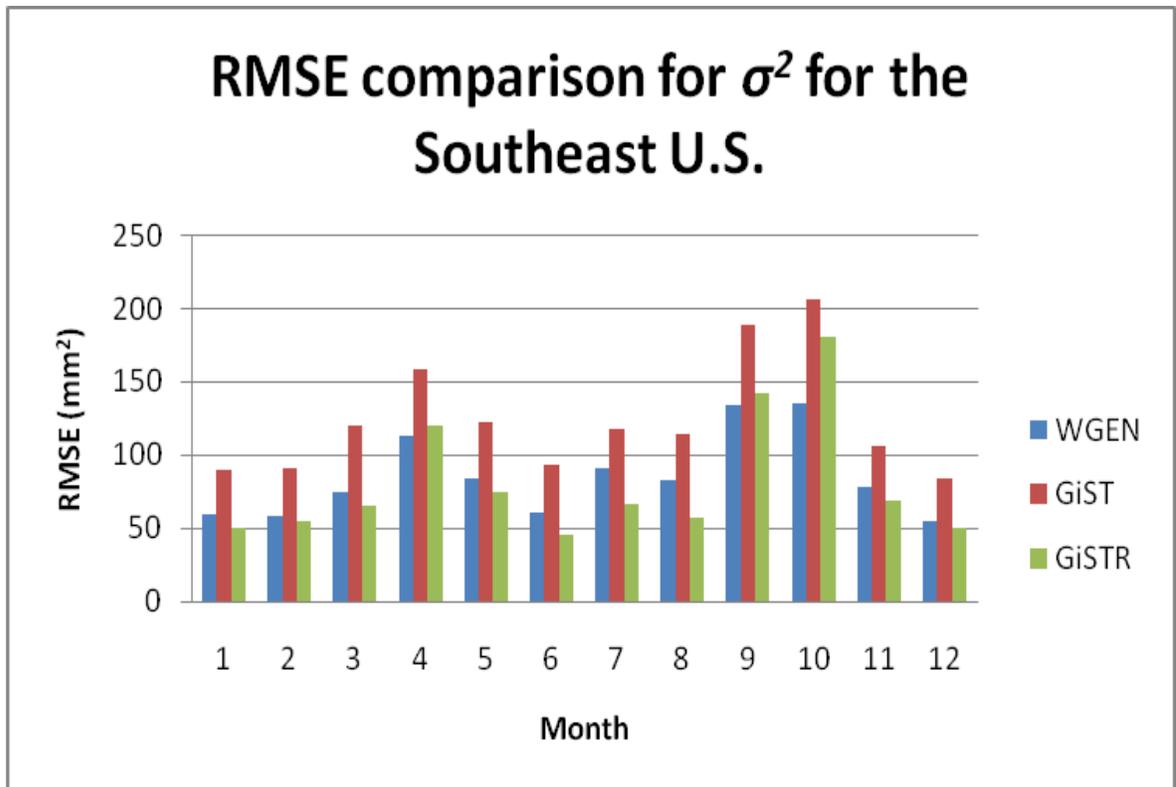
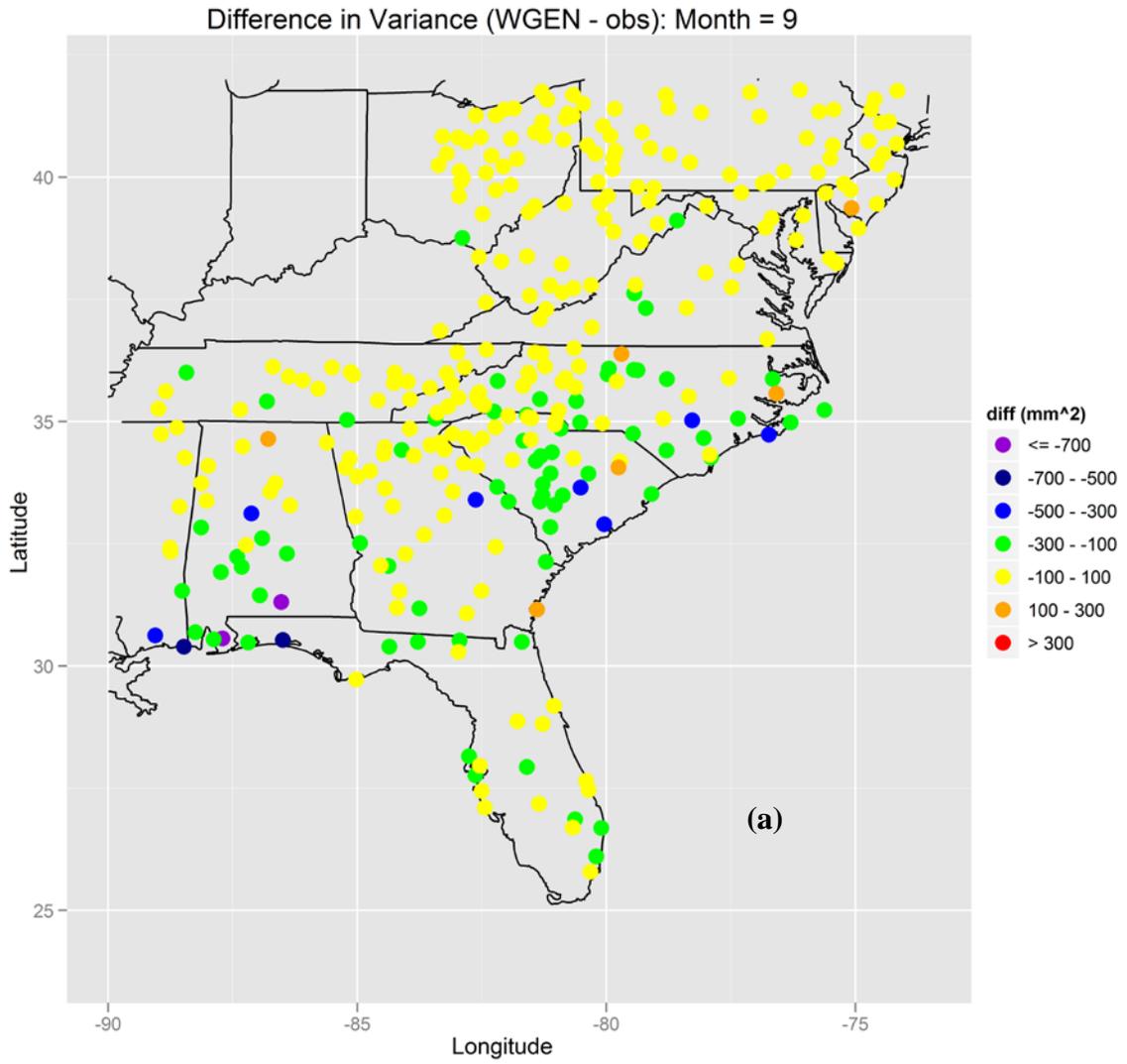
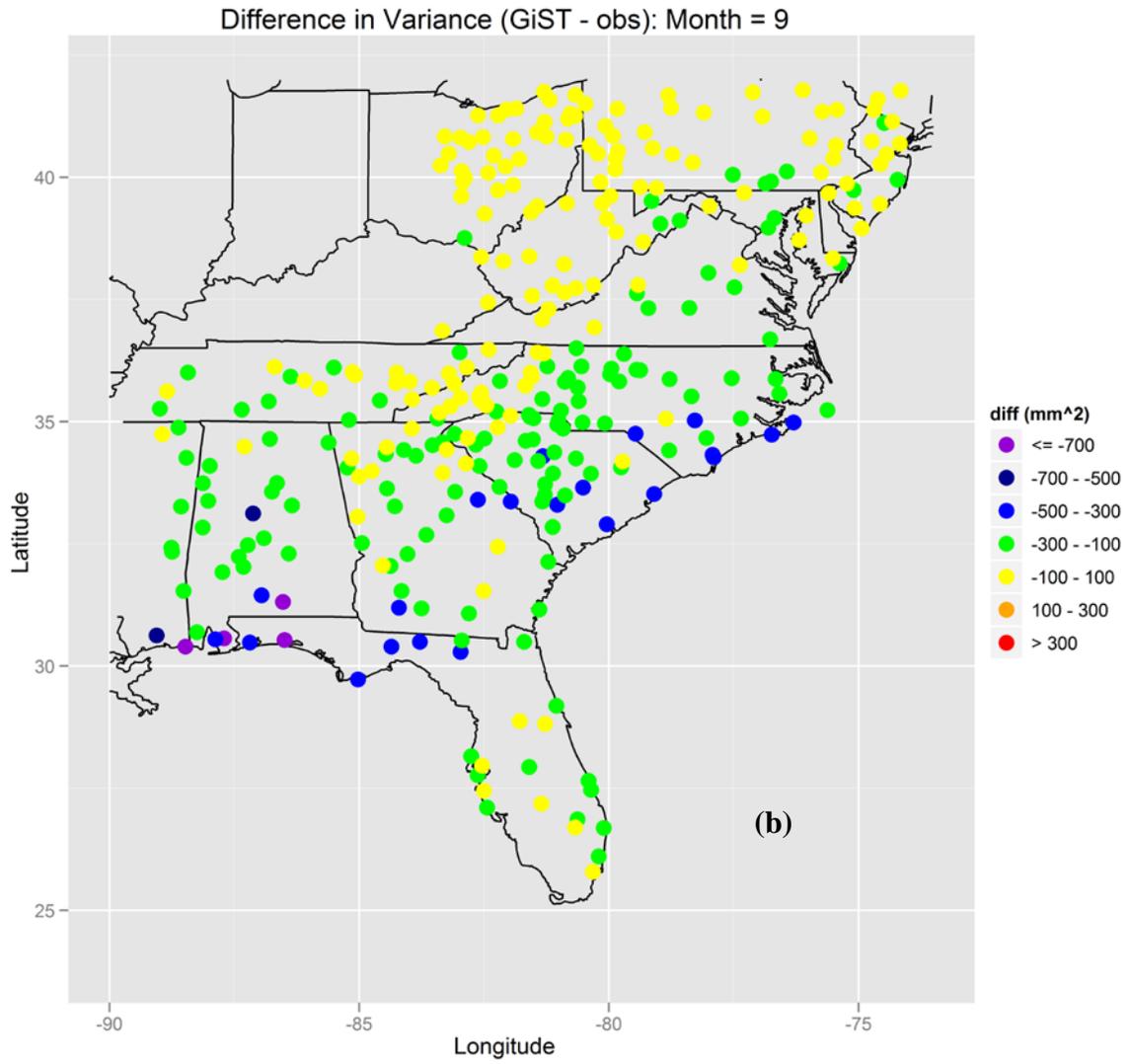
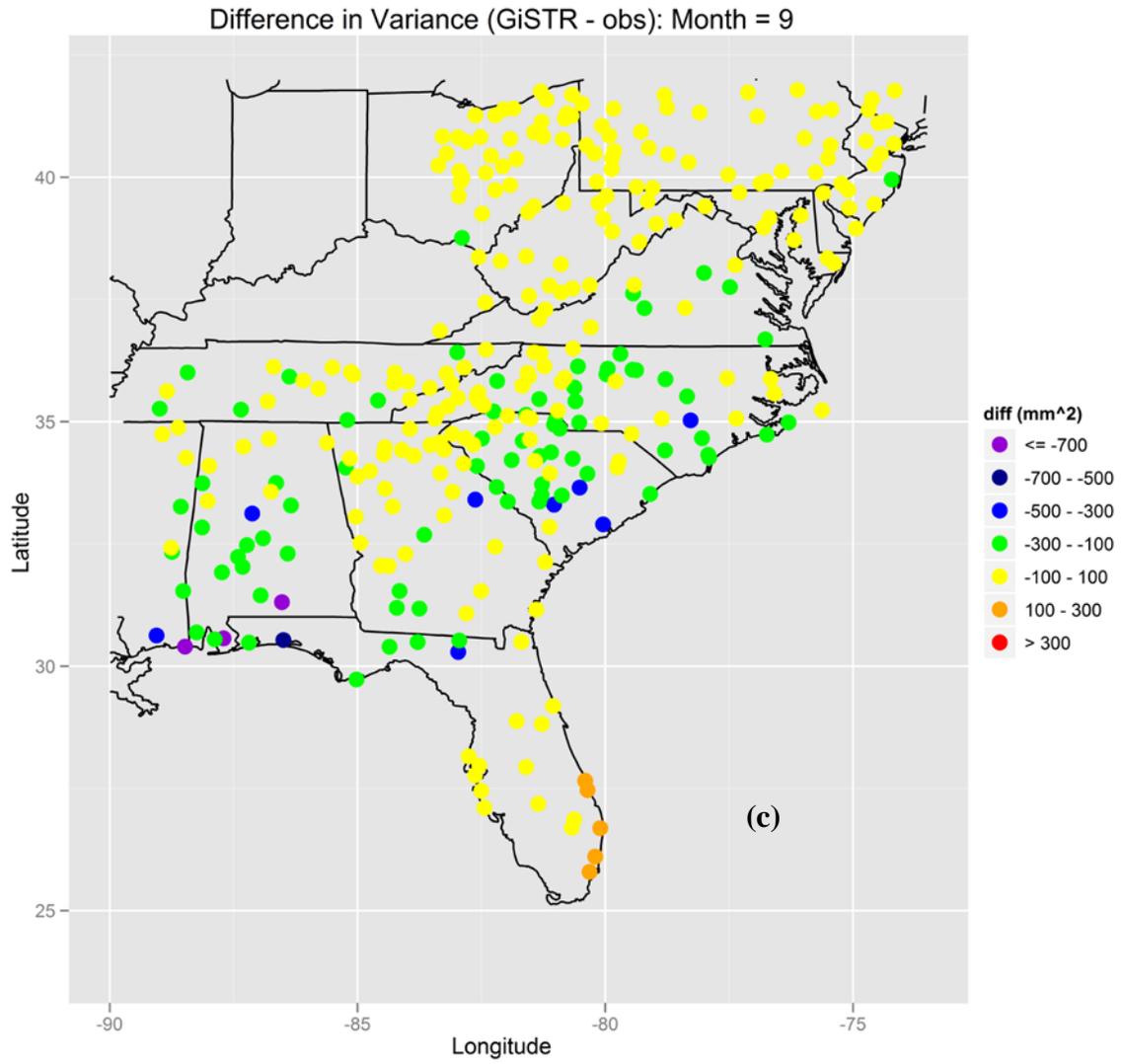


Figure 3.23. RMSE (mm²) comparison for the generated values of the variance of daily nonzero precipitation (σ^2) for the Southeast US domain.

Figure 3.24. Difference between generated and observed values of the variance of daily nonzero precipitation (σ^2)(mm²) across the Southeast U.S. domain for the month of August for WGEN (a), GiST (b), and GiSTR (c).







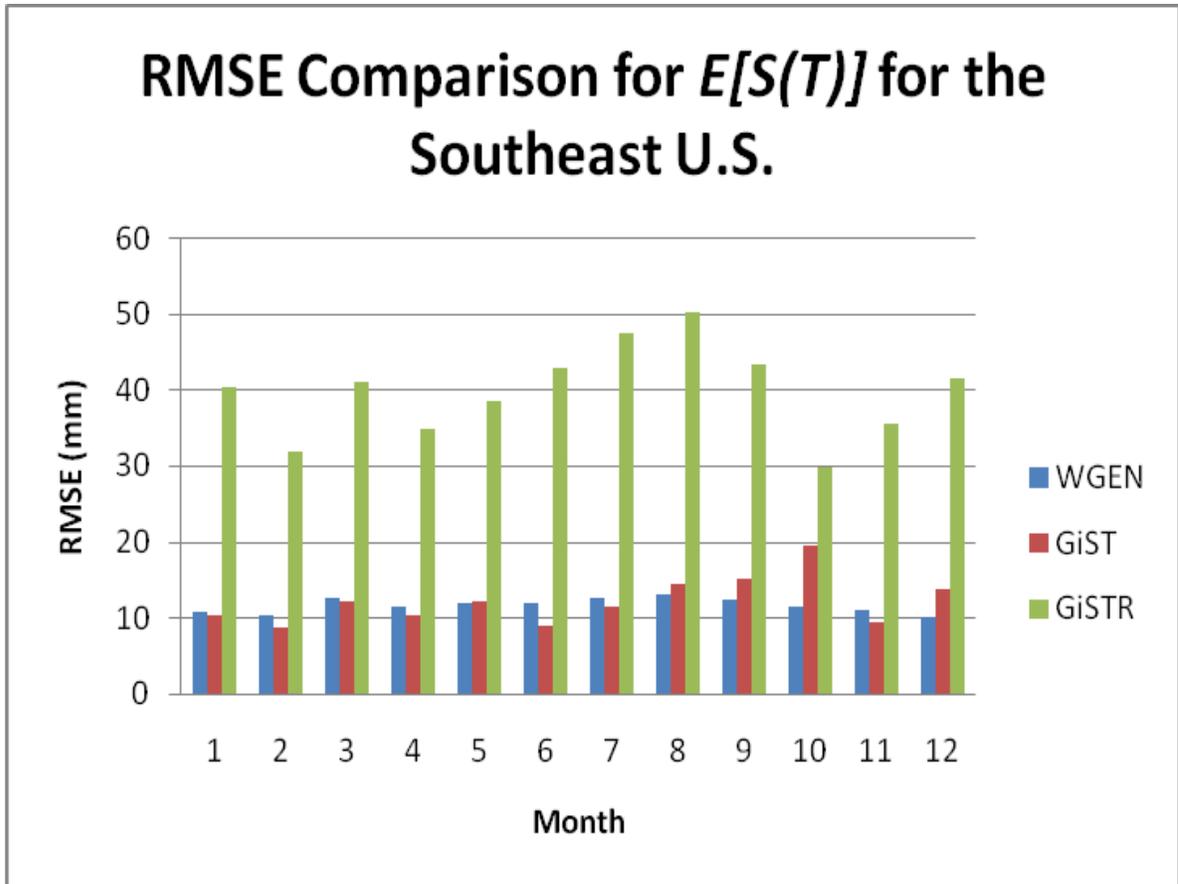


Figure 3.25. RMSE (mm) comparison for the generated values of the average total precipitation($E[S(T)]$) for the Southeast U.S. domain.

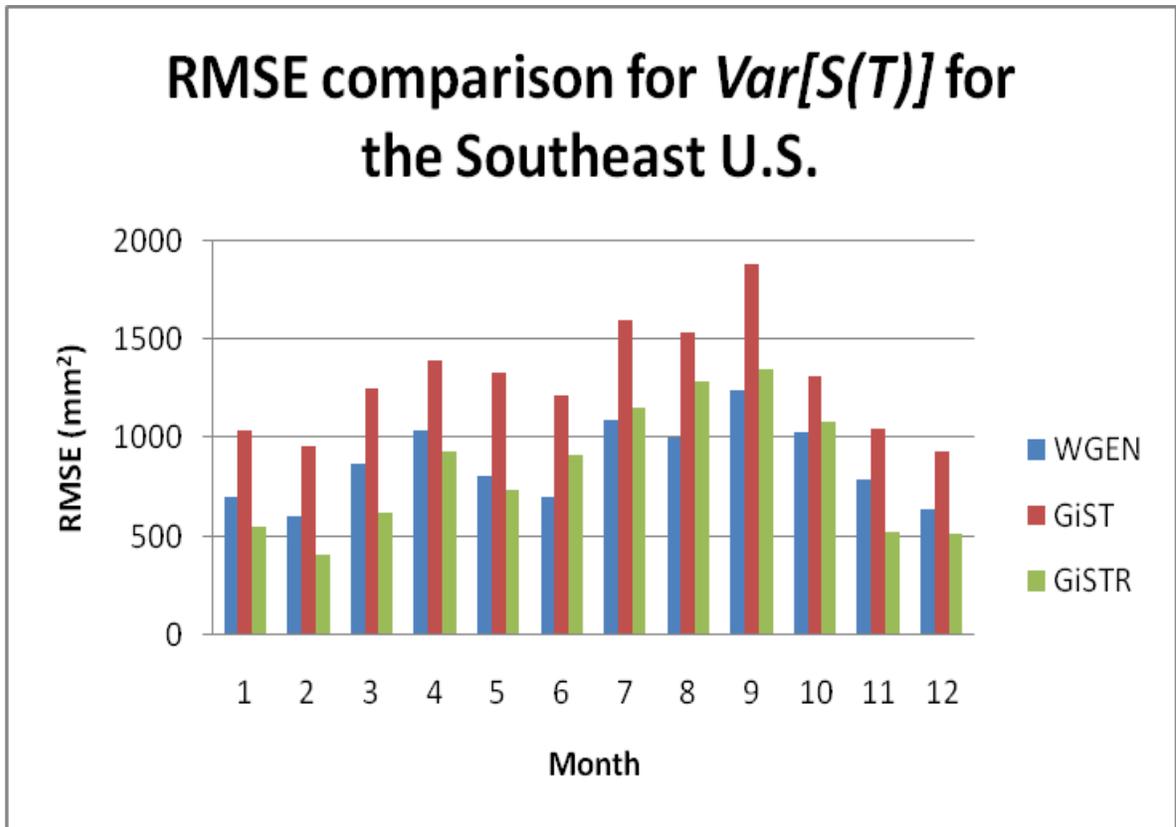
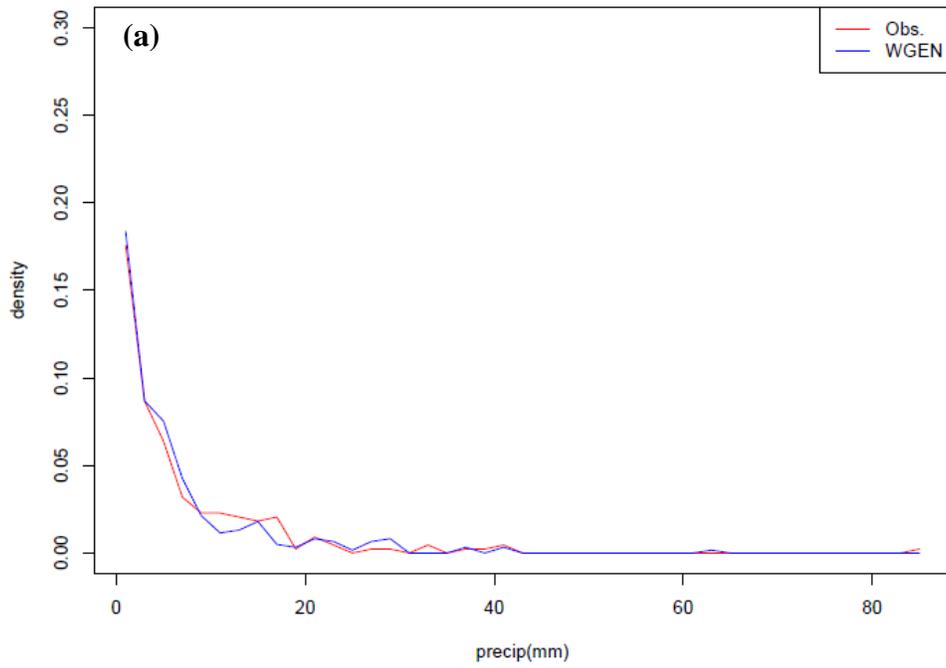


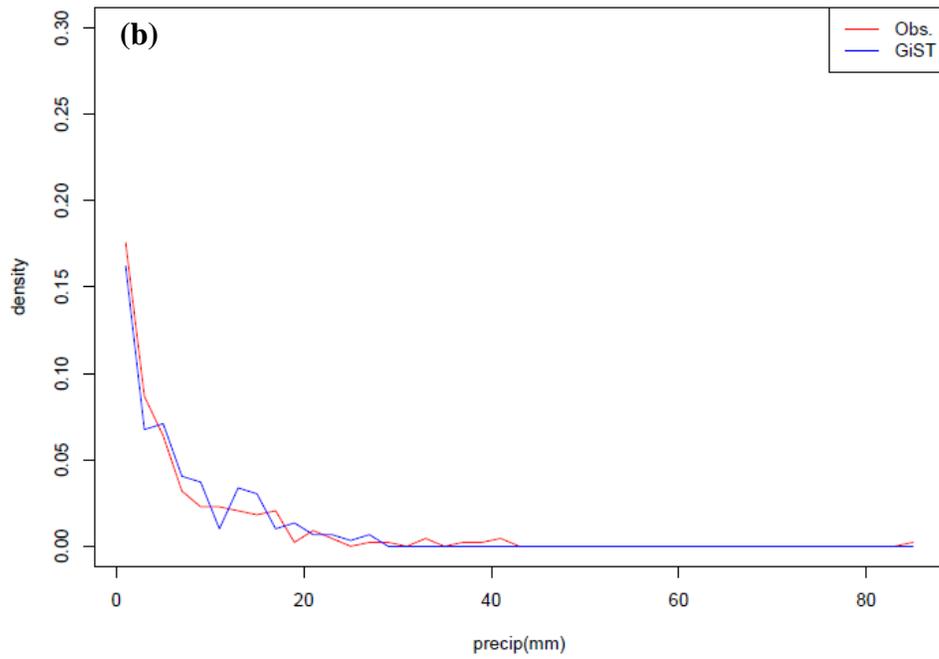
Figure 3.26. RMSE (mm^2) comparison for the generated values of the inter-annual variability ($Var[S(T)]$) for the Southeast US domain.

Figure 3.27. Generated vs. Observed PDF examples for January for station 310301 in Asheville, NC for each weather generator for WGEN (a), GiST (b), and GiSTR (c).

Comparison of Nonzero Rainfall PDF
Station: 310301 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 310301 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 310301 Month: 1

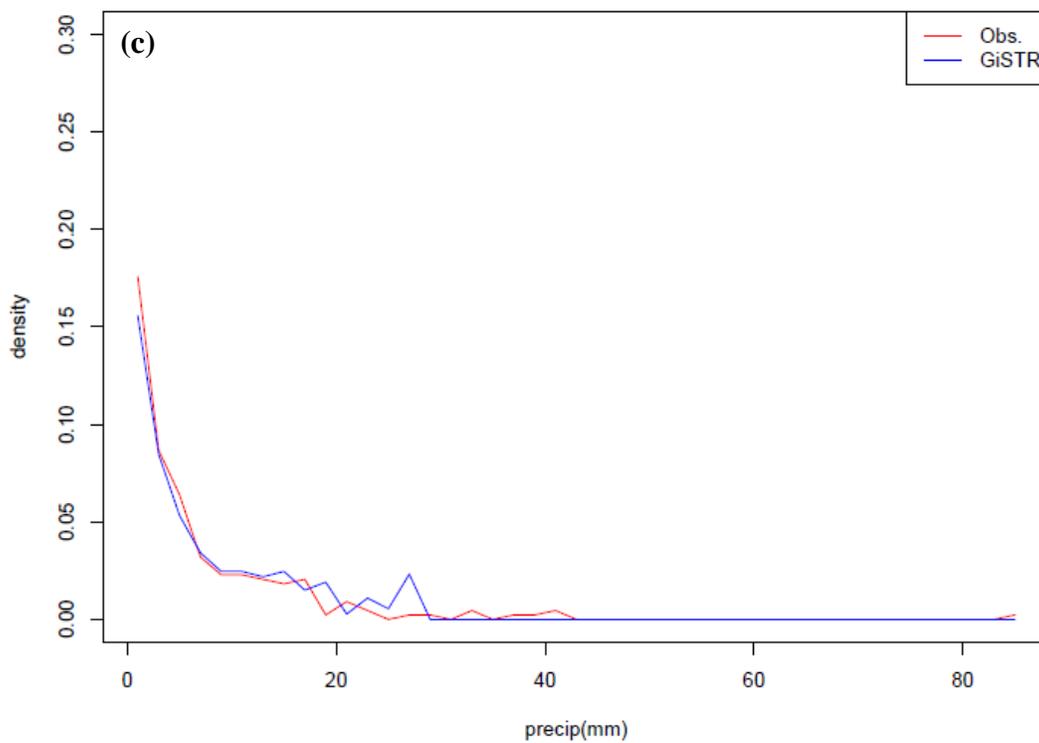
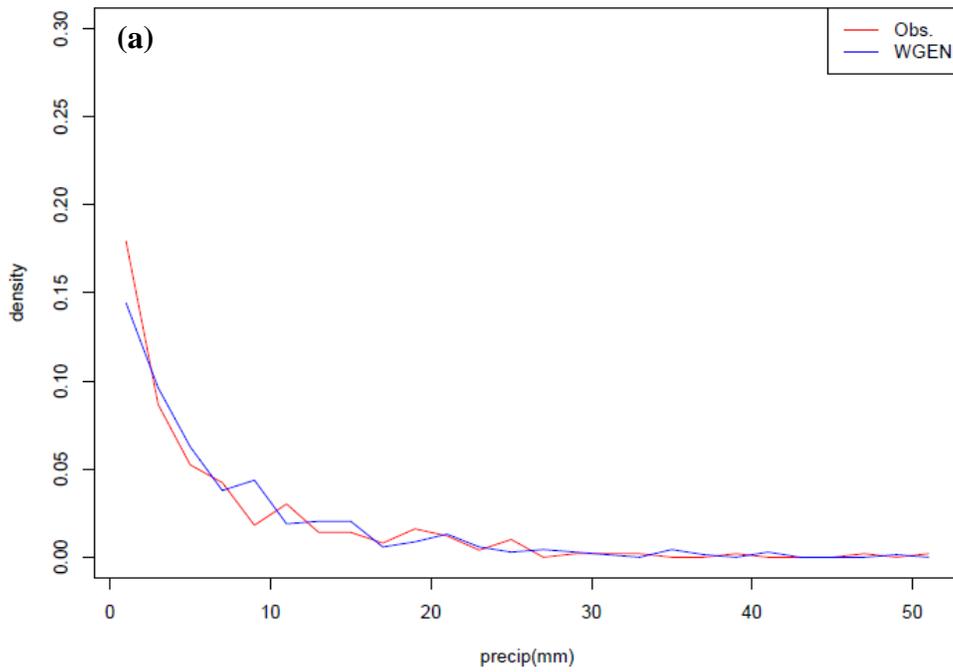
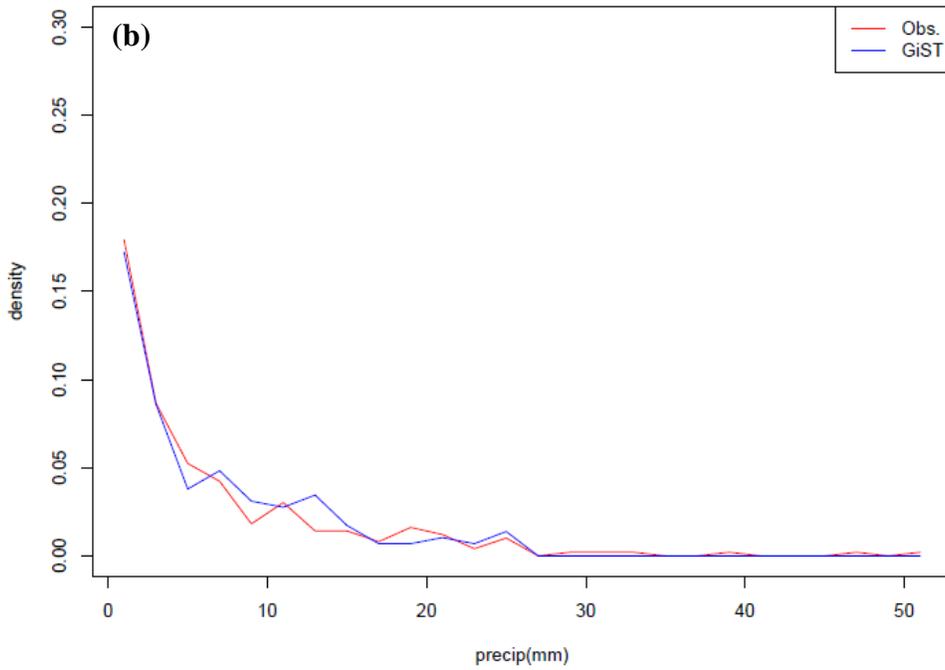


Figure 3.28. Generated vs. Observed PDF examples for January for station 460102 in Alderson, WV for each weather generator for WGEN (a), GiST (b), and GiSTR (c).

Comparison of Nonzero Rainfall PDF
Station: 460102 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 460102 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 460102 Month: 1

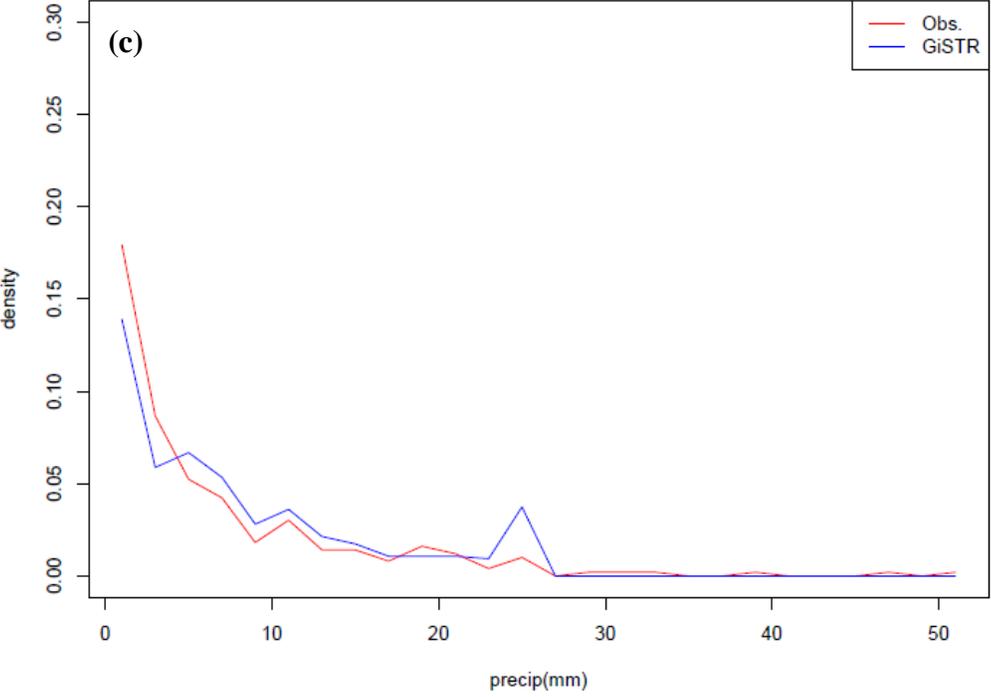
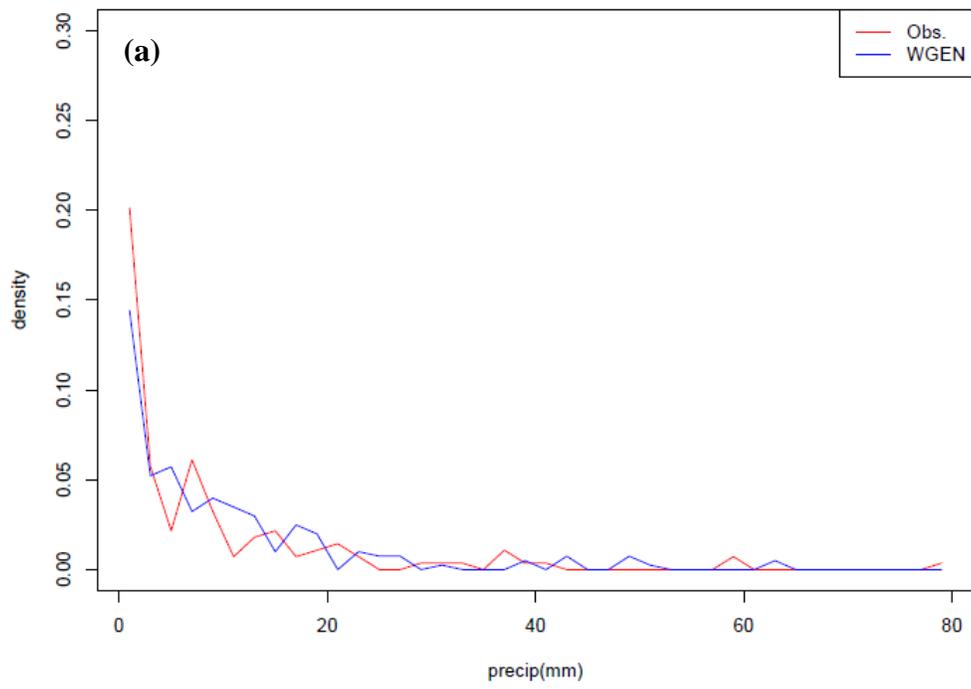
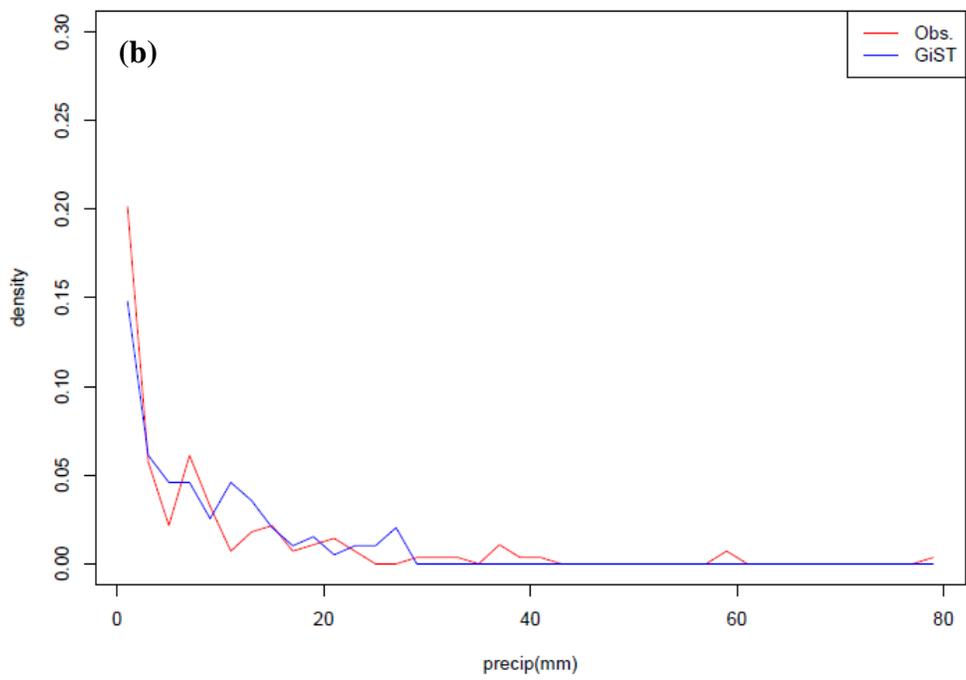


Figure 3.29. Generated vs. Observed PDF examples for January for station 089176 in Venice, FL for each weather generator for WGEN (a), GiST (b), and GiSTR (c).

Comparison of Nonzero Rainfall PDF
Station: 089176 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 089176 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 089176 Month: 1

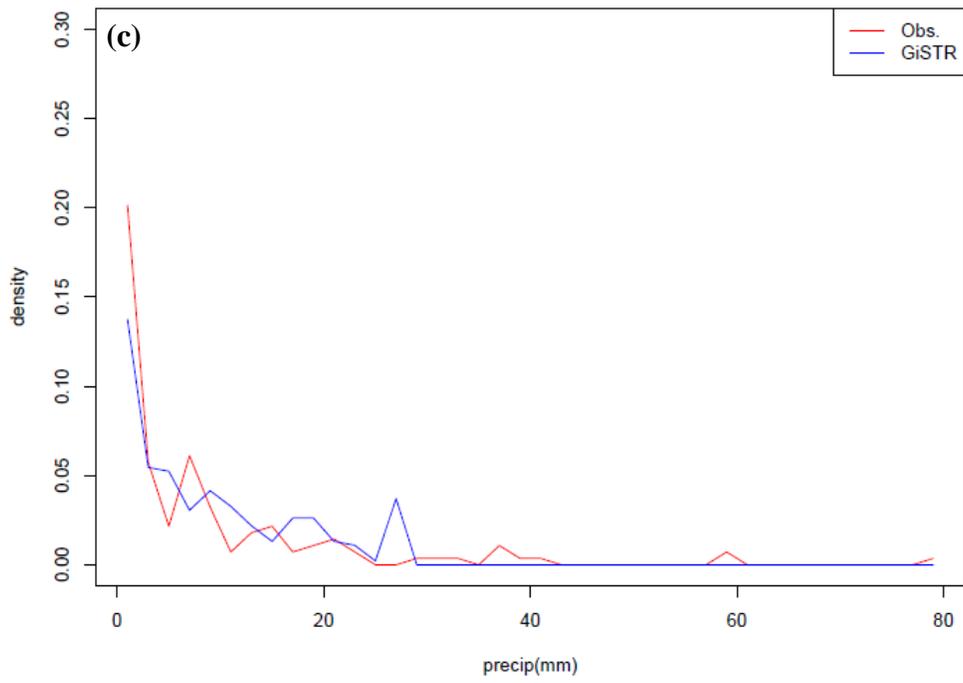
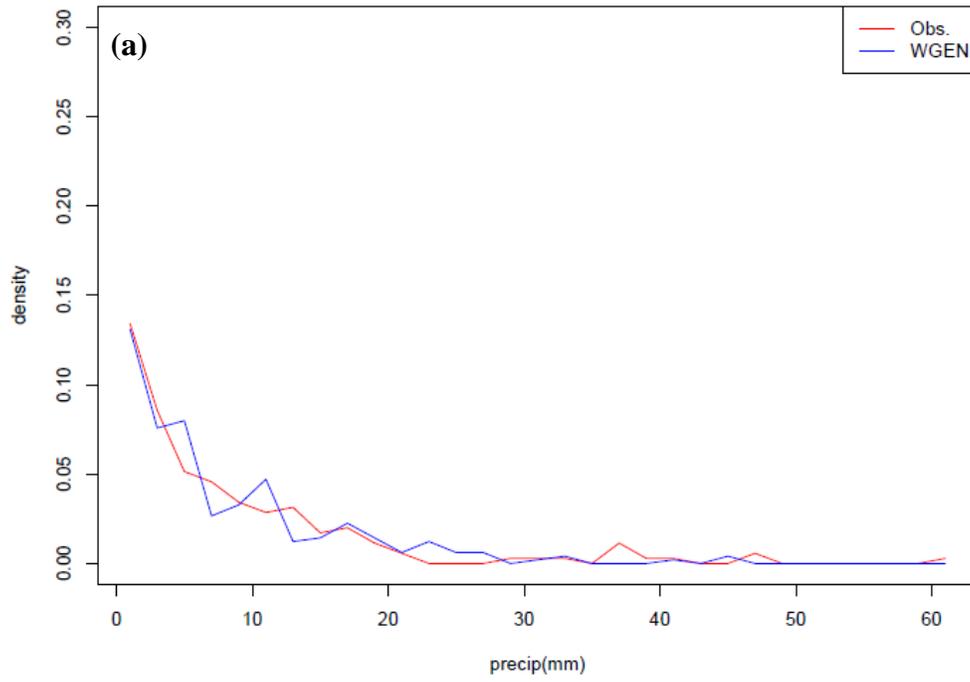
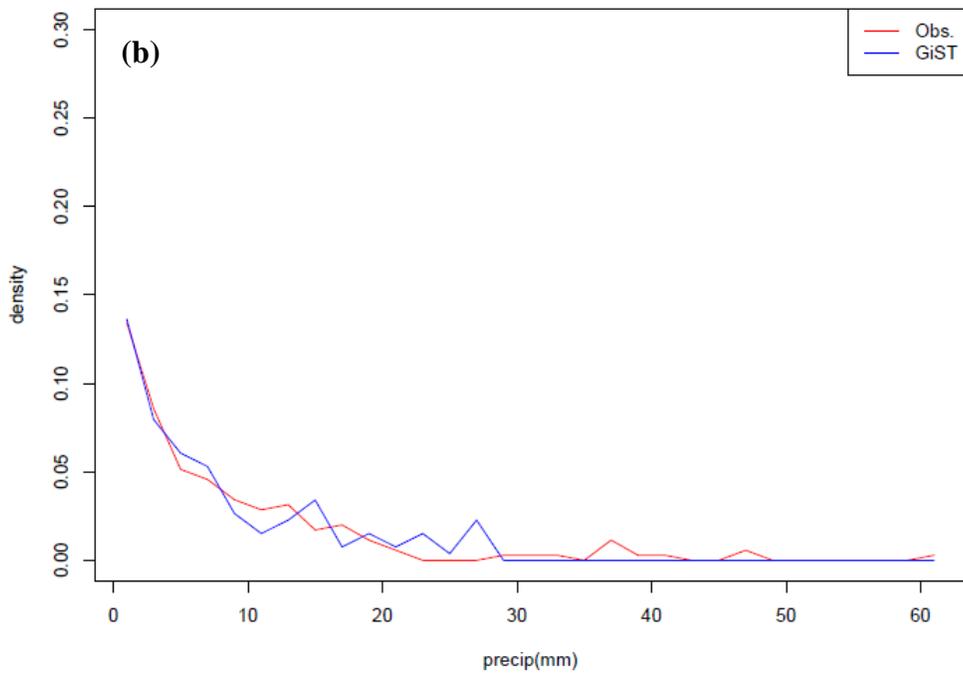


Figure 3.30. Generated vs. Observed PDF examples for September for station 310301 in Asheville, NC for each weather generator for WGEN (a), GiST (b), and GiSTR (c).

Comparison of Nonzero Rainfall PDF
Station: 310301 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 310301 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 310301 Month: 9

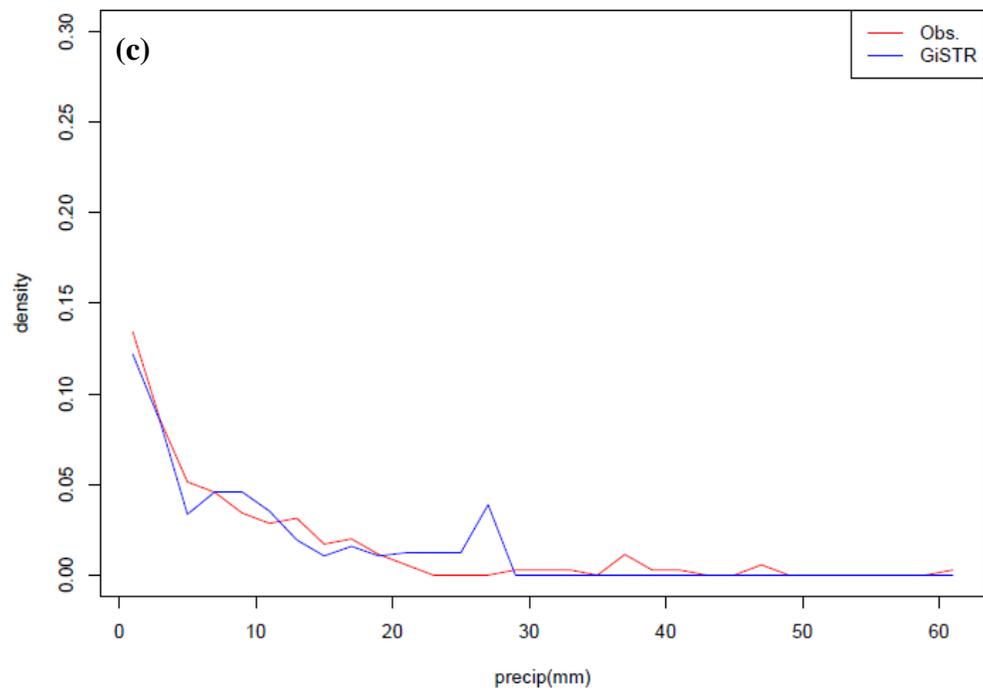
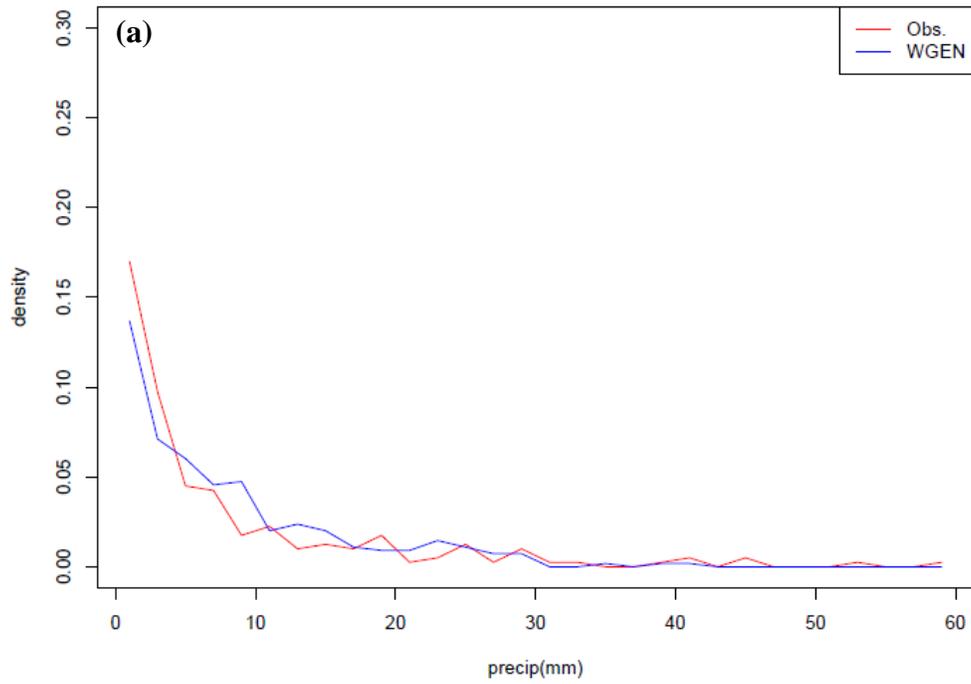
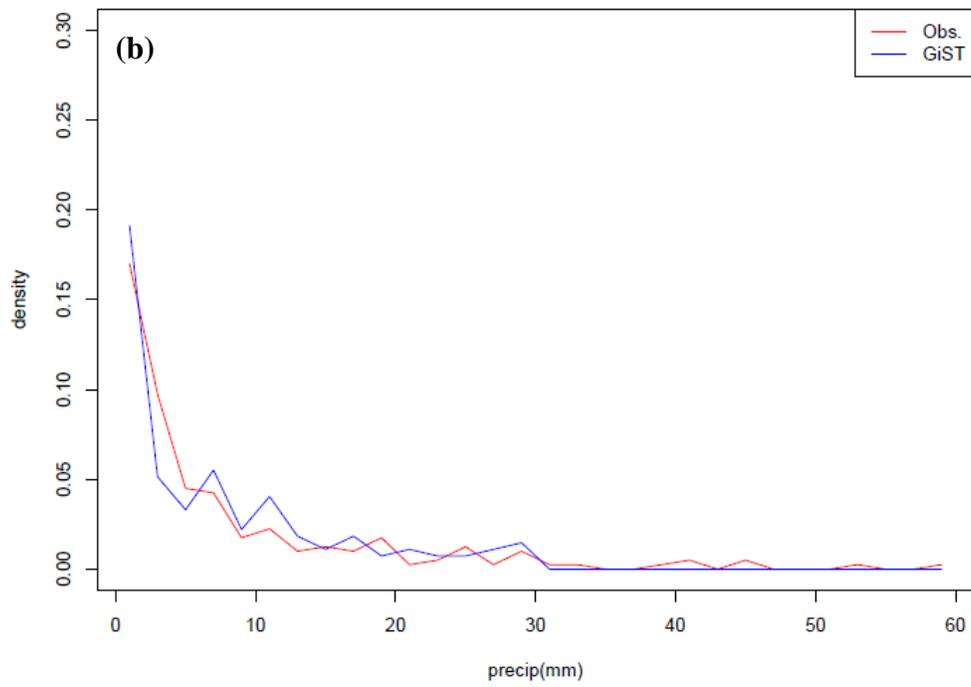


Figure 3.31. Generated vs. Observed PDF examples for September for station 460102 in Alderson, WV for each weather generator for WGEN (a), GiST (b), and GiSTR (c).

Comparison of Nonzero Rainfall PDF
Station: 460102 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 460102 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 460102 Month: 9

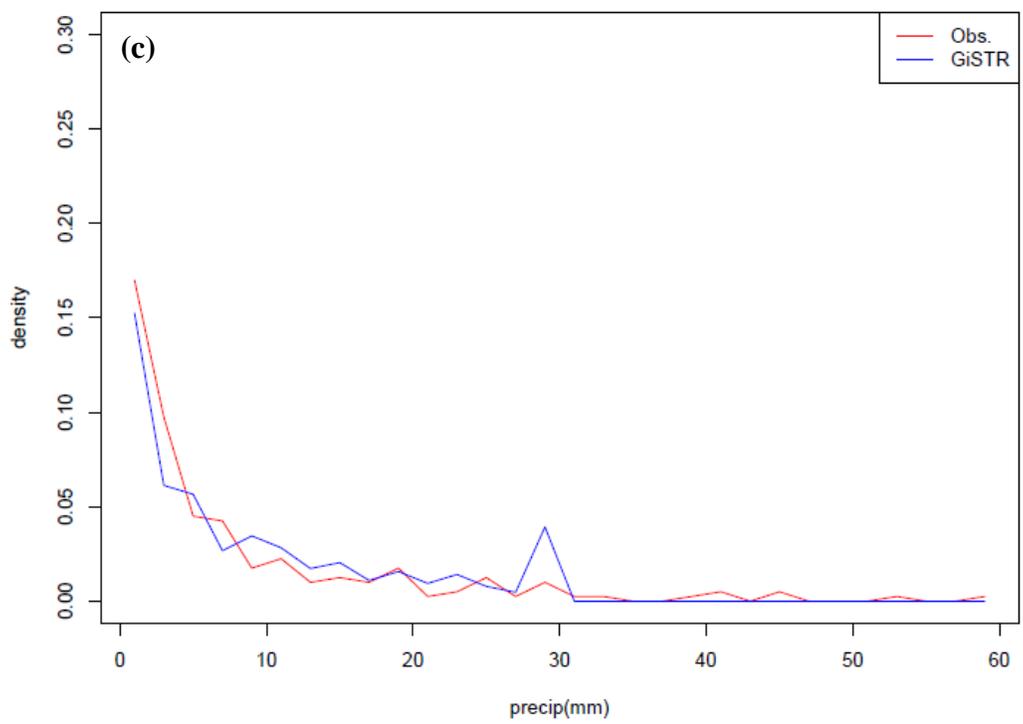
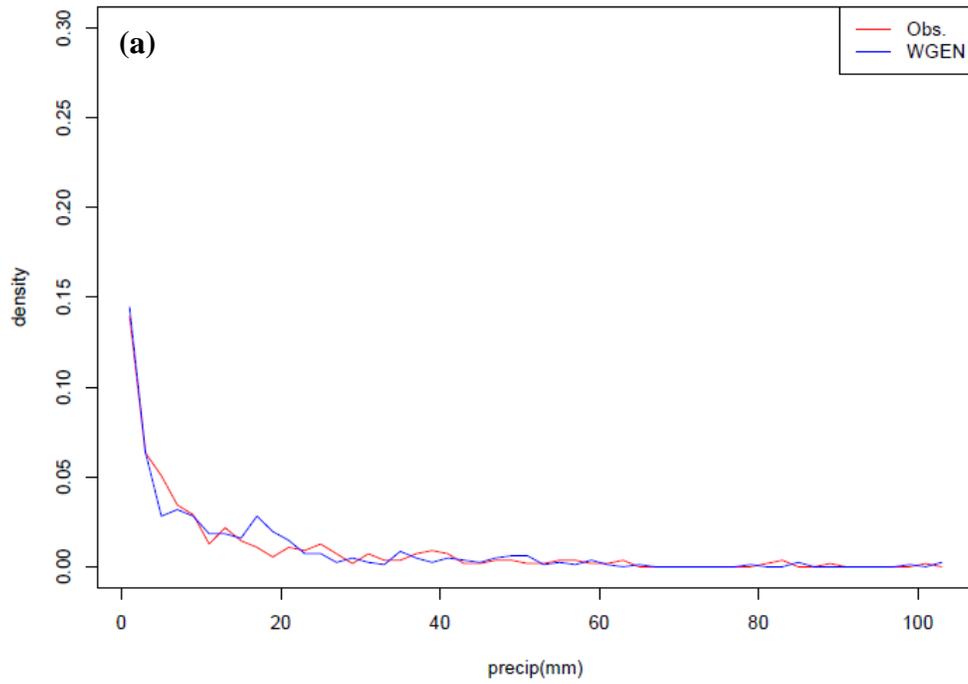
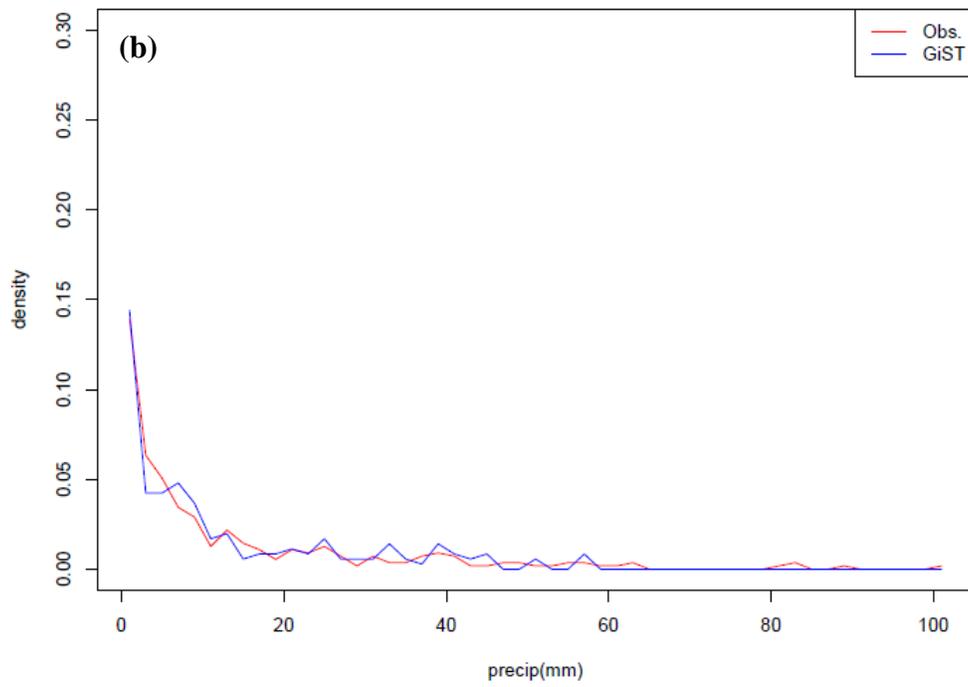


Figure 3.32. Generated vs. Observed PDF examples for September for station 089176 in Venice, FL for each weather generator for WGEN (a), GiST (b), and GiSTR (c).

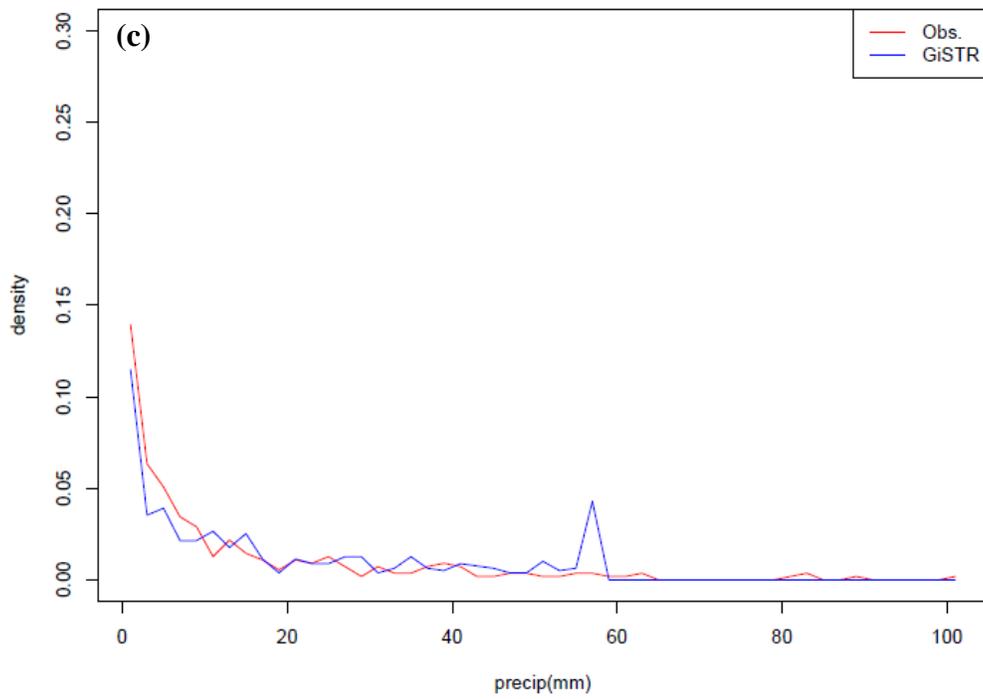
Comparison of Nonzero Rainfall PDF
Station: 089176 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 089176 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 089176 Month: 9



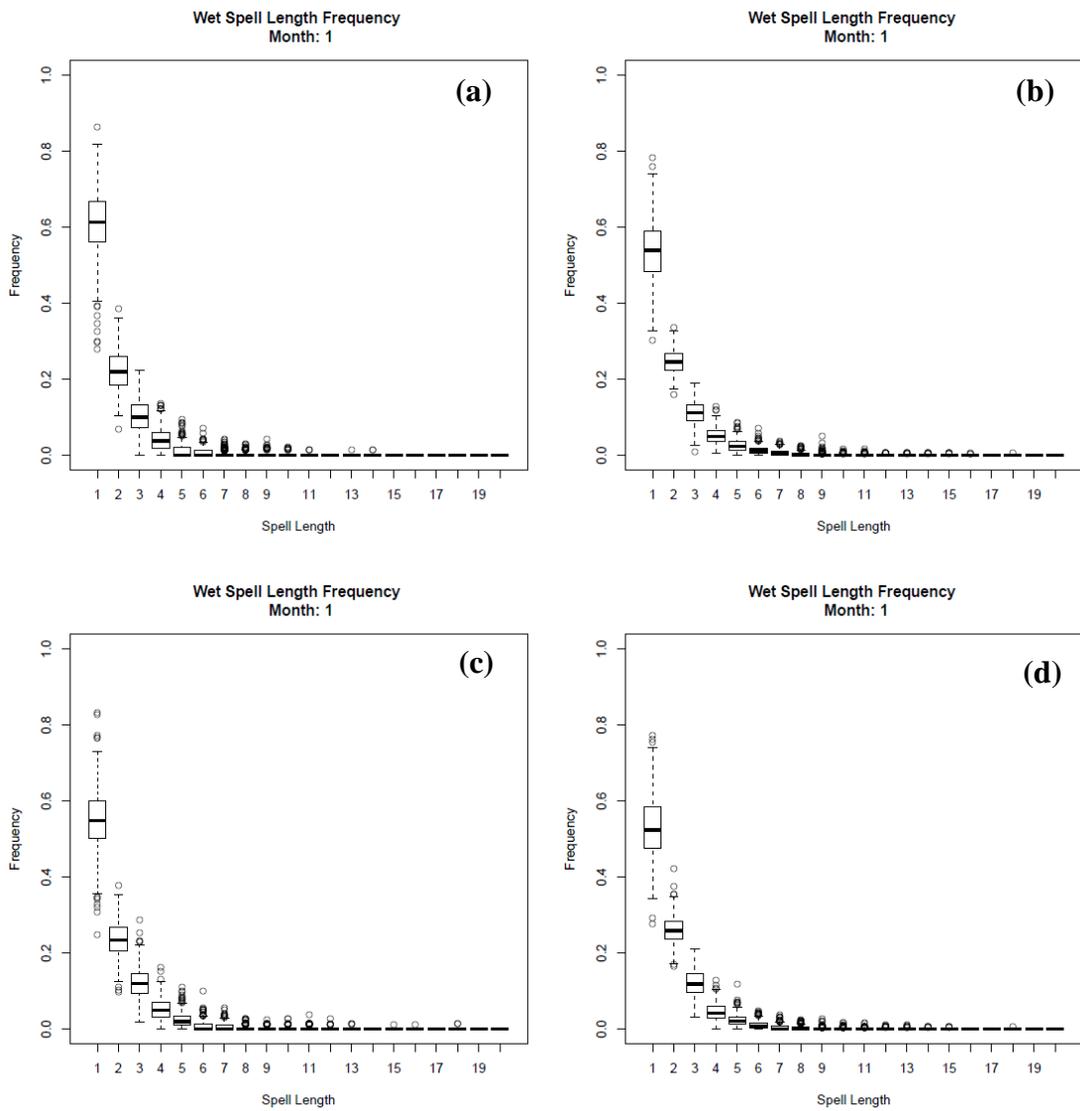


Figure 3.33. Boxplots of the frequency of January Wet Spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for the Southeast US domain.

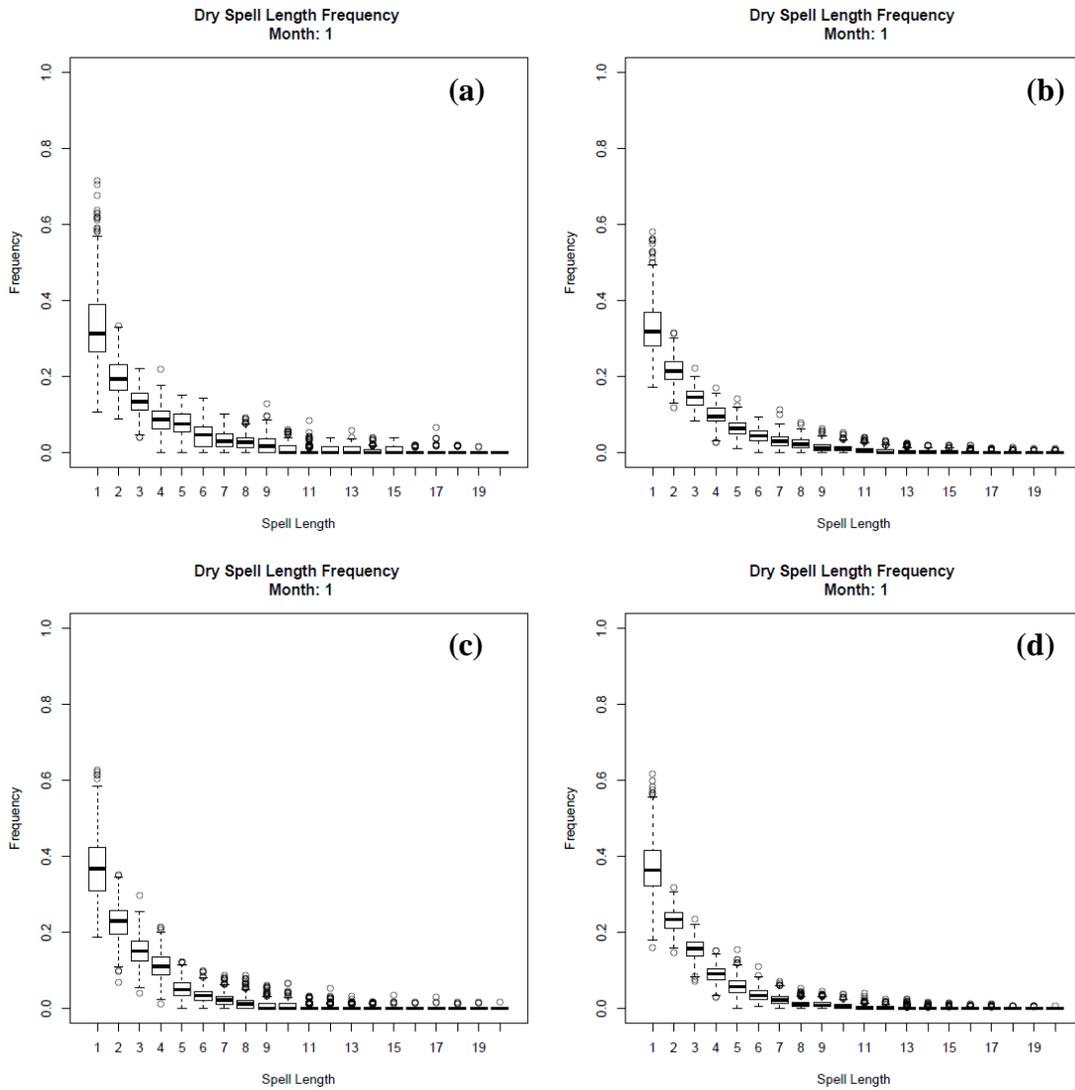


Figure 3.34. Boxplots of the frequency of January Dry Spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for the Southeast US domain.

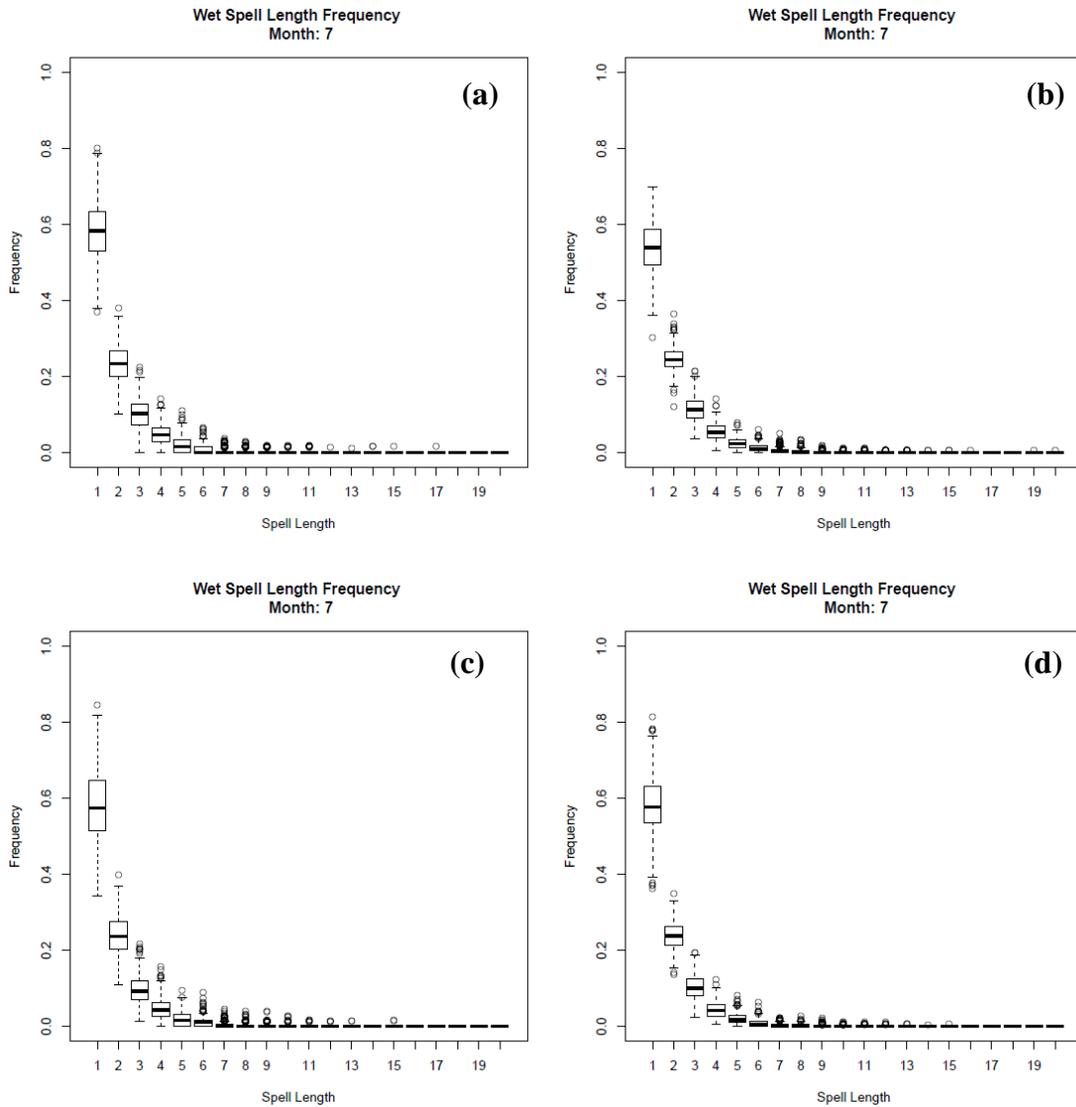


Figure 3.35. Boxplots of the frequency of July Wet Spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for the Southeast US domain.

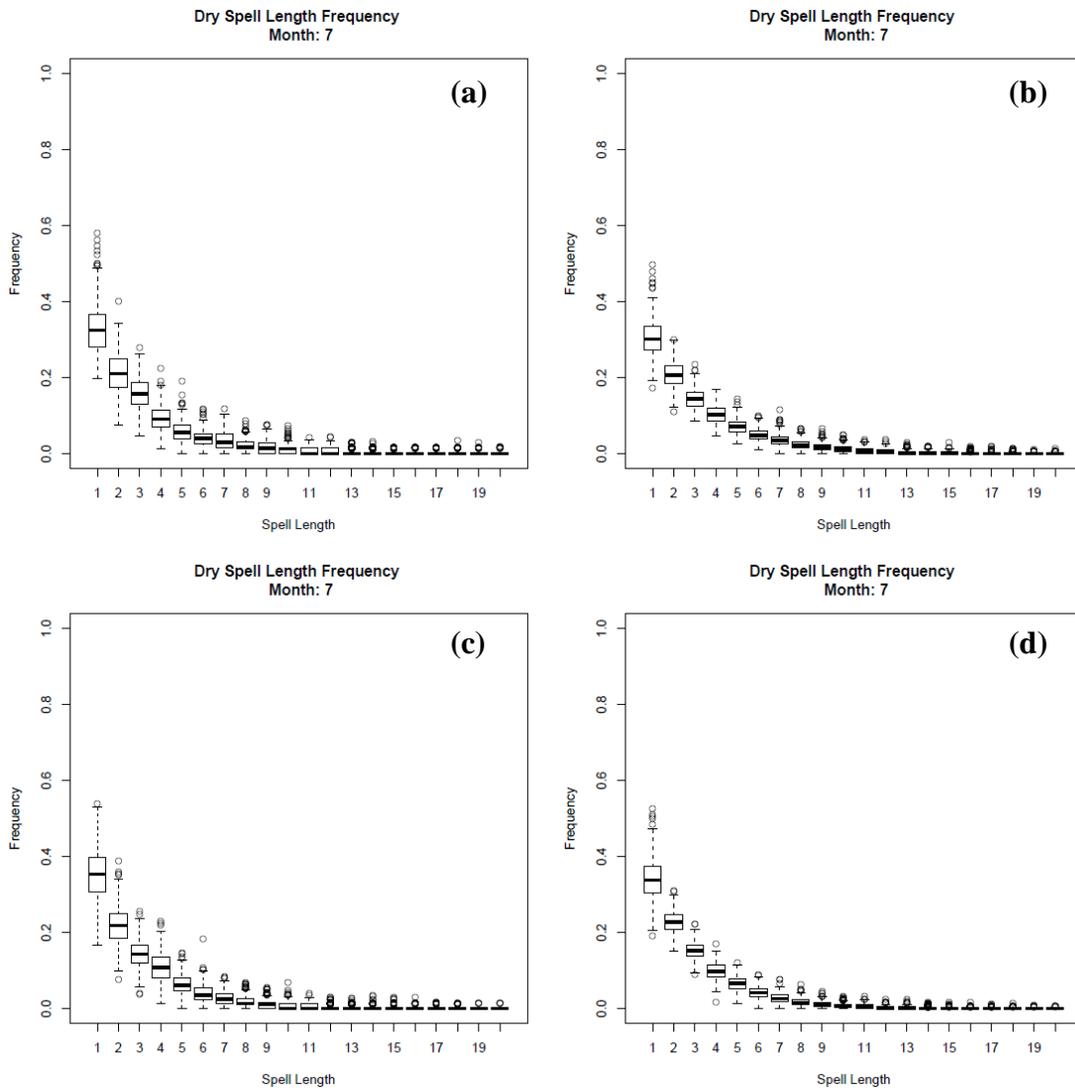


Figure 3.36. Boxplots of the frequency of July Dry Spells of various lengths for all stations for observations (a), WGEN (b), GiST (c), and GiSTR (d) for the Southeast US domain.

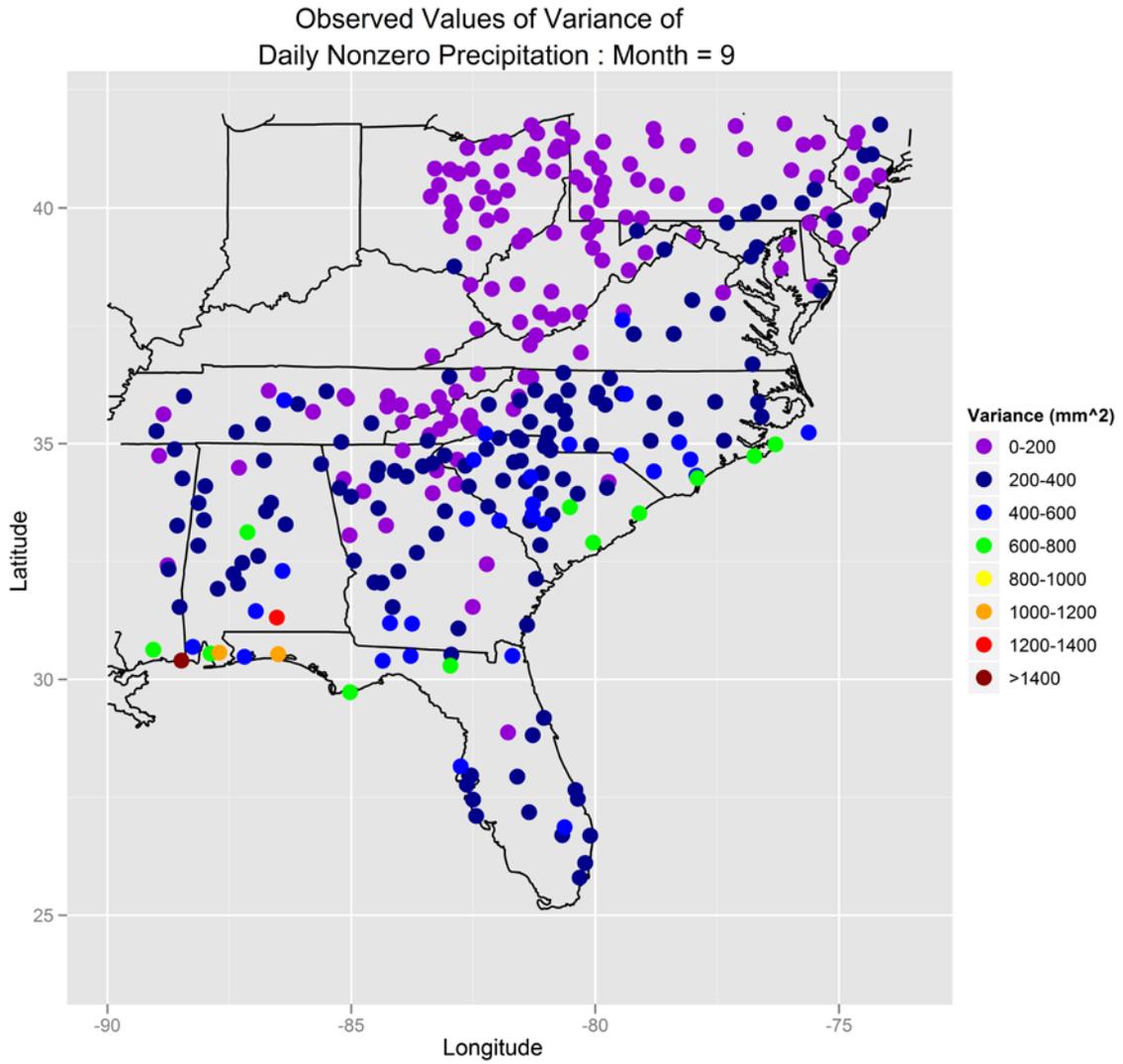


Figure 3.37. Observed values of the variance of daily nonzero precipitation (σ^2) (mm²) across the Southeast U.S. domain for September.

Chapter 4. Evaluation of Downscaling Variations – 1979-2000

While the previous chapter focuses on evaluating three different weather generators, this chapter focuses on evaluating four different variations on the general approach of downscaling precipitation with weather generators. These four variations are based on variations on two different assumptions that are used in the general approach. The first set of variations involves interpolating the values of each parameter for the present time for each GCM grid point from surrounding stations. The second set of variations addresses how the values of the gamma distribution parameters are determined.

The analysis of interpolation investigates the use of area averaging (AW) compared with locally weighted regression (LWR) as described by Wilks (2008). For the second set of variations, the method of moment estimators (MOM) is compared with the Greenwood and Durand approximation (DFIT) in determining the gamma distribution parameters for nonzero precipitation. Combining each produces four downscaling variations, which will each be evaluated using the same statistical methods and visual analysis as discussed in Chapter 3. In order to focus specifically on the different variations in the downscaling approach, only the GiSTR weather generator will be used for comparing techniques. GiSTR is used as the fixed generator since it adequately reproduces the observed spatial and temporal structure and better replicates the variability associated with extreme events, which is a historical problem for all weather generators (Wilks and Wilby, 1999). Since downscaling with any weather generator involves perturbing the parameters of each generator using the scaling relationship and then the weather generator of choice, it can be assumed that the error with regard to different variations used in the scaling relationship is the same regardless of the generator used.

The GCM used for this analysis is the Climate Forecast System Reanalysis (CFSR, Saha et al, 2010) with a resolution of approximately 2 degrees latitude and longitude, and each simulation will be evaluated against the observed statistics of each station in the Southeast U.S. domain for the period of 1979-2000. The CFSR data is used for this analysis because this version of CFS is operationally used for seasonal forecasting and also makes use of

coupled modeling of atmosphere, ocean, land-surface, and sea ice. The CFS Reforecast is not used in this analysis since it was not available at the time of the analysis. The four combinations considered in this study are the following:

- LWR-MOM – LWR for parameter interpolation and the MOM for the gamma distribution parameters
- AW-DFIT – AW for parameter interpolation and the DFIT for the gamma distribution parameters.
- LWR-DFIT – LWR for parameter interpolation and the DFIT for the gamma distribution parameters.
- AW-MOM – AW for parameter interpolation and the MOM for the gamma distribution parameters.

In addition to the analysis for differences between downscaling simulations, each simulation is compared to a Control simulation. This Control simulation reflects the use of GiSTR with observations being supplied to the generator rather than the use of the scaling relationship and the CFSR data. In short, this Control simulation is the same as the use of GiSTR with observations only as presented in Chapter 3. This additional analysis will allow for determining the increase in error that is associated with each downscaling variation and GiSTR compared to the error of GiSTR alone.

Following the comparison of downscaling techniques using GiSTR, a comparison of GiSTR and WGEN in a downscaling context using only one downscaling variation is presented. This comparison is intended to determine if the patterns between weather generators seen in Chapter 3 are consistent when used for downscaling. This combination of analyses tests the consistency of error between generators and downscaling variations, essentially determining if several of the assumptions in Chapter 4 and Chapter 5 are valid, or if future work should include a full evaluation of each weather generator and downscaling variation before considering applications to seasonal forecasting.

4.1. Downscaling Simulation Comparison

Following the analysis in the previous chapter the same set of parameters are used to evaluate each simulation. These parameters are also considered for evaluating the difference between each simulation and the Control.

4.1.1. Temporal Parameters

For each of the temporal parameters, the 1st order Markov transition probabilities (P_{01} and P_{11}), the unconditional probability of precipitation (π), and persistence (γ), there is very little difference observed in the error between simulations. For P_{01} , the difference in RMSE on average between each simulation is less than 0.002 for every month, as shown in Table 4.1. Similarly for P_{11} , the difference in RMSE on average between each simulation is less than 0.003 for every month as shown in Table 4.2. There is also no seasonal trend in the error of each simulation or difference in error between simulations for either parameter, which is also consistent with the Control.

In addition, testing for significance between each simulation for both parameters shows that there is no significant difference between each of these simulations. The two-sample T-test results for the RMSE of P_{01} and P_{11} between each downscaling simulation, summarized in Table 4.3, indicates that there is no significant difference between simulations. Comparing each of the error of each of these simulations to the error of the Control for each parameter shows several interesting results. On average, each simulation has smaller RMSE than the Control (by up to 0.0056) for values of P_{01} . In contrast, each simulation has larger RMSE than the Control (by up to 0.004) for values of P_{11} . While there are contrasting results in this comparison for each parameter, the two-sample T-tests comparing each downscaling simulation to the Control for each parameter, as summarized in Table 4.4, indicates that there is no significant difference in error between each simulation and the Control. The small difference between each simulation and the Control for each

parameter, as well as the lack of significant difference suggest that there is also no practical significance to these differences.

The results presented for the 1st order Markov transition probabilities are similar to results presented for the unconditional probability of rain (π) and persistence (γ). As shown by the RMSE comparison between simulations in Table 4.5, the difference in the RMSE of π between every simulation is less than 0.003 for every month. As shown by the RMSE comparison for γ between simulations in Table 4.6, the difference between each simulation is also less than 0.003 for all months. There are no seasonal trends in the difference in error between simulations for either parameter. These results are consistent with the results for the Markov transition probabilities.

Similarly to the analysis for the previous parameters, the two-sample T-tests for each parameter, summarized in Table 4.7, show that there is no significance in the RMSE between simulations. The p-values for each test were greater than 0.23, and the difference in the average RMSE was less than 0.0034 for both π and γ , indicating no significant difference in the RMSE of each parameter between each simulation. Comparing the error for each simulation for these parameters to the Control also shows some interesting results. For π and γ , each simulation has smaller RMSE than the Control by up to 0.0035. However, the T-tests comparing the RMSE for each parameter to the Control, summarized in Table 4.8, also indicates that there is no significant difference (p-values more than 0.22) in the error between each simulation and the Control. The results for this section indicate that there is no significant difference in the RMSE of each temporal parameter between downscaling simulations, or between each downscaling simulation and the Control. Given that these parameters reflect the occurrence of a wet day for each station, this implies that each simulation produces a similar number of wet days for each station. That is, if used to downscaling seasonal forecasts, each simulation will produce an equivalent number of wet days during a given period. Regardless of the similarities and lack of significant difference between simulations, the LWR-MOM and LWR-DFIT simulations have smaller error for these parameters, and therefore smaller error for the number of wet days for any given station and time period than the remaining simulations.

4.1.2. Spatial Structure Parameters

For all simulations, the spatial structure is not perturbed by the CFSR data. Given the resolution of CFSR, it cannot provide enough resolution to determine how the correlation of each individual station with every other station is influenced by the model, which is used by both GiST and GiSTR in the event generation process. For each of the simulations, the spatial structure (i.e. the correlation matrices) used in the generator is the spatial structure determined by the observations for 1971-2000. While the spatial structure is not directly perturbed, it is possible that the spatial structure is indirectly perturbed given how GiSTR generates precipitation events. GiSTR makes use of the 1st order Markov probabilities in the initial generation of precipitation events for the first two stations. In addition, the orthogonal Markov chain used by GiSTR considers both space and time, specifically the relationship between three stations in the current time, and information from the previous day for the station of interest. Therefore the perturbation to the 1st order Markov probabilities (i.e. P_{0l} and P_{1l}) may have some impact on the generated spatial structure. To explore the possible impacts on spatial structure, the RMSE of each of the correlation matrices for each simulation analyzed. For the correlation matrix of precipitation (a, ρ), the correlation matrix of precipitation events (b, ρ_{ev}), the correlation matrix of precipitation amounts (c, ρ_{am}), and the correlation matrix of extreme precipitation events (d, ρ_{ex}), the RMSE for each downscaling simulation across each month, shown in Figure 4.1, is approximately the same. Following the same analysis from the previous chapter, the observed and simulated average decay function correlograms are determined in order to evaluate the ability of each downscaling simulation to reproduce the appropriate decay in spatial structure with distance. The average decay function correlogram is the average correlation between stations in 5 km distance intervals. That is the value of the correlogram 15km is average correlation of stations between 10 and 15km apart. Figure 4.2 shows the observed and simulated average decay function correlogram for January for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). In January, each simulation shows a tendency to underestimate the correlogram for all four parameters with the smallest

underestimation for precipitation events (Figure 4.2b). What is also apparent from Figure 4.2 is that there is little difference in the correlogram produced by each simulation. These results are also consistent during the warm months. Figure 4.3 shows the observed and simulated average decay function correlogram for July for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). The results for July also indicate that each simulation produces a similar correlogram in all four cases and has a tendency to underestimate the observed correlogram. Each simulation provides the best replication of the correlogram for precipitation amounts (Figure 4.3c) and has the largest underestimation of the correlogram for precipitation events (Figure 4.3b). These results are consistent between months, but the results for both months show that the simulated correlograms for precipitation events have significantly more error than the correlograms produced by GiSTR outside of a downscaling context. The results for the correlogram comparison also indicate that the increase in ability to capture the observed correlogram more accurately in July is consistent with results shown in Section 3.2.2. for GiSTR. When used with observations, GiSTR accurately replicates the observed correlogram for precipitation events with little underestimation. Therefore, the application of CFS and the indirect influence on the spatial structure through the 1st order Markov chains in the event generation process influences the RMSE of the correlation matrix of precipitation events, by causing each simulation to more dramatically underestimate the correlation between stations at distances less than 1000 km. This indicates the possibility that there is significant difference in the RMSE of correlation matrix of precipitation events of one or more simulations from the Control. In addition, though not evident from Figures 4.2 and 4.3 the RMSE for correlation matrices for precipitation, precipitation amounts and precipitation extreme events may also be significant between each simulation and the Control.

Results from the two-sample T-test between simulations, summarized in Table 4.9, shows the simulations have an RMSE which is not significantly different from the RMSE of any of the other simulations. The average RMSE values for each generator for each of the four correlation matrices also shows that there is very little difference on average between each of the downscaling simulations. While the RMSE between simulations for each

correlation matrix is not significantly different, there are some cases where the RMSE of individual simulations for the correlation matrices is different from the Control. Table 4.10 summarizes the results of the two-sample T-test between each simulation and the Control. The test results indicate that the RMSE of ρ_{ev} for all four simulations is significantly larger than for the Control. For the other correlation matrices, only AW-DFIT had a significantly larger RMSE for ρ_{am} than the Control (p-value of 0.0019). Given that the RMSE of ρ_{ev} of each simulation was significantly larger than the Control, and there is no difference for the remaining spatial structure parameters, there is some indication of the impact of the perturbing the temporal parameters on values of ρ_{ev} produced by each simulation. Recall that each of the temporal parameters reflects the occurrence of precipitation events only, not the occurrence of specific amounts, extreme events, or the overall spatial structure of precipitation. Therefore, the perturbation of these temporal parameters only influences the structure of precipitation events. This influence becomes apparent when the results from Tables 4.4 and 4.8 are compared to the results from Table 4.9. Tables 4.4 and 4.8 shows that the error of π , γ , and P_{0j} is smaller than the Control, and the LWR-MOM and LWR-DFIT simulations offer larger improvements for each parameter compared to the remaining simulations. Table 4.9 shows that the difference between the each simulation and the Control for the RMSE of ρ_{ev} is smaller for the LWR-DFIT and LWR-MOM simulations. This similar pattern between the temporal parameters and ρ_{ev} indicates that, in absence of perturbing the spatial structure, downscaling variations which have lower error for the temporal parameters of precipitation will better replicate the spatial structure of precipitation events when GiSTR is used. This same result also applies to GiST, since it also makes use of the same orthogonal Markov chain and 1st order Markov probabilities for generating precipitation events. However, since the Markov transition probabilities are not used for generating precipitation amounts, the RMSE of ρ_{am} for all simulations is not significantly different from the Control, with the exception of the AW-DFIT variation.

Finally, as described in Chapter 2, GiSTR includes a separate section for extreme event generation, which follows the same structure as the regular event generation found in GiST and GiSTR. Therefore, GiSTR includes another set of 1st order and orthogonal Markov

probabilities specifically for generating extreme events. Following the downscaling procedure for precipitation events, everything related to the spatial structure of extreme events was not perturbed, but the 1st order Markov probabilities for extreme events were perturbed. The T-test results on the Markov transition probabilities for extremes, summarized in Table 4.11, indicates that the RMSE of each transition probabilities for the extremes for the LWR-DFIT and AW-DFIT was significantly larger than the LWR-MOM and AW-MOM simulations. The impact of the 1st order transition probabilities for extremes is corroborated by T-test results comparing each simulation to the Control summarized in Table 4.12. While there is no significant difference between each simulation and the Control for RMSE of ρ_{ex} , the difference from the Control for this parameter is larger for the LWR-DFIT and AW-DFIT than the remaining simulations. This suggests a similar impact on values of ρ_{ex} by the perturbation of the transition probabilities of extreme events as the impact of the normal Markov transition probabilities on values of ρ_{ev} .

4.1.3. Precipitation Amount Parameters

While the error of each simulation for most of the previous parameters are not significantly different from each other or from the Control, there are more distinct results for values of μ , σ^2 , $E[S(T)]$, and $Var[S(T)]$.

4.1.3.1. Mean Daily Nonzero Precipitation (μ)

Each simulation shows unique results for the error of the mean daily nonzero precipitation (μ). First, the LWR-MOM and LWR-DFIT variations have larger RMSE for this parameter compared to the other variations. The average RMSE for LWR-MOM, AW-DFIT, LWR-DFIT, and AW-MOM is 5.35, 3.78, 5.25, and 4.21 mm respectively. The LWR-MOM and LWR-DFIT variations also have the larger RMSE for every month than

other simulations, with the peak RMSE occurring in July through September, as shown by the RMSE comparison for each month in Figure 4.4.

The two-sample T-test result comparing each simulation, summarized in Table 4.13, indicates that the RMSE for LWR-MOM and LWR-DFIT is significantly larger than AW-DFIT and AW-MOM. The other combinations for this test, LWR-MOM versus LWR-DFIT and AW-DFIT versus AW-MOM, show no significant difference in RMSE. While there are significant differences in the RMSE of μ between individual simulations, the results of the two-sample T-test of each simulation against the Control (summarized in Table 4.14) indicate that every simulation has a significantly larger RMSE than the Control.

While there are significant differences between simulations shown by the two-sample T-tests for the RMSE of μ , there are also several interesting spatial patterns in the error of generated values of μ for each simulation. Consider as an example the error of μ across the domain for several months for each simulation. The error of μ associated with each station for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) in January, shown in Figure 4.5. This comparison indicates that each variation has a tendency to overestimate values of μ in western Pennsylvania, eastern Ohio, and West Virginia. However, Figure 4.5c also shows that the LWR-DFIT overestimates by 7 mm or more in the northeastern tip of the domain, while underestimating less in Alabama and Georgia than other simulations. This suggests that the RMSE of μ for LWR-DFIT is lower in the Carolinas, Alabama and Georgia than the RMSE for the entire domain in January (3.87 mm).

Figure 4.6 shows the error of μ for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for June. For this month, LWR-MOM (Figure 4.6a) and LWR-DFIT (Figure 4.6c) show a distinct tendency to underestimate values of μ in the western Carolinas. In addition, the LWR-MOM (Figure 4.6a), AW-DFIT (Figure 4.6b), and AW-MOM (Figure 4.6d) show a tendency to overestimate by more than 7 mm in Florida, while the LWR-DFIT (Figure 4.6c) overestimates by less than 2 mm for much of the same region. However, LWR-DFIT also overestimates μ by over 7 mm along the coast of the Carolinas and Georgia and underestimates values of μ by 4 to 7 mm in the Northeastern portion of the domain. September also shows some trends in the error of μ for both LWR-MOM and LWR-DFIT.

Finally, Figure 4.7 shows the error of μ for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for all stations in the domain for September. Both LWR-MOM (Figure 4.7a) and LWR-DFIT (Figure 4.7c) overestimate in the northeastern portion of the domain. However, while both overestimate in the same area, LWR-DFIT overestimates in this area by 18 to 24 mm and LWR-MOM overestimates by 8 to 12 mm. In addition, while there are these overestimates, LWR-DFIT seems to provide a better estimate across the central portion of the domain than the LWR-MOM, with the exception of the western Carolinas. While each of the LWR simulations shows some improvements in this month, it is apparent that both of the AW simulations provide the best estimate of μ during September in most of the domain. The spatial pattern between simulations demonstrated by the error of μ for January is common for most months across the domain, and the patterns shown in June and September also exists for July and August. Therefore, for months other than June through August it is likely that the RMSE for μ in the Carolinas, Georgia, Alabama, and Tennessee is less than the RMSE for the entire domain for LWR-DFIT. However for June through September, AW-DFIT provides the best estimate of μ across the same states. All four simulations also overestimate μ in June and September in Florida, which is consistent with the previous results for the evaluation of GiSTR in Chapter 3.

Both LWR-MOM and LWR-DFIT make use of LWR for the interpolation of parameters to GCM grid points, and both provide underestimates or overestimates of μ in the northeast corner of the domain. However, the maximum over- or underestimation occurs in different months between these two simulations; April for LWR-MOM and September for LWR-DFIT. Simulations AW-DFIT and AW-MOM, which make use of AW for the parameter interpolation to GCM grid points, do not show the same spatial pattern in this domain. Regardless of the patterns (both seasonal and spatial) shown by these results, it is apparent that the AW-DFIT has the lowest error for values of μ . Therefore, in seasonal forecasting and those studies of near-term climate trends, and for crop modeling in the same time periods, AW-DFIT is the recommended variation given that provides the most accurate values of the mean daily nonzero precipitation.

4.1.3.2. Variance of the Daily Nonzero Precipitation (σ^2)

While the RMSE comparison for values of μ indicates that AW-DFIT shows the smallest error over the entire domain, the error for the variance of the daily nonzero precipitation (σ^2) does not follow the same pattern. For the entire domain, the LWR-MOM and AW-MOM have smaller RMSE for values of σ^2 than the other simulations. The average RMSE for LWR-MOM, AW-DFIT, LWR-DFIT, and AW-MOM is 127.44, 150.28, 130.32, and 127.30 mm² respectively. Comparing the RMSE of σ^2 for all months, shown in Figure 4.8, it is apparent that the RMSE is smaller for LWR-MOM and AW-MOM in every month than other simulations, and the peak RMSE for each variation is in April, September, and October. The results of the two-sample T-tests between each simulation, summarized in Table 4.15, indicate that there is no significant difference in the RMSE of σ^2 between downscaling simulations. In addition, it is also apparent that the average RMSE is larger for the LWR-DFIT and AW-DFIT variations than the LWR-MOM and AW-MOM variations. While not a significant difference, this result suggests that using the MOM to estimate gamma parameters improves error with regards to σ^2 in the domain. Finally, the T-test results comparing the RMSE of each simulation to the Control, summarized in Table 4.16, shows that the RMSE is significantly larger than the Control for all simulations except the LWR-MOM simulation.

The RMSE comparison for each month, shown in Figure 4.8 highlights the months with peaks in the error of each downscaling simulation to be April and October. Figure 4.9 shows the error of σ^2 for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for all stations in the domain in April. For April, each downscaling simulation underestimates values of σ^2 by up to 300 mm² in most of the domain. However, there are several stations where each simulation underestimates values of σ^2 by more than 500 mm², primarily in Alabama and western Florida. Figure 4.10 shows the error of σ^2 for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for all stations in the domain in October. In October, there is a similar pattern where each simulation has underestimated in

most of the domain by 300 mm^2 or less and several locations where the underestimation is greater than 500 mm^2 along the northern Gulf Coast.

Regardless of which stations are underestimated, it is also evident that for those stations the underestimation of σ^2 is largest for AW-DFIT (up to 1300 mm^2) and AW-MOM (up to 1500 mm^2) in both April and October compared to the LWR-MOM (up to 1100 mm^2) and the LWR-DFIT (up to 1100 mm^2). This trend was also common for most other months, suggesting that while the AW interpolation produces more accurate values of μ in downscaling with weather generators, the LWR interpolation produces more accurate values of σ^2 for the most stations. However, while LWR-MOM and LWR-DFIT may have less error for values of σ^2 , there is no significant difference in the RMSE between downscaling simulations. Given that the LWR-MOM variation has the fewest number of stations where σ^2 is underestimated by more than 500 mm^2 , has one of the lowest average RMSE, and the RMSE for the LWR-MOM is not significantly different to the Control, the LWR-MOM variation has the lowest error for values of σ^2 .

4.1.3.3. Average Total Precipitation ($E[S(T)]$)

The average total precipitation ($E[S(T)]$) was shown in Chapter 2 (Equation 2.7) to be directly related to the mean daily nonzero precipitation (μ), and the pattern of error between each downscaling variation is very similar to the results shown in Section 4.1.3.1. The AW-DFIT variation has the smallest RMSE, which is consistent with the results in Section 4.1.3.1. The average RMSE of $E[S(T)]$ for LWR-MOM, AW-DFIT, LWR-DFIT, and AW-MOM is 65.73, 44.47, 64.61, and 52.50 mm respectively. Figure 4.11 shows a comparison of the RMSE of $E[S(T)]$ for all months for each simulation. From this comparison, it is apparent across all months LWR-MOM and LWR-DFIT have larger RMSE for values of $E[S(T)]$ than other simulations while AW-DFIT has the smallest RMSE. Testing the RMSE of each simulation against every other simulation, the T-test results, presented in Table 4.17, show similar to results to the T-tests for μ . The two-sample T-test between simulations shows that the RMSE for LWR-DFIT and LWR-MOM for $E[S(T)]$ is significantly larger

than the RMSE for AW-DFIT, but not significantly larger than the RMSE for AW-MOM. These results are all similar to the results presented in Section 4.1.3.1 and these similarities demonstrate the relationship between μ and $E[S(T)]$. In addition, the two-sample T-tests for the RMSE for each simulation against the Control, summarized in Table 4.18, shows that the RMSE of simulations LWR-DFIT and AW-MOM are significantly larger than the Control. However, the RMSE for $E[S(T)]$ for AW-DFIT and LWR-MOM are not significantly different compared to the Control.

The results for the evaluation of $E[S(T)]$ follow a similar result to the evaluation of μ for each simulation. Therefore, the result from Chapter 3 (that GiSTR overestimates $E[S(T)]$), that each downscaling simulation has a larger RMSE than the Control, and that each overestimates μ suggests that each downscaling simulation generally overestimates $E[S(T)]$ by a larger amount than the Control. In addition, the RMSE of $E[S(T)]$ for AW-DFIT and AW-MOM is on average less than the RMSE for LWR-MOM and LWR-DFIT. This indicates that the AW interpolation provides more accurate values of $E[S(T)]$ for most stations when used in downscaling. This is also consistent with the results for values of μ . This also emphasizes the earlier result presented in Section 4.1.3.1, that AW-DFIT is recommended for forecasts for situations where normal precipitation is anticipated on the large scale.

4.1.3.4. Inter-annual Variability ($Var[S(T)]$)

While the results for the average total precipitation ($E[S(T)]$) and mean daily nonzero precipitation (μ) are similar, the results for the inter-annual variability ($Var[S(T)]$) and the variance of daily nonzero precipitation (σ^2) are not. The AW-MOM variation has the smallest RMSE for values of $Var[S(T)]$. The average RMSE of LWR-MOM, AW-DFIT, LWR-DFIT, and AW-MOM is 2342.17, 1936.38, 2321.12, and 1846.90 mm^2 respectively. The RMSE comparison between simulations for each month, shown in Figure 4.12, indicates that the RMSE for LWR-MOM and LWR-DFIT is larger, and the RMSE for AW-DFIT and AW-MOM is smaller. The comparison also shows that the peak RMSE for each simulation

is in August through September. The T-test results comparing the RMSE between simulations, summarized in Table 4.19, are in contrast to the T-test results for σ^2 . The two-sample T-test indicates that LWR-MOM and LWR-DFIT have a significantly larger RMSE for $Var[S(T)]$ than AW-MOM. However, the testing also shows that the RMSE for other simulations are not significantly different. This is contradictory to the results presented in Section 4.1.3.2, which indicates no significant difference in the RMSE between simulations for values of σ^2 . The two-sample T-test results comparing each simulation to the Control for $Var[S(T)]$ are shown in Table 4.20. The RMSE of $Var[S(T)]$ for each downscaling simulation is significantly larger than the RMSE for the Control, which is consistent with the results for σ^2 .

As shown by Equation 2.8, $Var[S(T)]$ is dependent on both μ and σ^2 . Therefore in cases where the RMSE of μ is larger for LWR-MOM and LWR-DFIT, this will contribute most to the RMSE of $Var[S(T)]$ for those same simulations. Given that the error of μ for LWR-MOM and LWR-DFIT can be more than 7 mm for some stations, and that RMSE of σ^2 is not significantly different between simulations, the RMSE of μ has the largest effect on the RMSE of $Var[S(T)]$ for the domain. However, given that some simulations provide better estimates of μ and σ^2 in certain areas of the domain, it is possible that for specific states in the domain the simulations may have better values of the RMSE of each of these parameters.

4.1.4. Probability Distribution Function Analysis

The statistical comparison shows that all four simulations have significantly larger RMSE for most parameters than the Control. In addition the evaluation also shows that AW-DFIT and AW-MOM produce significantly more accurate values of the μ and $E[S(T)]$ than LWR-MOM and LWR-DFIT, with no significant difference in the RMSE between simulations for σ^2 , and some significant differences for $Var[S(T)]$. However, the PDFs produced by each downscaling simulation vary significantly across the domain and for individual months. For example, consider the PDFs produced by each variation for several stations across the domain in January. Figure 4.13 shows the PDF generated by LWR-MOM

(a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) compared to the observed PDF for station 083163 in Ft. Lauderdale, FL in January. It is apparent that LWR-MOM (Figure 4.13a) best matches the observed PDF of nonzero precipitation for this station. Recall from Chapter 2 that GiST and GiSTR both contain a capping value for every station in every month at 110% of the average maximum daily precipitation for that month and station in order to prevent unrealistically large values from occurring. The remaining three downscaling simulations have a tendency to hit this capping value on high precipitation events for this station more than LWR-MOM, and hit the cap most frequently for LWR-DFIT (Figure 4.13c). In contrast, for stations in North Carolina all the downscaling simulations tend to produce the PDF of observed nonzero precipitation similar to the Control. For example, the PDF comparison to the observed PDF for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for station 311975 in Concord, NC in January, shown in Figure 4.14, indicates that each simulation tends to replicate most of the observed PDF. While all the downscaling simulations provide a good fit to the observed PDF, LWR-DFIT (Figure 4.14c) provides the best fit for this station and others in North Carolina. This also in contrast to the PDFs produced for stations in New Jersey, in the northeast corner of the domain. Recall from Figure 4.3 that the values of μ in this area in January were overestimated by more than 7 mm by LWR-DFIT. Comparing the generated PDFs for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) to the observed PDF for station 283516 in Greenwood Lake, NJ in January, shown in Figure 4.15, indicates that the cap is present for all the simulations. However, LWR-DFIT (Figure 4.15c) fails to capture the PDF for lower precipitation amounts where the remaining simulations do capture the PDF for these amounts. Regardless of the fact that each simulation hits the cap that is inherent to GiSTR, this analysis indicates that LWR-DFIT overestimates values of μ because it fails to produce enough low intensity precipitation events. That is, LWR-DFIT generates more heavy precipitation for this station than the other downscaling simulations, which causes the drastic overestimation of μ in this region.

For September, the comparison of PDFs for the same stations also shows some unique patterns as the comparison in January. Figure 4.16 shows a comparison of the PDFs

generated by LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) compared to the observed PDF for station 083163 in Ft. Lauderdale, FL in September. Every downscaling simulation fails to capture smaller precipitation amounts, resulting in an overestimation of μ by all downscaling simulations for this station in September. This result is consistent for all the stations in Florida where μ is overestimated by over 7 mm, which is caused by GiSTR implementation in estimating heavy precipitation events as discussed in Chapter 3. Figure 4.17 compares the PDF produced by LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) to the observed PDF for station 311975 in Concord, NC in September. For this example, each of the downscaling simulations in September produces PDFs which capture most of the observed PDF. However, while each simulation captures most of the observed PDF of nonzero precipitation, LWR-DFIT (Figure 4.17c) best captures the PDF with the exception of the overestimation of precipitation at the cap. These patterns in the PDFs produced for station 311975 are similar for most stations in North Carolina. This suggests that while LWR-DFIT seems to hit the built in cap more frequently, it is still able to capture most of the PDF for stations in North Carolina. This in turn leads to better estimates of μ in North Carolina compared to the LWR-MOM in September. Finally, for stations in the northeast corner of the domain, the differences between simulations are clear for these stations. As an example consider the PDFs produced by LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) compared to the observed PDF in September for station 283516 in Greenwood Lake, NJ, shown in Figure 4.18. LWR-MOM (Figure 4.18a) and LWR-DFIT (Figure 4.18c) fail to capture low precipitation amounts and cause precipitation to occur too frequently at the cap for this station. Similarly to the Florida stations, the failure of the LWR-MOM and LWR-DFIT to capture the PDF and overestimate the frequency of values at the cap (which is common to all stations in the northeast corner for both simulations) causes values of μ in this area to be overestimated by 8 to 12 mm for LWR-MOM and 20 to 24 mm for LWR-DFIT. As mentioned in Chapters 2 and 3, the capping value for each of these locations is common to both GiST and GiSTR, which is used to prevent unrealistic values of precipitation from being produced by either generator. As noted in Chapter 3, GiSTR hits this cap more frequently than GiST, which contributes to the

RMSE of all the precipitation amount parameters. However, the error caused by the GiSTR generation process and capping value does not cause values of μ to be overestimated as much as when any downscaling variation is used. In situations where the simulation used does not capture the frequency of small precipitation amounts and overestimates the frequency of precipitation at the capping value, the result is an overestimation of values of μ which also causes values of $E[S(T)]$ to be overestimated. These results suggest that while GiSTR and the downscaling simulations both contribute to the RMSE of different precipitation amount parameters, removing or adjusting the capping value will likely improve the RMSE of these parameters by allowing large precipitation amounts to occur. In addition, in the northeastern corner of the domain the overestimation of μ resulting from failing to capture the frequency of small precipitation amounts only occurs for LWR-MOM and LWR-DFIT. While this occurs for more months for LWR-DFIT, both simulations make use of the LWR interpolation, suggesting that errors in this area are the result of the parameter interpolation more than the moment estimation.

4.1.5. Dry and Wet Spell Frequency Analysis

While there are several clear differences in the PDFs produced by each downscaling simulation for various locations in the domain, there are little differences in the frequency of the various dry and wet spells produced by each downscaling simulation. Regardless of the month considered, each downscaling simulation produces similar results to every other downscaling simulation. In every Markov chain generator those parameters that are used in precipitation event generation also control the frequency of dry and wet spells. Therefore, the similarities of the dry and wet spell frequencies between downscaling simulations is consistent with earlier testing which indicated no significant difference in the RMSE of P_{11} , P_{01} , π , and γ between downscaling simulations. While the dry and wet spell frequencies are similar between downscaling simulations, there are some differences between the observed dry and wet spell frequencies. Figure 4.19 shows the observed frequency of wet spells of various lengths for January (a) compared to those generated by LWR-MOM (b), AW-DFIT

(c), LWR-DFIT (d), and AW-MOM (e) for all stations in the domain. For wet spells in January, each downscaling simulation slightly underestimates one-day wet spells and slightly overestimates two-day wet spells. Figure 4.20 shows the observed frequency of dry spells of various lengths for January (a) compared to those generated by LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d), and AW-MOM (e) for all stations in the domain. Similarly to the January wet spells, for dry spells in January each downscaling simulation slightly overestimates one- and two-day dry spells. For both the wet and dry spell frequencies each simulation accurately captures the frequency of spells that are three days and longer.

Figure 4.21 shows the observed frequency of wet spells of various lengths for July (a) compared to those generated by LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d), and AW-MOM (e) for all stations in the domain. For July wet spells, each downscaling simulation show very little difference from the observations. Figure 4.22 shows the observed frequency of dry spells of various lengths for July (a) compared to those generated by LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d), and AW-MOM (e) for all stations in the domain. This comparison shows a similar pattern to wet spells in July, indicating very little difference between simulated and observed frequencies of dry spells regardless of length. This pattern of minor differences between observed and generated dry and wet spell frequencies is also consistent with the earlier testing on the RMSE of P_{11} , P_{01} , π , and γ between each downscaling simulation and the Control. The earlier testing indicated that the RMSE for several of these parameters for several downscaling simulations is not significantly different from the Control. Therefore the dry and wet spell frequencies simulated by each downscaling simulation have equivalent or smaller error than the dry and wet spell frequencies simulated by the Control.

Each of the analyses as part of this evaluation focuses on the difference between downscaling simulations across the Southeast U.S. Each variation shows a tendency to be more accurate for specific parameters. Given this result, there is at least one conclusion evident from this analysis. For most parameters reflecting precipitation occurrence, spatial structure and amounts, every other downscaling variation had more accuracy than the traditionally used AW-MOM variation. Of the three variations remaining, the LWR-MOM

variation best captures most parameters including spatial structure, temporal parameters and the variance of daily nonzero precipitation. However, given that the previous chapter evaluates individual weather generators outside of a downscaling context, the next section presents a comparison of two weather generators used in downscaling to determine which patterns from Chapter 3 remain between generators used for downscaling.

4.2. Weather Generator Comparison - Downscaling Context

In the previous section, the generator used to downscale was held constant in order to evaluate the effect of four different combinations of the two assumptions used in downscaling with weather generators. This section will return the weather generator evaluation in the context of downscaling. WGEN and GiSTR are used to downscale CFSR data for the same period as in the previous section (1979-2000) and for the same set of stations. WGEN is used in this analysis since it is often used as a standard for building other Markov chain weather generators. GiSTR is used for this analysis since the results of Chapter 3 indicate that it better replicates the variability associated with extreme events than WGEN, which is a historical program for most generators (Wilks and Wilby, 1999). For this analysis each downscaling assumption is fixed to one technique. Given that it has the smallest error for most parameters, the LWR is used to interpolate the observed weather generator parameters to each CFSR grid point and the MOM is used to determine the values of the gamma distributions used by each generator. Therefore, please note that the GiSTR downscaling simulation in this section is the same as LWR-MOM in Section 1. As with the previous section each generator has the observed parameter perturbed through the scaling relationship using the two fixed assumptions, each with a 30-member ensemble. This section focuses on whether the results shown in Chapter 3 are similar when each generator is applied in the context of downscaling.

4.2.1. Temporal Parameters

Comparing the 1st order Markov transition probabilities (P_{01} and P_{11}) produced by each generator in a downscaling context between generators shows a different result for the same comparison outside a downscaling context. Table 4.21 shows a comparison of the RMSE of P_{01} and P_{11} between the WGEN and GiSTR simulations for each month. For P_{01} , the RMSE for the WGEN simulation is similar to the GiSTR simulation on average and for every month. The RMSE for WGEN, while similar to GiSTR, is also larger than the RMSE for the GiSTR simulation for most months by up to 0.031. In contrast, for P_{11} the RMSE for the WGEN simulation is larger than the GiSTR simulation on average and for every month. Though there is no seasonal trend in the RMSE for either simulation, the RMSE for WGEN is larger than GiSTR by up to 0.24 for most months.

The results of the error comparison for the 1st order Markov probabilities are consistent with the results of the error comparison for the unconditional probability of rain (π) and persistence (γ). Table 4.22 shows a comparison of the RMSE of π and γ between the WGEN and GiSTR simulations for each month. First, for values of π , the WGEN simulation has the largest RMSE on average and across every month in the domain. Similarly to the results for P_{11} , the RMSE of π for the WGEN simulation is larger than the GiSTR simulation by up to 0.23 for every month, with no seasonal trend in the RMSE for either simulation. The RMSE for values of γ also shows a similar pattern to the results for P_{11} . The WGEN simulation has a larger average RMSE for this parameter than the GiSTR simulation. In addition, the RMSE of the WGEN simulation is larger for every month by up to 0.18, with no seasonal trend in the RMSE for either simulation.

Given that the direct relationship between P_{01} , P_{11} , π , and γ described in Chapter 2 (Equations 2.5 and 2.6), the RMSE of P_{01} and P_{11} produced by each of these downscaling simulations influences the RMSE of π and γ through this relationship. Therefore, since the RMSE of P_{01} and P_{11} in WGEN increases when used for downscaling, the RMSE of π and γ also increases. These results are in contrast to the results presented in Chapter 3, where WGEN was shown to have smaller error than GiSTR for each of these parameters. Testing

the difference in the RMSE between the WGEN simulation and the GiSTR simulation also shows several interesting results. The two-sample T-test results for each parameter, summarized in Table 4.23, indicate that the RMSE of P_{11} , π , and γ is significantly larger for WGEN than GiSTR. This result is different than the results presented in Chapter 3, where WGEN had significantly smaller RMSE for all these parameters in the same domain. Therefore, this indicates that applying WGEN in downscaling causes an increase in the RMSE for these parameters and causes a decrease in the RMSE for these parameters for GiSTR. However, it is important to note that this comparison only uses one of the four combinations of the downscaling assumptions discussed in Section 1. Therefore, further simulations with the other combinations of downscaling assumptions are needed to confirm if these results are common across all assumptions, or if WGEN is sensitive to the choice of downscaling variation.

4.2.2. Spatial Structure Parameters

Following the approach presented in Section 4.1.2, the spatial structure was not directly perturbed by the downscaling variation used in either WGEN or GiSTR. While the previous discussion highlighted how the spatial structure produced by GiSTR could still be impacted by the perturbation of the temporal parameters, the same influence does not apply to the WGEN simulation. Since WGEN does not consider both the spatial and temporal structure of precipitation in the event generation process, the perturbation applied to the temporal parameters has no influence on the spatial structure produced by WGEN. In contrast to the results for P_{01} , P_{11} , π , and γ , the results for the spatial structure are consistent to the results presented in Chapter 3. Figure 4.23 shows the RMSE of each simulation for the correlation matrix of precipitation (a, ρ), the correlation matrix of precipitation events (b, ρ_{ex}), the correlation matrix of precipitation amounts (c, ρ_{am}), and the correlation matrix of precipitation extreme events (d, ρ_{ex}) for all months. Regardless of the parameter considered, the RMSE of the spatial structure for WGEN is larger than the RMSE for GiSTR. Figure 4.24 shows the observed average decay function correlogram compared to those created by each

simulation for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d) for January. Similarly to the results in Section 3.2.2, GiSTR better replicates the observed correlogram than WGEN for all four situations, but best replicates the observed correlogram for precipitation events (Figure 4.24b). The results for July also show similar results to the generator comparison in Section 3.2.2. Figure 4.25 shows the observed average decay correlogram compared to those created by each simulation for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). As with January, the correlograms generated by the GiSTR simulation better replicate the observed correlogram than WGEN for all four parameters. However, GiSTR best replicated the observed correlogram for precipitation amounts (Figure 4.25c). As mentioned in Section 3.2.2, the observed correlogram decreases more rapidly with distance in summer months compared to winter months because of the influence of isolated convection during summer. The rapid decrease in the observed correlogram allows the WGEN simulation to better replicate the observed correlogram for stations more than 500km apart. However, the two-sample T-test on the RMSE of each correlation matrix for each downscaling simulation, summarized in Table 4.24, indicates that the RMSE for WGEN is significantly larger than GiSTR for every parameter. These results are consistent with the results presented in Chapter 3 and with the correlograms produced by each simulation, which is expected since the spatial structure was not directly affected by the parameter perturbation applied in the downscaling process with either weather generator. The RMSE for each spatial structure parameter for the WGEN simulation was tested against the RMSE for each spatial structure parameter produced by WGEN used outside of a downscaling context (that is, the RMSE of WGEN shown in Section 3.2.2). The results of this additional analysis show that the RMSE for each spatial structure parameter in this WGEN simulation is not significantly different from the RMSE of the spatial structure parameters presented in Chapter 3 (p-values > 0.33). This result is expected since WGEN does not consider the spatial structure during the event generation process, which indicates that the perturbation process applied to the temporal parameters of WGEN does not impact the spatial structure.

4.2.3. Precipitation Amount Parameters

For the precipitation amount parameters (μ , σ^2 , $E[S(T)]$, and $Var[S(T)]$) comparing the downscaling simulations of WGEN and GiSTR also shows significantly different results between generators than described in Section 3.2.3.

4.2.3.1. Mean Daily Nonzero Precipitation (μ)

In Section 3.2.3.1, WGEN is shown to have smaller error than GiSTR for values of the mean daily nonzero precipitation (μ) across the domain. However, when used in a downscaling simulation, the error for values of μ is larger for WGEN than GiSTR. The average RMSE of μ for WGEN and GiSTR is 5.93 and 5.35 mm respectively. The comparison of the RMSE across all months, shown in Figure 4.26, indicates that the RMSE of μ for WGEN downscaling simulation is similar to the RMSE for the GiSTR downscaling simulation except in certain months. In April, the RMSE for WGEN is 19.5 mm, while in June, July and August the RMSE for WGEN is less than 3 mm. While the RMSE difference between WGEN and GiSTR in June, July and August reflects the results in Chapter 3, the RMSE difference for April suggests that there may be a few stations where the value of μ is unrealistically overestimated or underestimated in the domain. Figure 4.27 shows the difference between the WGEN simulated values of μ and the observed values for each station in April. The WGEN simulation has a tendency to overestimate values of μ by more than 7 mm in the northeast corner of the domain and in coastal Georgia. The tendency to overestimate in the northeast corner of the domain is common to both simulations. However, the actual difference between the simulated and observed values for WGEN is between 30 and 50 mm in this region, while GiSTR overestimates in the same region by 14 to 22 mm. The overestimation for the coastal stations in Georgia in April is also common to both simulations, but the WGEN simulation overestimates values of μ for these stations by more than 180 mm and the GiSTR simulation overestimates by between 20 to 40 mm. The fact that the WGEN simulated values of μ have an error 350% larger than the GiSTR simulation

for these stations, it is possible that the RMSE of this month for the WGEN simulation (which > 200% larger than the average RMSE) is the result of the overestimation at these stations. This result is consistent with results from Section 3.2.3.1, which indicates that WGEN has a smaller RMSE than GiSTR when used outside of a downscaling context.

4.2.3.2. Variance of Daily Nonzero Precipitation (σ^2)

The results from Section 3.2.3.2 indicate that WGEN has a higher error for the variance of the daily nonzero precipitation (σ^2) than GiSTR outside of a downscaling context. When used for downscaling, WGEN still has higher error than GiSTR for values of σ^2 . The average RMSE of σ^2 for WGEN and GiSTR is 268.74 and 127.44 mm² respectively. Figure 4.28 compares the RMSE of each simulation across all months in the domain. In a similar result to the comparison for the RMSE of μ , the RMSE of each simulation for values of σ^2 is similar between simulations for all months, except September, March and July. Given that the RMSE for the WGEN simulation is up to 320% larger than the average RMSE for these months, it is possible that the RMSE for these months is the result of large overestimations or underestimations for individual stations. Figure 4.29 shows the difference between the simulated and observed values of σ^2 for the WGEN simulation for September. The difference between the generated and observed values of σ^2 in September shows that WGEN simulates values of σ^2 within 500mm² of the observed values. However, there are several stations on the Gulf Coast of Florida where WGEN overestimates by over 300 mm². The exact error of the WGEN simulation for each of these stations is over 3500 mm², while the GiSTR simulation overestimates values of σ^2 for the same stations in September by less than 500 mm². This particular tendency was also common to the months of March and July, suggesting that the RMSE of the WGEN simulation for these months is the result of unrealistic values of σ^2 for these stations. Given the result for WGEN in this section and the previous section, the RMSE of μ and σ^2 for the WGEN simulation may be improved by removing these outlying values.

4.2.3.3. Average Total Precipitation ($E[S(T)]$)

The results for the error of the average total precipitation ($E[S(T)]$) are similar to the results for the mean daily nonzero precipitation (μ). The WGEN simulation has the higher error for $E[S(T)]$ than the GiSTR simulation. The average RMSE of $E[S(T)]$ for the WGEN and GiSTR downscaling simulations is 79.73 and 65.73 mm respectively. The comparison of the RMSE between simulations across all months, shown in Figure 4.30, indicates that the RMSE of each generator is similar for most months. However, for April the RMSE of the WGEN simulation is 222.72 mm, while the GiSTR simulation has a RMSE less than 71.78 mm. Following the relationship between μ and $E[S(T)]$, described in Chapter 2 (Equation 2.7), the RMSE of $E[S(T)]$ for the WGEN simulation in April is related to the RMSE of μ for the same simulation. That is, the high error of μ for the WGEN simulation at individual stations also contributes the RMSE of $E[S(T)]$ for the month of April and the increase in the average RMSE. Therefore, removing those stations where the WGEN simulation produces unrealistically high values of μ will likely improve the error of the WGEN simulation for both μ and $E[S(T)]$. However, the current results for this parameter are in contrast to the results presented in Chapter 3, which indicate that WGEN had a lower error than GiSTR outside of a downscaling context.

4.2.3.4. Inter-annual Variability ($Var[S(T)]$)

Given the relationship between the inter-annual variability ($Var[S(T)]$), the mean daily nonzero precipitation (μ), and the variance of the daily nonzero precipitation (σ^2), the error comparison in this section shows that the error of the WGEN simulation is larger for $Var[S(T)]$ than for the GiSTR simulation. The average RMSE of WGEN and GiSTR is 11032.19 and 2314.74 mm² respectively. Figure 4.31 shows the RMSE of the WGEN and GiSTR simulations for every month. The comparison of these values indicates that each simulation has similar RMSE for all month except April and September. In April and September the RMSE for the WGEN simulation is over 137% larger than the average RMSE

for the WGEN simulation, and over 850% larger than the RMSE of the same months for the GiSTR simulation. In each of these months, the RMSE of μ and σ^2 for the WGEN simulation was also much larger than the average RMSE for that simulation. Given the direct relationship between these two parameters and $Var[S(T)]$, along with the earlier discussion regarding the error for individual stations, it is apparent that outlying error values for both μ and σ^2 caused the RMSE of $Var[S(T)]$ for the WGEN simulation to also be unrealistically large in April and October. Therefore, it is likely that removing these outlying stations for both μ and σ^2 will improve the RMSE of $Var[S(T)]$ for the WGEN simulation. Regardless of this result, the error comparison presented in this analysis shows a similar result to those presented in Chapter 3. While the error of each simulation for each of the amount parameters shows specific results, the two-sample T-test results, summarized in Table 4.25, indicates that there is no significant difference in the RMSE of each parameter between each generator.

4.2.4. Probability Distribution Function Analysis

While the RMSE of most parameters between each downscaling simulation is similar, the RMSE for most parameters increases between WGEN when not applied to downscaling (Chapter 3) and the WGEN downscaling simulation. This increase is what causes GiSTR and WGEN to have comparable accuracy when used for downscaling. However, the questions remains as to why the WGEN simulation errors for μ and σ^2 which are much larger for individual stations and months than for the GiSTR simulation and other stations and months. These particular results are related to the ability of the WGEN simulation to replicate the PDF for these stations. Recall from Section 4.2.3.1 that for stations in coastal Georgia, the value of μ was overestimated at these stations by more than 180 mm in April by the WGEN simulation, and by between 20 and 40 mm by the GiSTR simulation. Figure 4.32 shows the comparison between the PDF produced by the WGEN simulation and the observed PDF for station 097847 in Savannah, GA in April. The generated PDF for this station shows that the WGEN simulation failed to capture the observed PDF of nonzero precipitation for

this station in April and the value of μ at this station is overestimated by 240 mm. In contrast, this station was overestimated by 40 mm by the GiSTR simulation. This problem was common to both of these coastal Georgia stations and stations in northeast corner of the domain where values of μ are overestimated by up to 50 mm. A similar problem can also be shown for those stations where the WGEN simulation overestimates the value of σ^2 by over 3500 mm². Recall from Section 2.3.2 that the WGEN simulation shows this overestimate for stations in Florida, while the GiSTR simulation overestimates σ^2 by up to 500 mm² in the same region. Figure 4.33 shows the comparison between the PDF produced by the WGEN simulation and the observed PDF for station 080945 in Bradenton, FL in September. Examining the PDFs shows that it while it captures most precipitation that occurs, there are multiple instances where precipitation amounts are generated which are much larger than the observed maximum precipitation for this station. This pattern is common for PDF produced by the WGEN simulation for all the stations in Florida where σ^2 is overestimated by more than 3500 mm². This suggests that the extremely heavy precipitation produced by the WGEN downscale simulation for these stations causes values of σ^2 to be overestimated on the Gulf Coast of Florida.

The result of the failure to capture the PDF in these locations for April and September caused the RMSE for μ and σ^2 to be overestimated by the WGEN simulation. This also contributed to the RMSE of $E[S(T)]$ and $Var[S(T)]$ for WGEN simulation to be comparable to the GiSTR downscale simulation. Removing these stations will likely improve the domain average RMSE of μ and σ^2 for April and September respectively. However, in most other months there are no outliers in the error like in April or September, such as in the example for January for the error of μ and σ^2 shown in Figure 4.34. This suggests that the RMSE of WGEN when used for downscaling is still comparable the RMSE of GiSTR used for downscaling for the amount parameters without these outlier stations. To determine if removing the stations associated with these outliers will improve the results, an additional analysis showing the improvement for both simulations with these outliers removed, is shown in the next section.

4.2.5. Changes with Outliers Removed

For this analysis, those stations where the WGEN simulation overestimated μ by more than 180 mm or overestimated σ^2 by more than 3500 mm² are removed. The analysis shown in Section 4.2.3 is repeated to show any improvements and changes in both simulations with these outlier stations removed. Removing these outlying stations changes the result for values of μ and $E[S(T)]$. For these parameters, the WGEN simulation has smaller RMSE on average than GiSTR for μ and a larger RMSE on average for σ^2 . For values of μ , the average RMSE of WGEN and GiSTR is 5.05 and 5.13 mm respectively. In addition, for values of $E[S(T)]$, the average RMSE of WGEN and GiSTR is 68.37 and 63.38 mm respectively. The same seasonal trends shown in Section 2.3 for μ and $E[S(T)]$ are also present with the outliers removed. Figure 4.35 shows the RMSE comparison between simulations for values of μ with the outliers removed. In addition, Figure 4.36 shows the same RMSE comparison between simulations for values of $E[S(T)]$ with outliers removed. In both comparison plots, it is apparent that RMSE associated with each parameter for the WGEN simulation peaks in April, though with a smaller magnitude with the outliers removed. However, with the outliers removed, the WGEN simulation has smaller error for the mean daily nonzero precipitation (μ), which is similar to results for the comparison between generators from Section 3.2.3.1. However, given that the RMSE for the average total precipitation ($E[S(T)]$) is still smaller for the GiSTR simulation, which is in contrast to the generator comparison results in Section 3.2.3.3. While those stations where μ is overestimated by more than 180 mm were removed, those stations in the northeast corner of the domain where the WGEN simulation overestimated μ by between 30 and 50 mm are not removed. The error of these stations for μ , aggregated to the monthly parameter, $E[S(T)]$, like causes the average RMSE of the WGEN simulation to remain larger than the GiSTR simulation for this parameter.

For the remaining parameters, the variance of the daily nonzero precipitation (σ^2) and the inter-annual variability ($Var[S(T)]$), the error of the GiSTR simulation remains smaller than the WGEN simulation with the outlier stations removed. The average RMSE of σ^2 for

WGEN and GiSTR is 183.58 and 124.08 mm² respectively. The average RMSE of $Var[S(T)]$ for WGEN and GiSTR is 4233.94 and 2154.23 mm² respectively. Figure 4.37 shows the RMSE comparison between simulations with the outliers removed for σ^2 by month. Similarly, Figure 4.38 shows the RMSE comparison between simulations with the outliers removed for $Var[S(T)]$ by month. From both comparison plots, it is apparent that the unrealistic of the WGEN simulation associated with September and April is reduced with the outliers removed. Another interesting result of this analysis is the change in the significance testing. Testing the RMSE for each simulation and every amount parameter originally showed that there was no significant difference between parameters. This result is still consistent for μ and $E[S(T)]$, but the two-sample T-test results, summarized in Table 4.26, show that the RMSE for the WGEN simulation is significantly larger than the GiSTR simulation for both σ^2 and $Var[S(T)]$ when outliers are removed. However, it is also important to note that the GiSTR simulation still has lower error for both parameters, which is consistent with the results for the evaluation of each generator presented in Chapter 3. The issues with outlying stations highlight an important implication for using any of the weather generators or downscaling variations presented in this Chapter. While these variations and weather generators each have their own spatial and seasonal error patterns, the fact that precipitation amounts produced for individual stations can have these erroneous values shows the importance of noting where any generator or downscaling process suffers high error for the applications of local users. For instance, a drastic overestimation of precipitation amounts caused by this WGEN simulation would produce erroneous crop and streamflow modeling results, or overestimate the potential for above normal precipitation in a season without communicating the uncertainty of any downscaling process to the end user. However, the fact that these stations have such erroneous results for both simulations (though there is less error in GiSTR) is important for those users requiring forecasts at these stations. While the exact cause for the gross errors at these locations is not determined in this analysis, it is likely that either the LWR interpolation failed to estimate parameters at the CFSR grid point, or that the MOM estimation produced erroneous parameters for the gamma distribution, or that both failed to produce a realistic value of the input parameters at these

stations. In these cases, the capping value present in GiSTR may have prevented the error at these stations from reaching or exceeding the error produced by the WGEN simulation. Given that the downscaling variation used may have caused the result, this suggests that a different variation may change this result and that several variations may be necessary in a forecast situation in order to communicate to the end user the best variation for their area of interest.

4.2.6. Dry and Wet Spell Frequency Analysis

Given that removing the outliers, as shown in Section 4.2.5, did not affect the results associated with the temporal parameters, this section focuses on dry and wet spell frequencies associated with all stations in the domain. In Chapter 3 each generator is shown to replicate the dry and wet spell frequencies with very little difference between each generator and between each generator and the observations. However, the simulation of the dry and wet spell frequencies by each downscaling simulation in this section shows unique errors that are not shown in Chapter 3. Figure 4.39 show the observed frequency of wet spells of various lengths in January (a) compared to those produced by the WGEN simulation (a) and the GiSTR simulation (b) for all stations in the domain. For wet spells in January, the WGEN simulation underestimates the frequency of one day wet spells compared to the observations and the GiSTR simulation. Figure 4.40 shows the observed frequency of dry spells of various lengths in January (a) compared to those produced by the WGEN simulation (b) and the GiSTR simulation (c) for all stations in the domain. The WGEN simulation shows very little difference compared to the observations for dry spells in January, while the GiSTR simulation underestimates the frequency of one-day dry spells.

Figure 4.41 shows the observed frequency of wet spells of various lengths in July compared to those produced by the WGEN simulation (a) and the GiSTR simulation (b) for all stations in the domain. Similarly to the January wet spells, the WGEN downscale underestimates the frequency of one-day wet spells in July while the GiSTR downscale shows little difference from observed July wet spell frequencies. Figure 4.42 shows the

observed frequency of dry spells of various lengths in July (a) compared to those produced by the WGEN simulation (b) and the GiSTR simulation (c) for all stations in the domain. Similarly to the January dry spells, the WGEN downscale simulation shows very little difference from observations for July dry spells and the GiSTR downscale underestimates the frequency of one-day dry spells.

The underestimation of one day wet spells reflects the increased RMSE of P_{01} and P_{11} for the WGEN downscale shown in Section 4.2.1. In addition, the increased RMSE of P_{11} between the simulation of WGEN in Chapter 3 and the WGEN simulation in this section is also reflected in overestimation of the frequency of wet spells longer than ten days. All the results, with the exception of the evaluation of spatial structure show that when downscaling is applied, the difference between weather generators do not follow the same error patterns described in Chapter 3. This suggests that further evaluation using all the available combinations of the downscaling variations and weather generators is necessary to determine the exact combination which best captures precipitation in the Southeast U.S. While further analysis is still required, based on this research the GiSTR generator and the LWR-MOM variation have the lowest error for most parameters across the Southeast U.S.

4.3. Summary Discussion

4.3.1. Downscaling Simulation Comparison

Using GiSTR as the fixed generator, this evaluation focused primarily on the evaluation of various downscaling variations applied to the assumptions of the scaling relationship. The combinations were LWR-MOM, AW-DFIT, LWR-DFIT, and AW-MOM. The results of this analysis, summarized in Table 4.27, indicate that on average the RMSE of all parameters resulting from the other three variations of the downscaling assumptions are more accurate than the commonly used AW-MOM variation. Of the three combinations

remaining, the LWR-MOM variation best captures most parameters reflecting precipitation events and amounts.

While individual variations are best for specific parameters, tests for significance showed that there was no significant difference in the RMSE for the temporal parameters or the spatial structure between downscaling variations. However, for those parameters reflecting precipitation amounts (μ , σ^2 , $E[S(T)]$, and $Var[S(T)]$), the testing revealed that specific variations provide a significant improvement in the RMSE of these parameters across the domain. For μ the two-sample T-tests indicate that the variations using AW interpolation have significantly smaller RMSE than those using LWR interpolation. However, the testing also shows that for values of $E[S(T)]$ that the AW-DFIT variation has significantly smaller RMSE than those using the LWR interpolation. However, while the RMSE of $E[S(T)]$ for the AW-DFIT is smaller than the RMSE for the AW-MOM variation, there is no significant difference in the RMSE between the two simulations. Tests for significance on the variance of daily nonzero precipitation (σ^2) and the inter-annual variability ($Var[S(T)]$) show that there is no significant difference in the error of σ^2 between simulations and that LWR variations have a significantly larger error for $Var[S(T)]$ than the AW-MOM variation.

While there are patterns between downscaling simulations for the RMSE of various parameters, there are also patterns in the error of these parameters in the domain for each downscaling simulation. In the downscaling simulations using the LWR there is a tendency to overestimate values of μ by more than 7 mm in the northeast corner of the domain. This tendency is not shown by either of the downscaling simulations using the AW. The LWR as described by Wilks (2008) makes use of the elevation of both the stations and the elevation of each GCM grid point in the regression of each weather generator parameter, while the AW does not consider elevation. Given the resolution of CFSR it is possible that the elevation of the land surface model is too coarse to provide the appropriate elevation in the northeast corner of the domain. The northeast corner of the domain does have strong variations in elevation which may not be captured by the CFSR topography. However, this is likely not the only sources of error since the values of μ are within 7 mm of the observed values for

most months in the Appalachian Mountains from western Pennsylvania through Georgia. An additional source of error for values of μ may result from an edge effect on the LWR downscaling simulations in the northeast corner of the domain. However, this is somewhat unlikely since there are no similar errors shown on the edges or corners of the domain. Regardless of what caused these errors for μ in the LWR downscaling simulations, it is likely that considering an RMSE for individual states may show that the RMSE of μ for the LWR downscaling simulations is less than the average in the central portion of the domain. The tendency to underestimate values of the variance of daily nonzero precipitation is still present for much of the domain, with an apparent pattern of underestimating more for coastal stations during the peak of the Atlantic hurricane season. Each simulation has the largest underestimation on the north Gulf Coast and in southern Alabama, with values of σ^2 underestimated by over 500 mm². However, the LWR-MOM variation has the fewest number of stations where this occurs. The ultimate conclusion from this analysis is that the LWR-MOM variation provides the lowest error for most parameters of interest in weather generator downscaling, including the temporal parameters of precipitation occurrence, the spatial structure, and the variance of daily nonzero precipitation. As such, the LWR-MOM variation was applied to the weather generator comparison in the context of downscaling.

4.3.2. Weather Generator Comparison – Downscaling Context

In order to determine if the patterns between generators when used for downscaling are consistent with the patterns shown in Chapter 3, WGEN are also used for downscaling the CFSR data and compared to the GiSTR downscaling. Since the results of Section 1 show that LWR-MOM has the lowest error for most parameters, this is used as the constant variation for this analysis. However, while it is expected that the same patterns would hold between generators regardless of the downscaling context, the downscaling performed with each generator indicates several differences between the downscaling simulations and the results presented in Chapter 3 in the Southeast U.S. The results of this analysis, summarized in Table 4.28, show the best simulation for each parameter.

While the RMSE of the spatial structure parameters, σ^2 , and $Var[S(T)]$ reflect the results shown in Chapter 3 (which indicate that WGEN has larger error than GiSTR for each parameter), for almost all the remaining parameters in a downscaling context GiSTR shows smaller RMSE on average across the domain as compared to the WGEN simulation. For all the temporal parameters, the RMSE for WGEN increases when downscaling is applied, and becomes larger than the GiSTR simulation. The increase in the RMSE for WGEN for the temporal parameters is also reflected in the dry and wet spell frequencies across the domain. The overestimation of the frequency of wet spells longer than 8 days and the underestimation of the frequency of wet spells shorter than 2 days indicates there is a general overestimation of P_{11} across the domain. However, this only occurs when downscaling is applied, suggesting that for WGEN the chosen combination for the downscaling assumption has a strong influence on downscaling precipitation events with WGEN, while it has much less influence on GiSTR.

For μ and $E[S(T)]$, the GiSTR downscaling simulation has smaller RMSE as compared to WGEN. In Chapter 3, WGEN is shown to have smaller RMSE for both parameters. While the difference between each downscaling simulation for both parameters is not significant, the downscaling variation used also effects the values of both of these parameters. However, while the increase of the RMSE of the temporal parameters for the WGEN downscaling simulation is common across the domain, the RMSE of μ and $E[S(T)]$ for the WGEN downscaling simulation has been shown to be the result of the failure of WGEN to capture the PDF at two specific locations in April, which likely contributes to the RMSE for these parameters across the domain. However, it is unlikely that this alone is the only cause for the increased error. Given the previous indication of the overestimation of P_{11} by the WGEN downscaling simulation across the domain, a general increase in the RMSE of μ and $E[S(T)]$ could be the result of an increase in the number of days of precipitation generated in the domain. Therefore it is possible that the amount generation process is being called too often, resulting in an overestimation of μ and $E[S(T)]$. A common issue between the GiSTR and WGEN downscaling simulations is the overestimation of values of μ in the northeast corner of the domain. This is consistent with the evaluation of downscaling

combinations, since both downscaling simulations make use of LWR which has been shown to have this problem.

Another pattern observed was the change in significance of RMSE of σ^2 between the WGEN downscaling and control simulations as compared to analysis from Chapter 3. Recall, that the RMSE of σ^2 between the WGEN control simulation and the GiSTR control simulation was shown to not be significantly different. However, analysis of the downscaling has shown that WGEN has significantly larger RMSE of σ^2 than GiSTR. Similarly to the results for values of μ , there are several locations where the generated values of σ^2 are much larger than the average, particularly in September. These locations more accurately capture the observed values of μ , but also produce precipitation amounts far larger than the historical maximum precipitation amount at that location. The result is an overestimation of values of σ^2 at these locations, which affects the RMSE of these values throughout the domain. This in turn also affects the RMSE of $Var[S(T)]$. The comparison of $Var[S(T)]$ between the GiSTR and WGEN downscaling simulation is consistent with the results presented in Chapter 3, but the magnitude of the difference is much larger when downscaling is applied (37.08 mm² in Chapter 3 versus 8717.45 mm² in this chapter).

In an additional analysis, the stations where the WGEN simulation overestimated the mean daily nonzero precipitation (μ) by more than 180 mm or the variance of the daily nonzero precipitation (σ^2) was overestimated by more than 3500 mm² are removed. The result of this analysis improves the RMSE of all the precipitation amount parameters. However, removing the outliers only causes the RMSE of μ produced by both simulations to reflect the results shown in Chapter 3. That is, removing outlying stations from both simulations allows the WGEN simulation to have a smaller RMSE for μ than the GiSTR simulation, which is shown in Chapter 3. However, it was also shown for each simulation that there was no significant difference between simulations for each of the amount parameters.

While there are several clear patterns in these results, there is one common trend to all of these results. Regardless of the downscaling variation or generator used, the error of the downscaling process is larger than when each generator is with observations only. In

addition, while there are mixed results on which downscaling combination provides the highest accuracy for all the parameters described, LWR-MOM, LWR-DFIT, and AW-DFIT all have better accuracy for most parameters than the AW-MOM. AW-MOM is commonly found in literature as the general approach to downscaling with weather generators. While the AW-DFIT combination provides the most accurate values for mean daily nonzero precipitation (μ) and the average monthly total precipitation ($E[S(T)]$) across the Southeast U.S. in general, it is possible that the LWR/MOM and the LWR/DFIT may be more accurate in the central portion of the domain, since this would not be effected by the increase in error in the northeast corner. It is possible that this difference may be related to a combination of an edge effect and topography. These edge and topography effects are not likely primary causes of increased error. Regardless, it is likely that the error of the LWR-DFIT and LWR-MOM variations is improved with the stations in the northeast corner of the domain removed. While there are several clear patterns in the error of each downscaling combinations with regards to each parameter, the difference between the GiSTR and WGEN downscaling simulations suggest that the downscaling combination has an impact on the results of each generator. As such, future work in this area should be to explore all the possible combinations of the downscaling assumptions and weather generator to determine the best combination for providing accurate estimates of precipitation in different regions of the Southeast U.S. In addition, while the simulations are evaluated against the observations for 1979-2000 at each station, repeating this analysis with observations from the 1971-2000 time period used in Chapter 3 does not change the results in this Chapter. While the weather generator evaluation could be performed for the 1979-2000 time period, the 1971-2000 time period is used in Chapter 3 since this 30 year period reflects a standard period for determining normal climatology. Therefore this period is used to determine if each generator can replicate the climate in this standard period, while allowing the 2001-2009 period to be used for analyzing the potential error with regards to forecast downscaling weather generators. In order to determine if the trends shown in this chapter are similar regardless of the time period, the next chapter presents a similar analysis to this chapter using CFSR as forecast for the period 2001-2009.

Table 4.1. RMSE Comparison for each downscaling simulation across the Southeast U.S. for value of P_{0I} .

Month	LWR-MOM	AW-DFIT	LWR-DFIT	AW-MOM
1	0.070	0.075	0.067	0.065
2	0.050	0.048	0.044	0.055
3	0.062	0.067	0.060	0.071
4	0.040	0.041	0.044	0.045
5	0.057	0.056	0.066	0.063
6	0.056	0.059	0.060	0.060
7	0.055	0.054	0.053	0.053
8	0.060	0.054	0.052	0.055
9	0.067	0.076	0.066	0.070
10	0.059	0.075	0.069	0.069
11	0.063	0.060	0.057	0.057
12	0.076	0.069	0.074	0.079
average	0.059	0.061	0.059	0.062

Table 4.2. RMSE Comparison for each downscaling simulation across the Southeast U.S. for value of P_{1I} .

Month	LWR-MOM	AW-DFIT	LWR-DFIT	AW-MOM
1	0.057	0.062	0.061	0.065
2	0.068	0.066	0.064	0.067
3	0.062	0.055	0.059	0.060
4	0.066	0.072	0.071	0.072
5	0.063	0.057	0.065	0.061
6	0.067	0.071	0.066	0.064
7	0.072	0.071	0.072	0.076
8	0.055	0.055	0.057	0.059
9	0.069	0.064	0.069	0.077
10	0.081	0.071	0.081	0.075
11	0.058	0.056	0.061	0.054
12	0.062	0.054	0.066	0.059
average	0.065	0.063	0.066	0.066

Table 4.3. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the Markov Transition Probabilities (P_{01} and P_{11}). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second.

Simulations compared		P_{01}		P_{11}	
		Difference in Average RMSE	P-value	Difference in Average RMSE	P-value
LWR-MOM	AW-DFIT	-0.0018	0.34	0.0022	0.246
LWR-MOM	LWR-DFIT	0.00027	0.475	-0.0010	0.36
LWR-MOM	AW-MOM	-0.0024	0.27	-0.00087	0.39
AW-DFIT	LWR-DFIT	0.0020	0.32	-0.0032	0.14
AW-DFIT	AW-MOM	-0.00066	0.44	-0.0030	0.16
LWR-DFIT	AW-MOM	-0.0027	0.25	0.00017	0.47

Table 4.4. Two-sample T-test results comparing each downscaling simulation to the Control for the Markov Transition Probabilities. The difference in the average RMSE is the Simulation – Control. P-values in red indicate that RMSE for that simulation is significantly larger than the Control.

Simulation used	P_{01}		P_{11}	
	Difference in Average RMSE	P-value	Difference in Average RMSE	P-value
LWR-MOM	-0.0053	0.25	0.0029	0.48
AW-DFIT	-0.0036	0.20	0.00077	0.39
LWR-DFIT	-0.0056	0.073	0.0040	0.073
AW-MOM	-0.0029	0.22	0.0038	0.098

Table 4.5. RMSE Comparison for each downscaling simulation across the Southeast U.S. for value of the unconditional probability of rain (π).

month	LWR-MOM	AW-DFIT	LWR-DFIT	AW-MOM
1	0.046	0.045	0.045	0.042
2	0.023	0.021	0.023	0.030
3	0.040	0.040	0.038	0.045
4	0.024	0.029	0.024	0.024
5	0.032	0.032	0.041	0.034
6	0.025	0.027	0.027	0.030
7	0.023	0.023	0.025	0.024
8	0.038	0.030	0.030	0.034
9	0.038	0.046	0.039	0.040
10	0.034	0.055	0.049	0.051
11	0.035	0.036	0.041	0.037
12	0.054	0.046	0.051	0.051
average	0.034	0.036	0.036	0.037

Table 4.6. RMSE Comparison for each downscaling simulation across the Southeast U.S. for value of persistence (γ).

month	LWR-MOM	AW-DFIT	LWR-DFIT	AW-MOM
1	0.10	0.12	0.10	0.11
2	0.11	0.11	0.099	0.11
3	0.11	0.10	0.10	0.11
4	0.097	0.10	0.11	0.11
5	0.11	0.10	0.11	0.11
6	0.11	0.12	0.12	0.12
7	0.12	0.12	0.12	0.12
8	0.10	0.10	0.10	0.10
9	0.13	0.13	0.13	0.14
10	0.13	0.12	0.13	0.13
11	0.11	0.11	0.099	0.097
12	0.11	0.11	0.12	0.12
average	0.11	0.11	0.11	0.11

Table 4.7. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the unconditional probability of rain (π) and persistence (γ). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second.

Simulations compared		π		γ	
		Difference in Average RMSE	P-value	Difference in Average RMSE	P-value
LWR-MOM	AW-DFIT	-0.0016	0.35	-3.5E-05	0.50
LWR-MOM	LWR-DFIT	-0.0018	0.33	0.00089	0.42
LWR-MOM	AW-MOM	-0.0026	0.26	-0.0025	0.29
AW-DFIT	LWR-DFIT	-0.00024	0.48	0.00092	0.42
AW-DFIT	AW-MOM	-0.00097	0.41	-0.0025	0.29
LWR-DFIT	AW-MOM	-0.00073	0.43	-0.0034	0.23

Table 4.8. Two-sample T-test results comparing each downscaling simulation to the Control for the unconditional probability of rain (π) and persistence (γ). The difference in the average RMSE is the Simulation – Control. P-values in red indicate that RMSE for that simulation is significantly less than the Control.

Simulation used	π		γ	
	Difference in Average RMSE	P-value	Difference in Average RMSE	P-value
LWR-MOM	-0.0035	0.43	-0.0026	0.23
AW-DFIT	-0.0019	0.34	-0.0026	0.28
LWR-DFIT	-0.0017	0.35	-0.0035	0.22
AW-MOM	-0.00091	0.41	-6E-05	0.49

Table 4.9. Average RMSE for each downscaling simulation for each correlation matrix.

Parameter	LWR-MOM	AW-DFIT	LWR-DFIT	AW-MOM
ρ	0.13	0.14	0.14	0.13
ρ_{ev}	0.089	0.093	0.091	0.092
ρ_{am}	0.13	0.15	0.14	0.14
ρ_{ex}	0.10	0.11	0.11	0.11

Table 4.10. Two-sample T-test results comparing each downscaling simulation to the Control for the correlation matrix of precipitation events (ρ_{ev}). The difference in the average RMSE is the Simulation – Control. P-values in red indicate that RMSE for that simulation is significantly larger than the Control.

Simulation Used	Difference in Average RMSE	P-value
LWR-MOM	0.016	0.0011
AW-DFIT	0.020	0.00024
LWR-DFIT	0.018	0.00078
AW-MOM	0.019	0.00029

Table 4.11. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the Markov transition probabilities for extreme events. P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second simulation.

Simulations compared		P_{01}		P_{11}	
		Difference in Average RMSE	P-value	Difference in Average RMSE	P-value
LWR-MOM	AW-DFIT	-0.073	7.07E-06	-0.089	5.35E-06
LWR-MOM	LWR-DFIT	-0.071	1.09E-05	-0.092	2.93E-06
LWR-MOM	AW-MOM	-0.0046	0.38	-0.0026	0.44
AW-DFIT	LWR-DFIT	0.0019	0.42	-0.0037	0.32
AW-DFIT	AW-MOM	0.068	8.68E-06	0.086	7.54E-06
LWR-DFIT	AW-MOM	0.066	1.13E-05	0.090	4.21E-06

Table 4.12. Two-sample T-test results comparing each downscaling simulation to the Control for the correlation matrix for precipitation extreme events (ρ_{ex}). The difference in the average RMSE is the Simulation – Control. P-values in red indicate that RMSE for that simulation is significantly larger than the Control.

Simulation Used	Difference in Average RMSE	P-value
LWR-MOM	-0.0012	0.46
AW-DFIT	0.0085	0.24
LWR-DFIT	0.0056	0.32
AW-MOM	0.0018	0.44

Table 4.13. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the mean daily nonzero precipitation (μ). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second.

Simulations compared		Difference in Average RMSE (mm)	P-value
LWR-MOM	AW-DFIT	1.57	0.0046
LWR-MOM	LWR-DFIT	0.11	0.43
LWR-MOM	AW-MOM	1.144	0.035
AW-DFIT	LWR-DFIT	-1.47	0.0066
AW-DFIT	AW-MOM	-0.43	0.19
LWR-DFIT	AW-MOM	1.03	0.047

Table 4.14. Summary of Two-sample T-test results comparing downscaling simulations to the Control for the RMSE of the mean daily nonzero precipitation (μ). P-values in red indicate the simulation has a significantly larger RMSE than the Control.

Simulation Used	Difference in Average RMSE (mm)	P-value
LWR-MOM	2.86	0.047
AW-DFIT	1.29	0.00073
LWR-DFIT	2.75	2.83E-05
AW-MOM	1.72	0.00045

Table 4.15. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the variance of daily nonzero precipitation (σ^2). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second.

Simulations compared		Difference in Average RMSE (mm ²)	P-value
LWR-MOM	AW-DFIT	-22.84	0.14
LWR-MOM	LWR-DFIT	-2.79	0.44
LWR-MOM	AW-MOM	0.14	0.50
AW-DFIT	LWR-DFIT	20.04	0.17
AW-DFIT	AW-MOM	22.98	0.16
LWR-DFIT	AW-MOM	2.944	0.44

Table 4.16. Summary of Two-sample T-test results comparing downscaling simulations to the Control for the RMSE of the variance of daily nonzero precipitation (σ^2). P-values in red indicate that either that the simulation has a significantly larger RMSE than the Control.

Simulation Used	Difference in Average RMSE (mm ²)	P-value
LWR-MOM	46.26	0.44
AW-DFIT	69.10	0.0011
LWR-DFIT	49.06	0.0065
AW-MOM	46.12	0.016

Table 4.17. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the average total precipitation ($E[S(T)]$). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second.

Simulations compared		Difference in Average RMSE (mm)	P-value
LWR-MOM	AW-DFIT	21.26	0.015
LWR-MOM	LWR-DFIT	1.13	0.45
LWR-MOM	AW-MOM	13.23	0.086
AW-DFIT	LWR-DFIT	-20.14	0.014
AW-DFIT	AW-MOM	-8.03	0.17
LWR-DFIT	AW-MOM	12.11	0.094

Table 4.18. Summary of Two-sample T-test results comparing downscaling simulations to the Control for the RMSE of the average total precipitation ($E[S(T)]$). P-values in red indicate that either that the simulation has a significantly larger RMSE than the Control.

Simulation Used	Difference in Average RMSE (mm)	P-value
LWR-MOM	26.024	0.092
AW-DFIT	4.76	0.22
LWR-DFIT	24.90	0.0013
AW-MOM	12.79	0.031

Table 4.19. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the inter-annual variability ($Var[S(T)]$). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second.

Simulations compared		Difference in Average RMSE (mm^2)	P-value
LWR-MOM	AW-DFIT	378.36	0.055
LWR-MOM	LWR-DFIT	-6.38	0.49
LWR-MOM	AW-MOM	467.84	0.033
AW-DFIT	LWR-DFIT	-384.73	0.052
AW-DFIT	AW-MOM	89.49	0.29
LWR-DFIT	AW-MOM	474.22	0.032

Table 4.20. Summary of Two-sample T-test results comparing downscaling simulations to the Control for the RMSE of the inter-annual variability ($Var[S(T)]$). P-values in red indicate that either that the simulation has a significantly larger RMSE than the Control.

Simulation Used	Difference in Average RMSE (mm^2)	P-value
LWR-MOM	1477.85	0.032
AW-DFIT	1099.50	1.69E-08
LWR-DFIT	1484.23	3.97E-06
AW-MOM	1010.01	1.7E-06

Table 4.21. Comparison of the RMSE of Markov Transition probabilities (P_{01} and P_{11}) between the GiSTR simulation and the WGEN simulation.

Month	P_{01}		P_{11}	
	GiSTR	WGEN	GiSTR	WGEN
1	0.070	0.076	0.057	0.28
2	0.050	0.081	0.068	0.28
3	0.062	0.069	0.062	0.28
4	0.040	0.051	0.066	0.23
5	0.057	0.043	0.063	0.25
6	0.056	0.067	0.067	0.26
7	0.055	0.051	0.072	0.28
8	0.060	0.065	0.055	0.30
9	0.067	0.077	0.069	0.26
10	0.059	0.047	0.081	0.24
11	0.063	0.047	0.058	0.24
12	0.076	0.067	0.062	0.27
average	0.059	0.062	0.065	0.27

Table 4.22. Comparison of the RMSE of the unconditional probability of rain (π) and persistence (γ) between the GiSTR simulation and the WGEN simulation.

Month	π		γ	
	GiSTR	WGEN	GiSTR	WGEN
1	0.046	0.23	0.10	0.23
2	0.023	0.25	0.11	0.22
3	0.039	0.22	0.11	0.23
4	0.024	0.17	0.097	0.22
5	0.032	0.18	0.11	0.23
6	0.025	0.20	0.11	0.25
7	0.023	0.22	0.12	0.26
8	0.038	0.23	0.10	0.28
9	0.038	0.21	0.13	0.24
10	0.034	0.16	0.13	0.22
11	0.035	0.18	0.11	0.22
12	0.054	0.21	0.11	0.23
average	0.034	0.20	0.11	0.24

Table 4.23. Summary of the two-sample T-test results for Markov Transition probabilities (P_{01} and P_{11}), the unconditional probability of rain (π), and persistence (γ). P-values in red indicate that the RMSE of that parameter for WGEN is significantly larger than the RMSE of that parameter for GiSTR.

Parameter	Difference in Average RMSE (WGEN - GiSTR)	P-value
P_{01}	0.0024	0.31
P_{11}	0.20	8.31E-15
π	0.17	8.14E-12
γ	0.12	2.04E-13

Table 4.24. Summary of two-sample T-test results for the RMSE of the correlation matrix of precipitation (ρ), the correlation matrix of precipitation events(ρ_{ev}), the correlation matrix of precipitation amounts(ρ_{am}), and the correlation matrix of precipitation extreme events (ρ_{ex}). P-values in red indicate that the RMSE of that parameter for WGEN is significantly larger than the RMSE for GiSTR.

Parameter	Difference in Average RMSE (WGEN - GiSTR)	P-value
ρ	0.13	0.00014
ρ_{ev}	0.21	5.75E-08
ρ_{am}	0.10	7.04E-05
ρ_{ex}	0.051	0.0040

Table 4.25. Summary of the two-sample T-test results for the mean daily nonzero precipitation (μ), the variance of the daily nonzero precipitation (σ^2), the average total precipitation ($E[S(T)]$), and the inter-annual variability($Var[S(T)]$). P-values in red indicate that the RMSE of that parameter for WGEN is significantly larger than the RMSE of that parameter for GiSTR.

Parameter	Difference in Average RMSE (WGEN - GiSTR)	P-value
μ	0.58 mm	0.34
σ^2	141.30 mm ²	0.055
$E[S(T)]$	13.99 mm	0.21
$Var[S(T)]$	8717.45 mm ²	0.072

Table 4.26. Summary of the two-sample T-test results for the mean daily nonzero precipitation (μ), the variance of the daily nonzero precipitation (σ^2), the average total precipitation ($E[S(T)]$), and the inter-annual variability($Var[S(T)]$). after removing outlier stations. P-values in red indicate that the RMSE of that parameter for WGEN is significantly larger than the RMSE of that parameter for GiSTR.

Parameter	Difference in Average RMSE (WGEN - GiSTR)	P-value
μ	-0.083 mm	0.46
σ^2	59.50 mm ²	0.0053
$E[S(T)]$	4.99 mm	0.33
$Var[S(T)]$	2079.71 mm ²	0.0095

Table 4.27. Summary of Best Downscaling Variation on Average for the different parameters evaluated.

Parameter	Best Downscaling Variation
Probability of a wet day following a dry day (P_{01})	LWR/MOM and LWR/DFIT
Probability of a wet day following a wet day (P_{11})	AW/DFIT
Unconditional Probability of Rain (π)	LWR/MOM
Persistence (γ)	LWR/MOM
Precipitation correlation matrix (ρ)	LWR/MOM
Precipitation event correlation matrix (ρ_{ev})	LWR/MOM
Precipitation amount correlation matrix (ρ_{am})	LWR/MOM
Precipitation extreme event correlation matrix (ρ_{ex})	LWR/MOM
Mean daily nonzero precipitation amounts (μ)	AW/DFIT
Variance of daily nonzero precipitation amounts (σ^2)	LWR/MOM
Average monthly total precipitation ($E[S(T)]$)	AW/DFIT
Variance of the monthly total precipitation / Inter-annual Variability ($Var[S(T)]$)	AW/MOM

Table 4.28. Comparison Table between the weather generators (full details in Chapter 3) and downscaling simulations in the Southeast U.S. Comparison includes GiSTR and WGEN only.

Parameter	Best weather generator	Best downscaling simulation	Best downscaling simulation (no outliers)
Probability of a wet day following a dry day (P_{01})	WGEN	both	both
Probability of a wet day following a wet day (P_{11})	WGEN	GiSTR	GiSTR
Unconditional Probability of Rain (π)	WGEN	GiSTR	GiSTR
Persistence (γ)	WGEN	GiSTR	GiSTR
Precipitation correlation matrix (ρ)	GiSTR	GiSTR	GiSTR
Precipitation event correlation matrix (ρ_{ev})	GiSTR	GiSTR	GiSTR
Precipitation amount correlation matrix (ρ_{am})	GiSTR	GiSTR	GiSTR
Precipitation extreme event correlation matrix (ρ_{ex})	GiSTR	GiSTR	GiSTR
Mean daily nonzero precipitation amounts (μ)	WGEN	GiSTR	WGEN
Variance of daily nonzero precipitation amounts (σ^2)	GiSTR	GiSTR	GiSTR
Average monthly total precipitation ($E[S(T)]$)	WGEN	GiSTR	GiSTR
Variance of the monthly total precipitation / Inter-annual Variability ($Var[S(T)]$)	GiSTR	GiSTR	WGEN

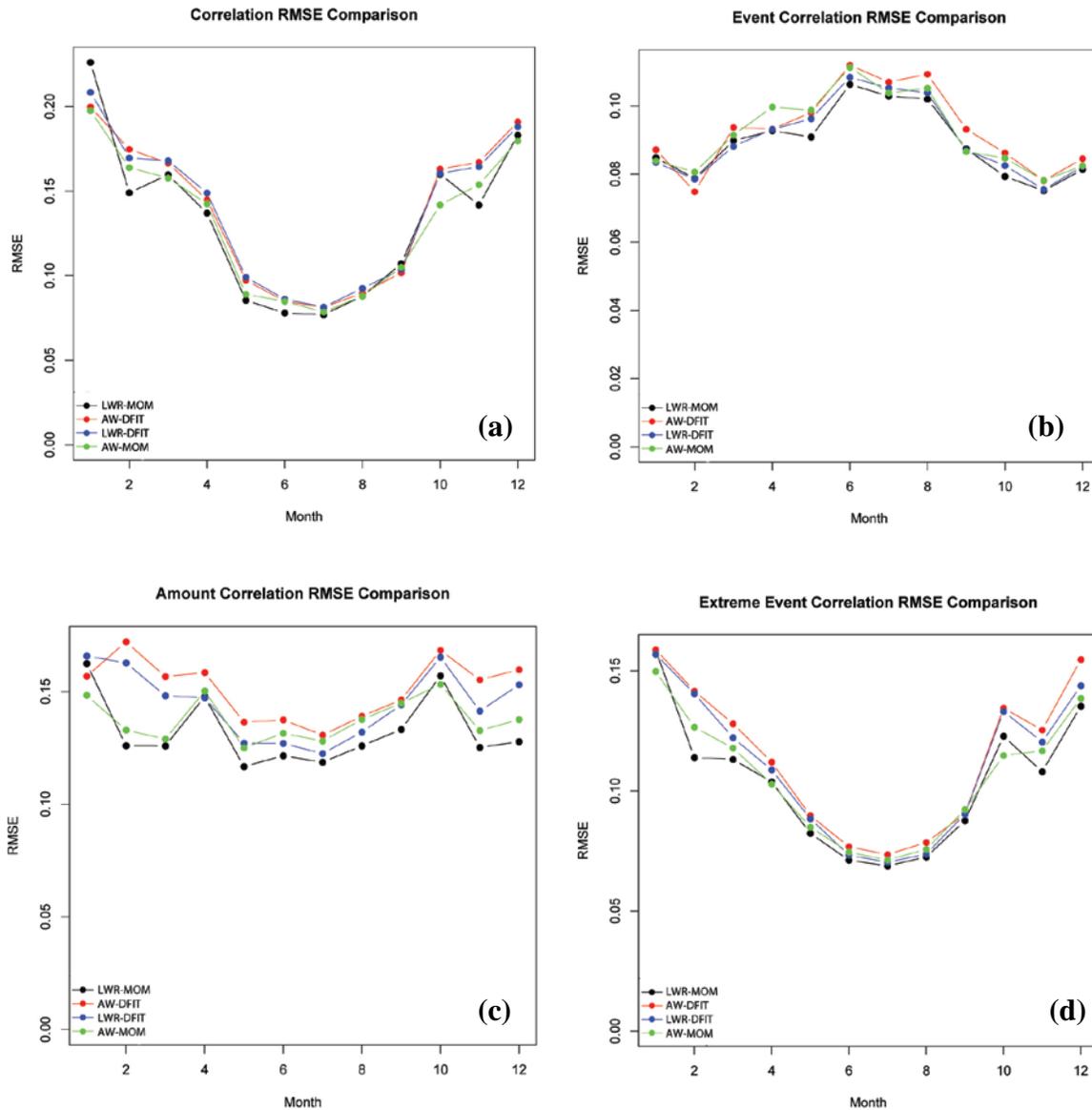


Figure 4.1. RMSE of each downscaling simulation by month for the correlation matrix of precipitation (ρ) (a), the correlation matrix of precipitation events (ρ_{ev}) (b), the correlation matrix of precipitation amounts (ρ_{am}) (c), and the correlation matrix of precipitation extreme events (ρ_{ex}) (d).

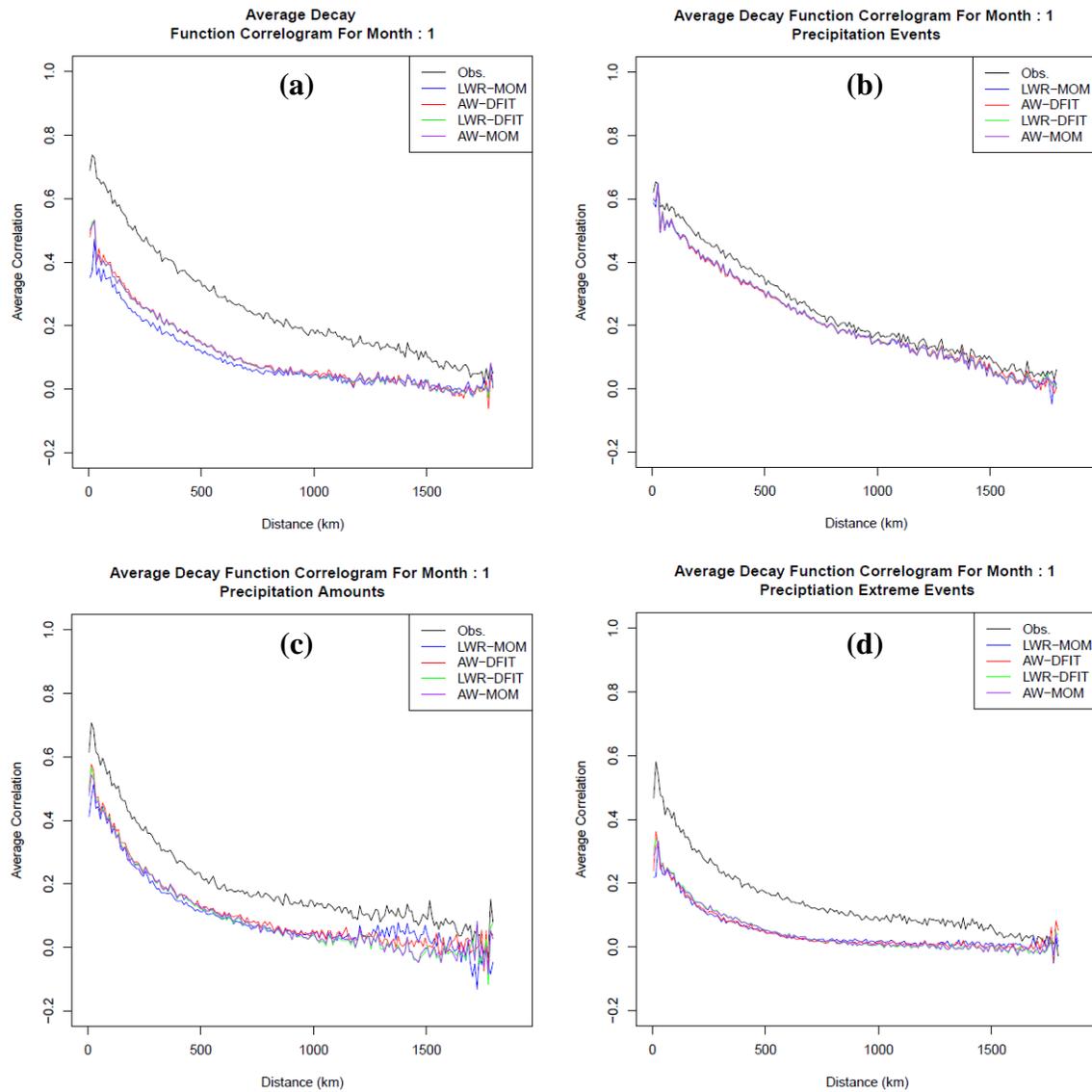


Figure 4.2. Average Decay Function Correlogram for each simulation and observations for January for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 1979-2000 time period.

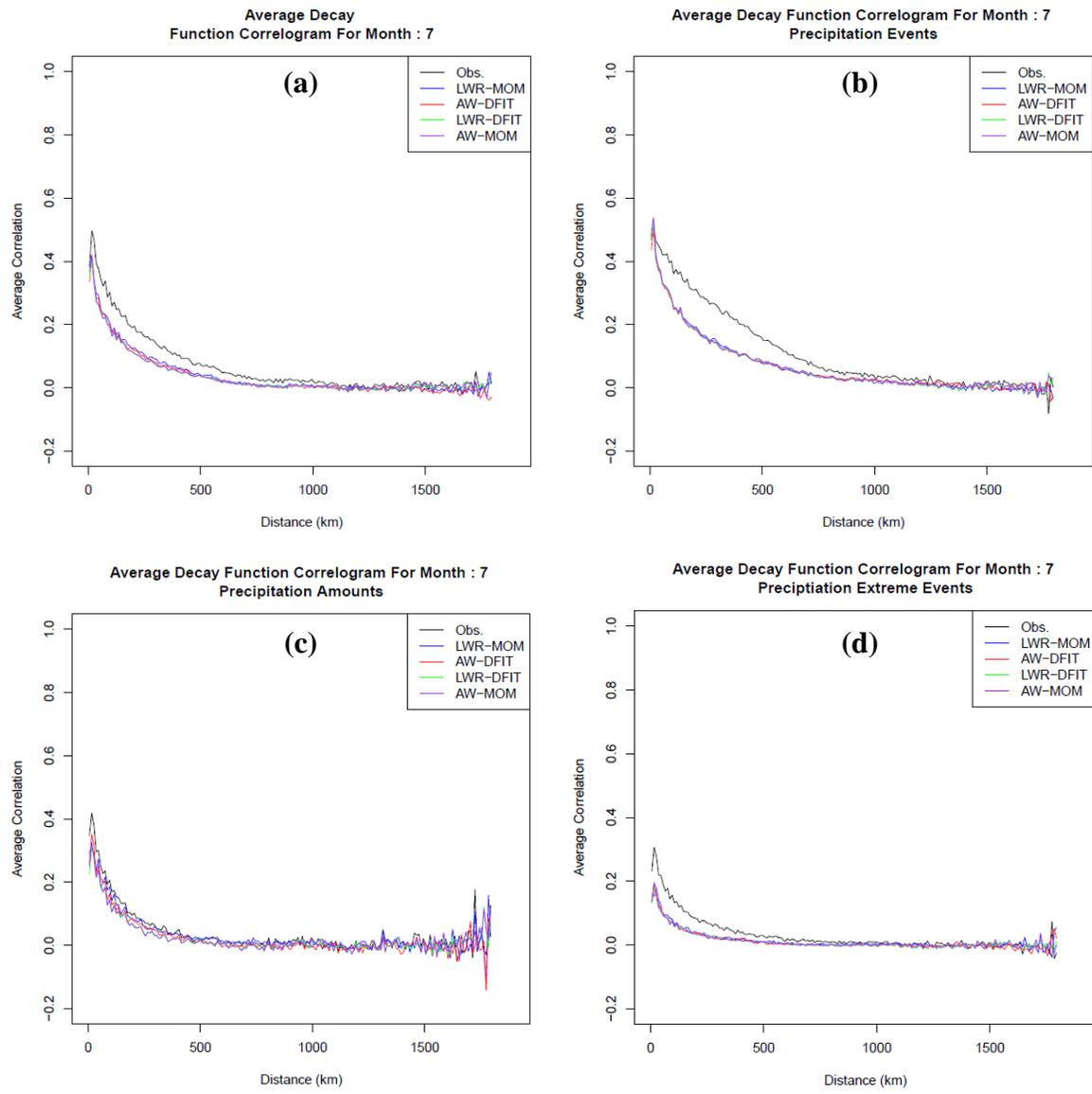


Figure 4.3. Average Decay Function Correlogram for each simulation and observations for July for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 1979-2000 time period.

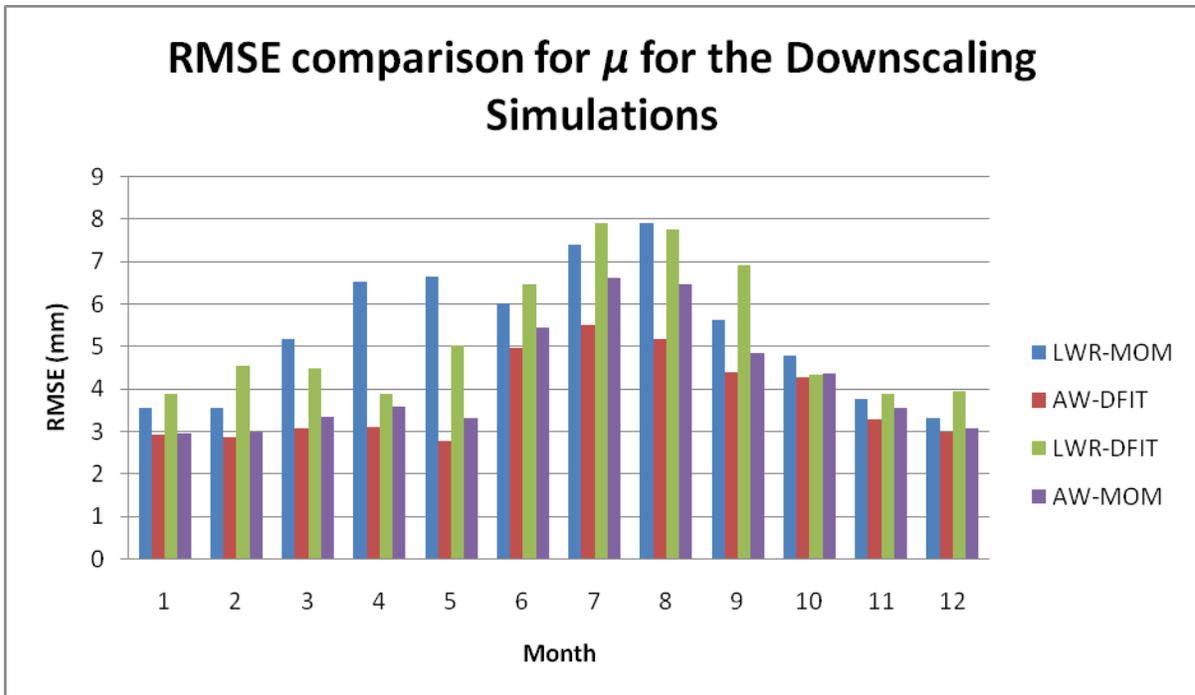
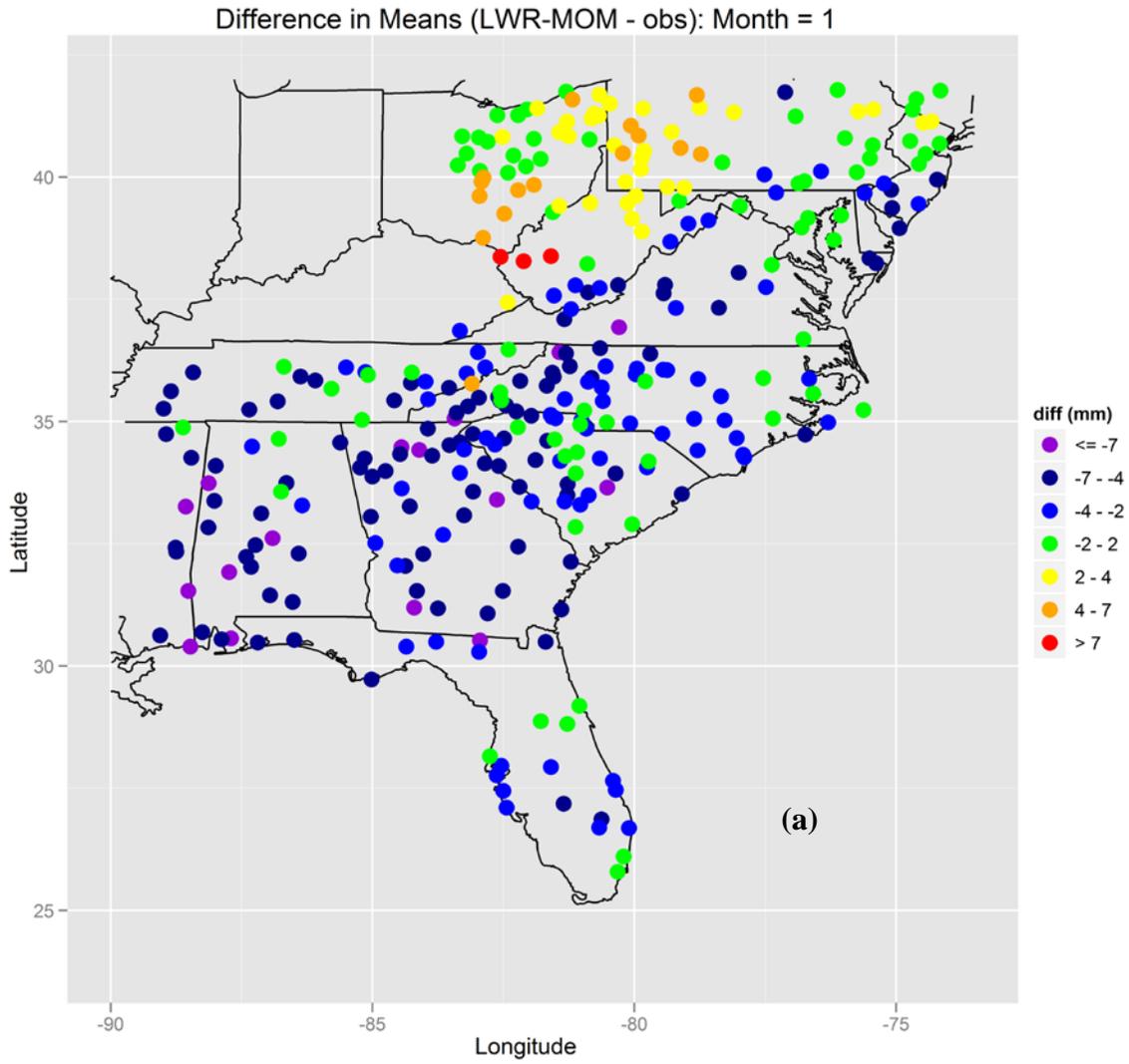
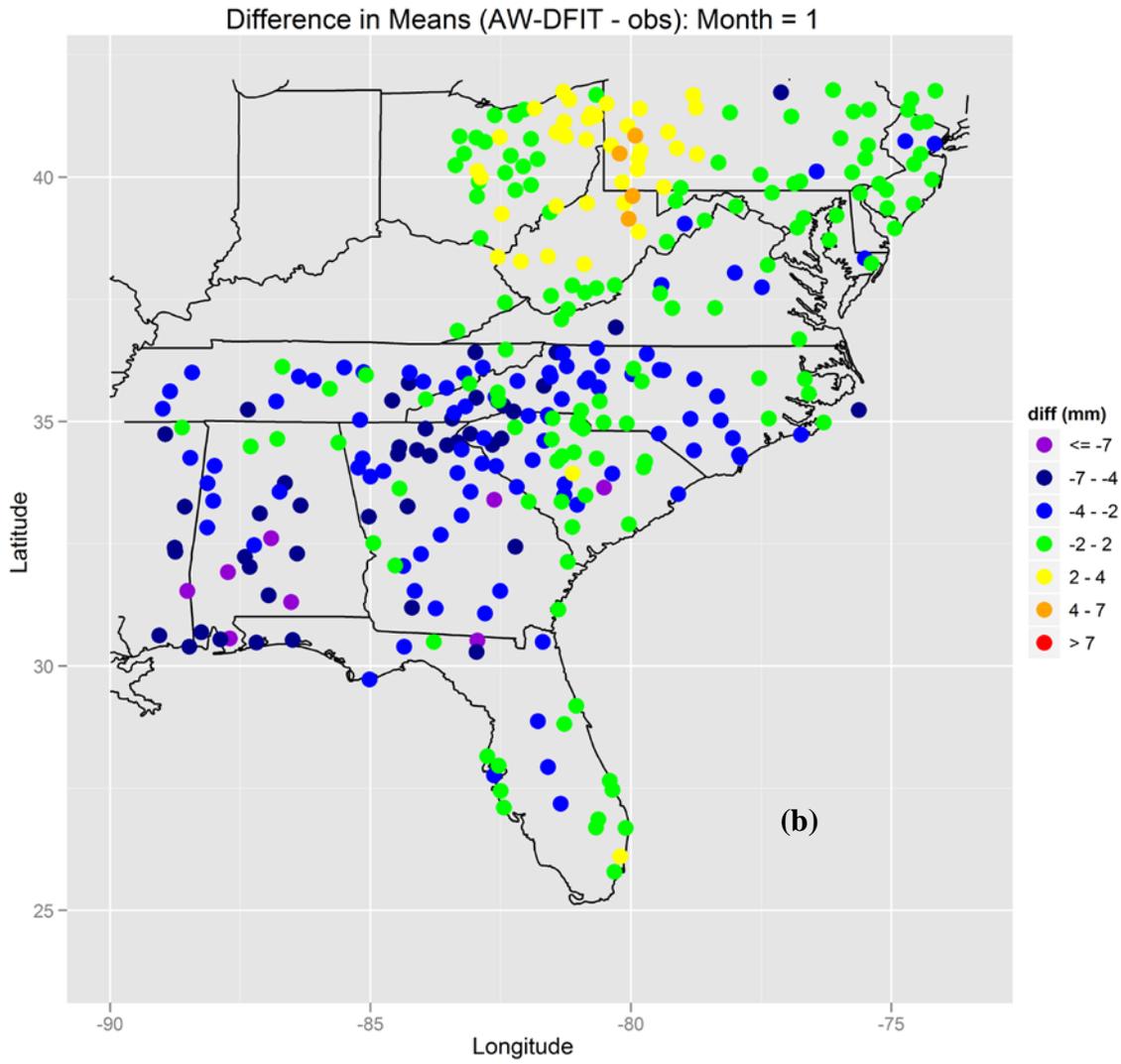
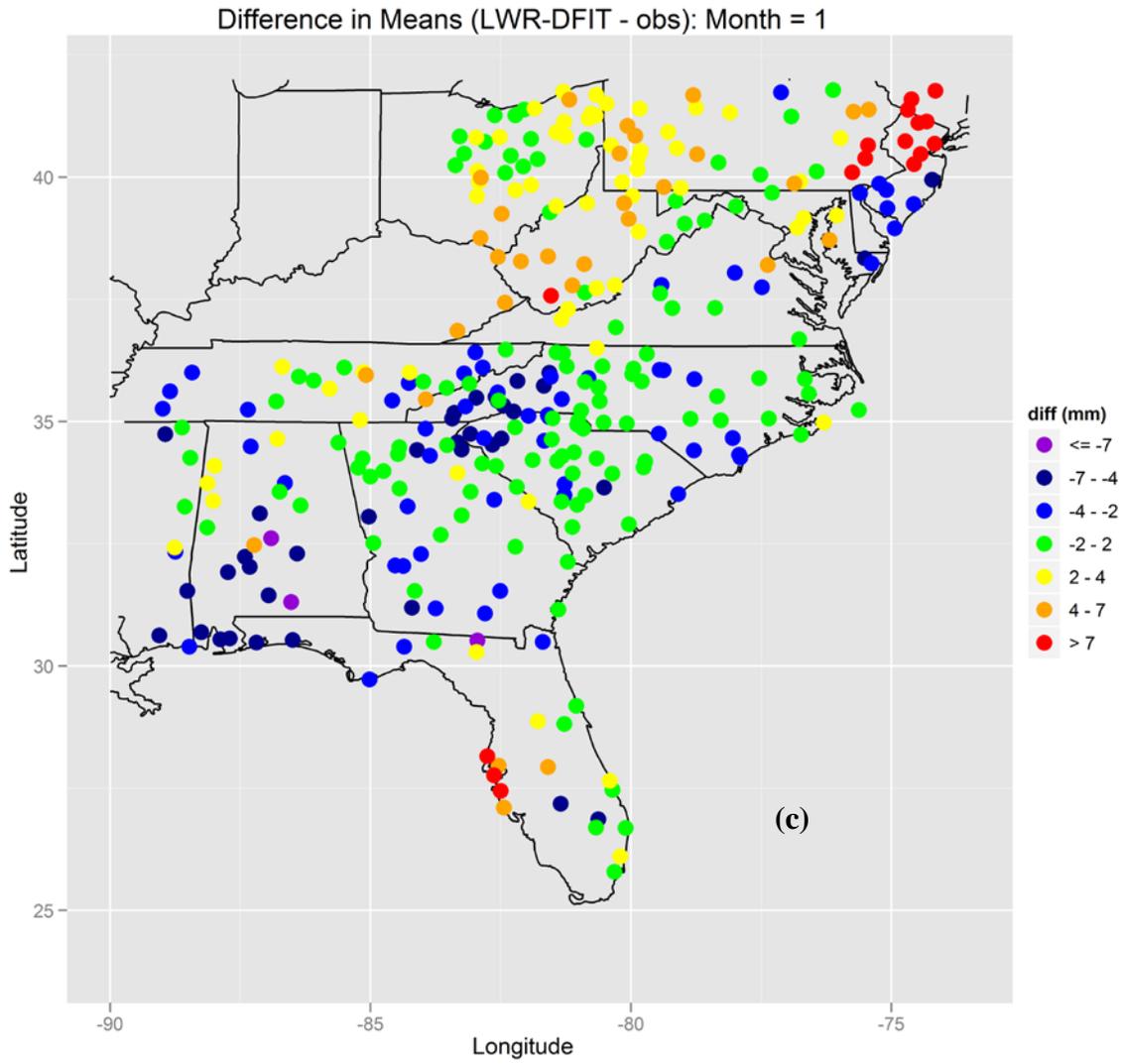


Figure 4.4. RMSE Comparison between downscaling simulations for mean daily nonzero precipitation (μ).

Figure 4.5. Difference between generated and observed values of mean daily nonzero precipitation (μ) for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for January.







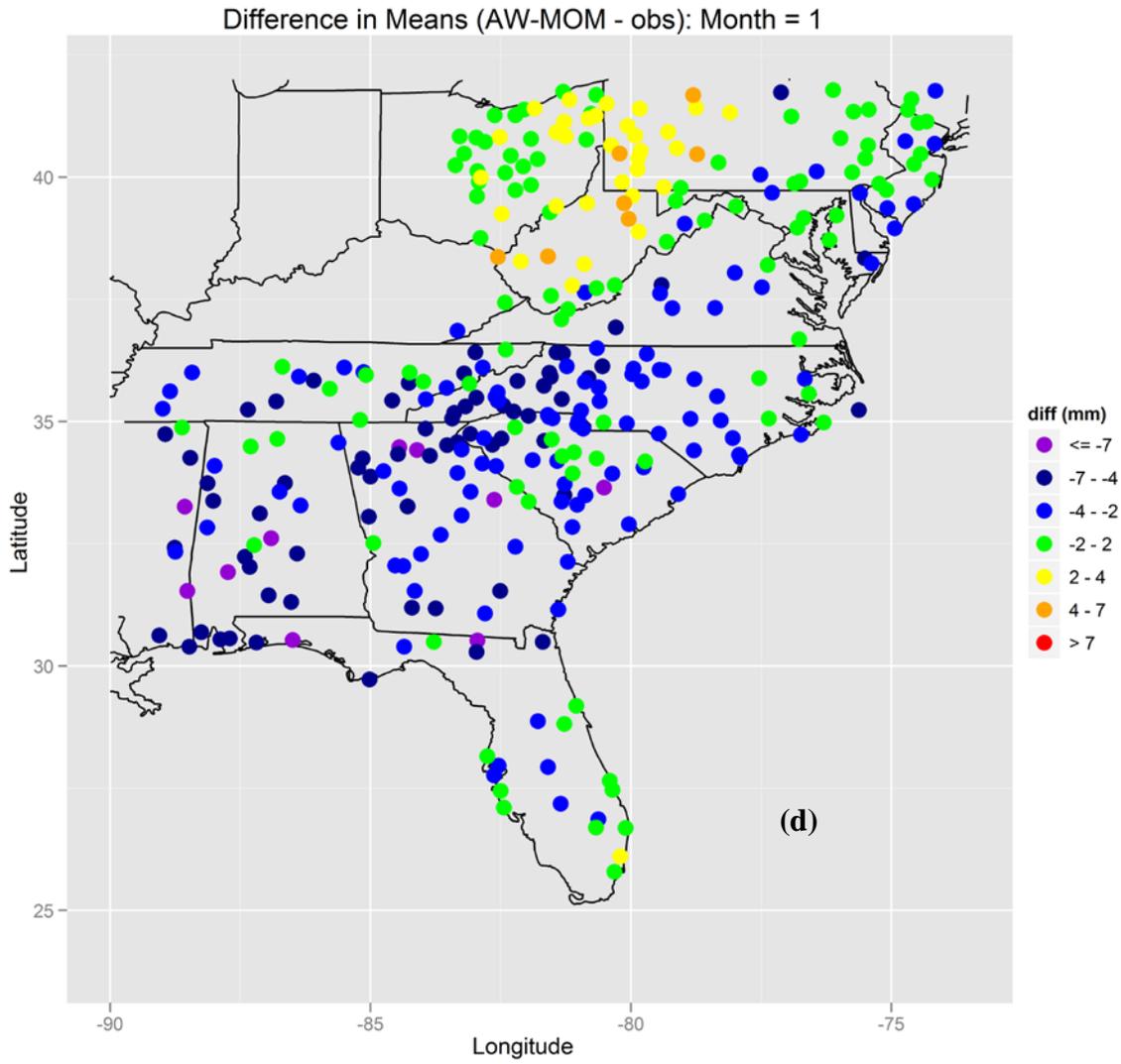
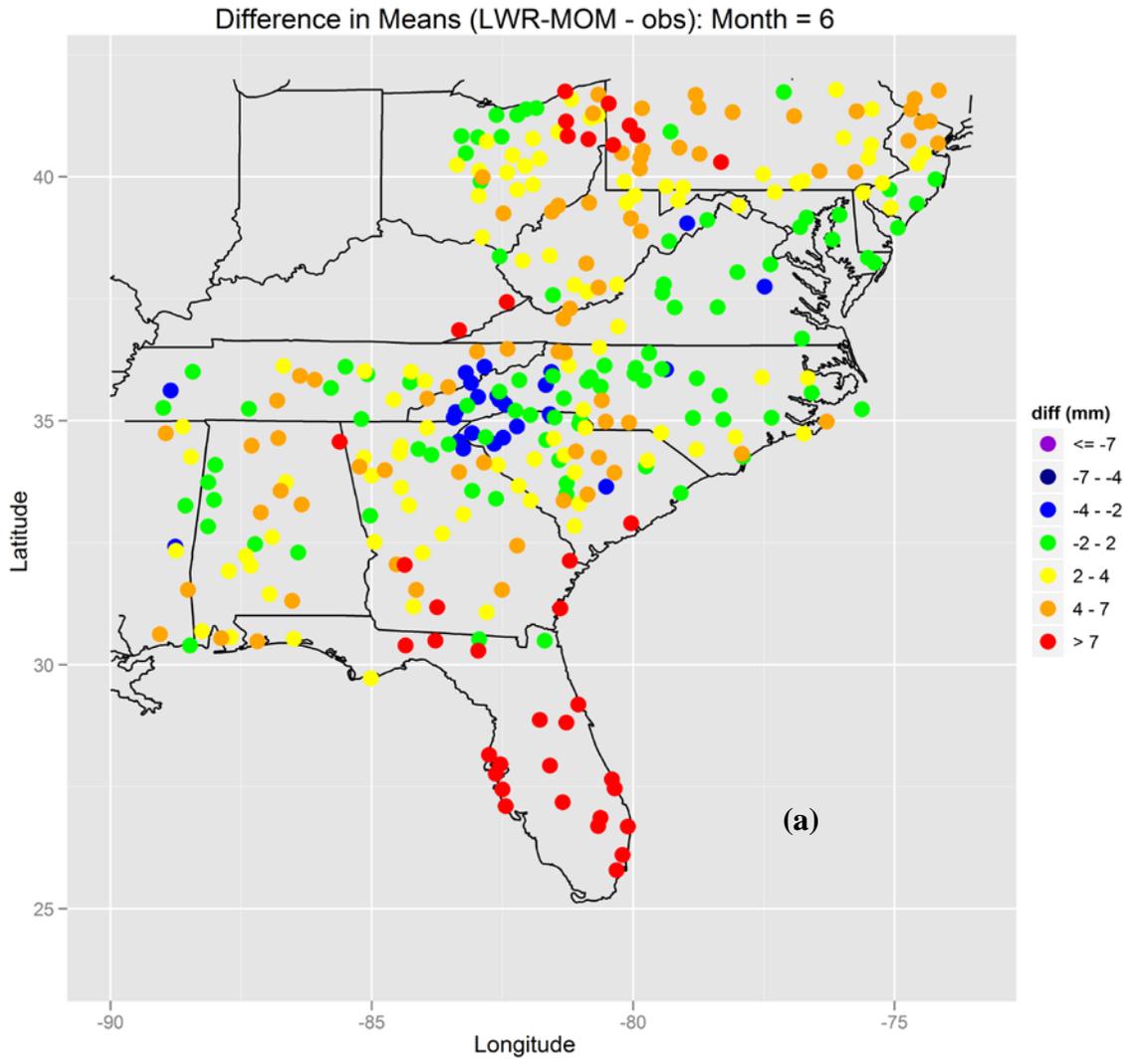
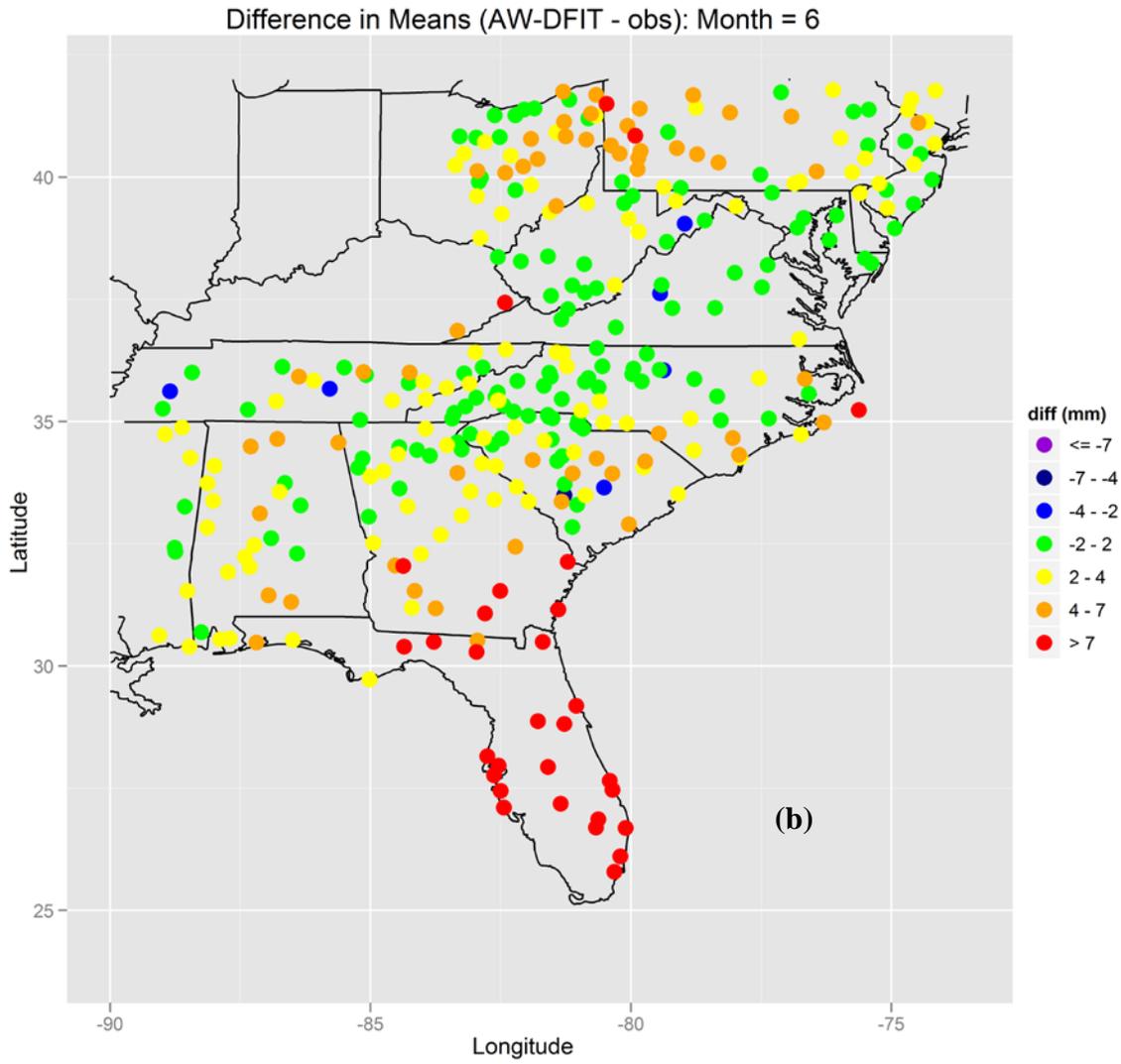
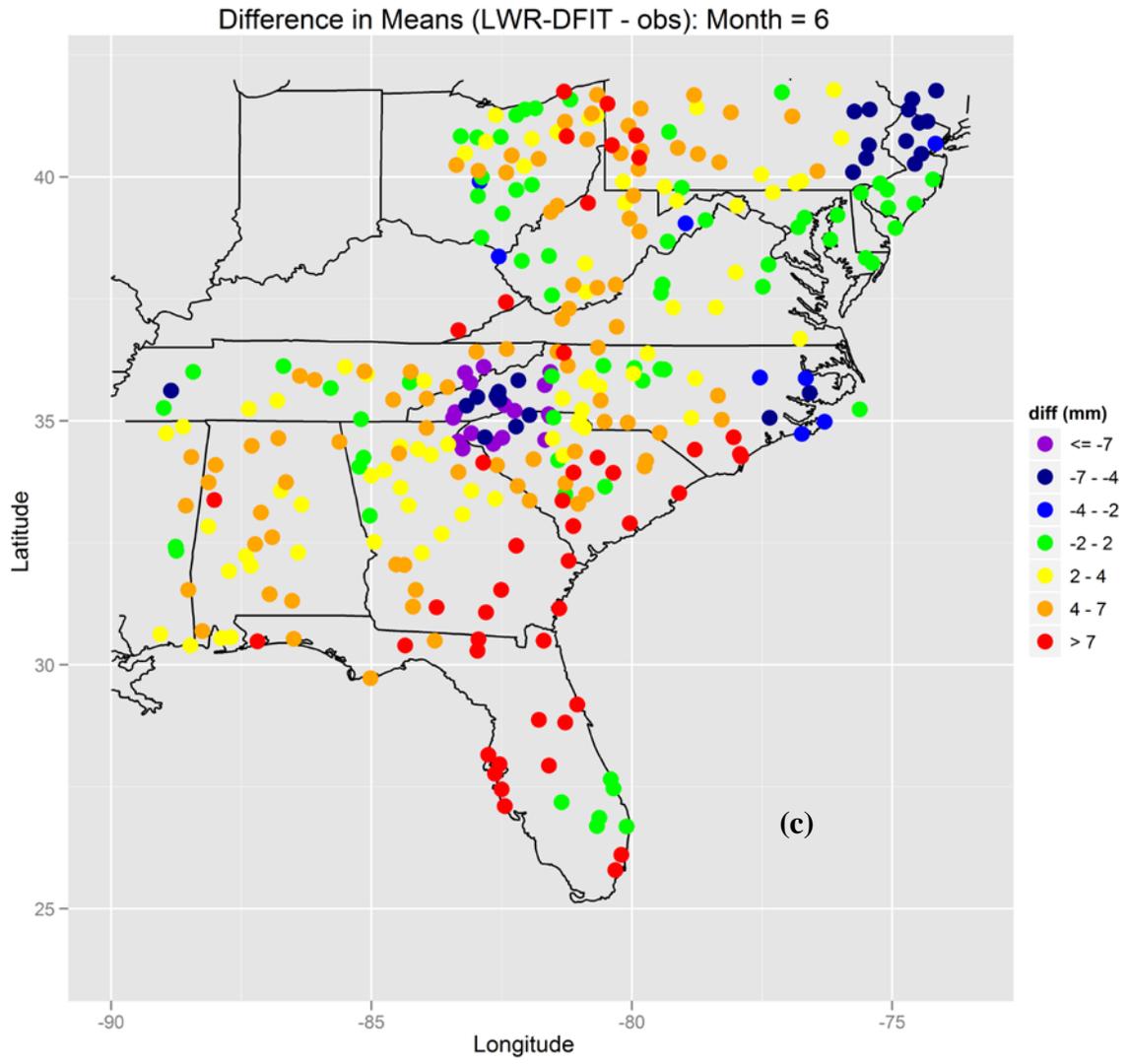


Figure 4.6. Difference between generated and observed values of mean daily nonzero precipitation (μ) for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for June.







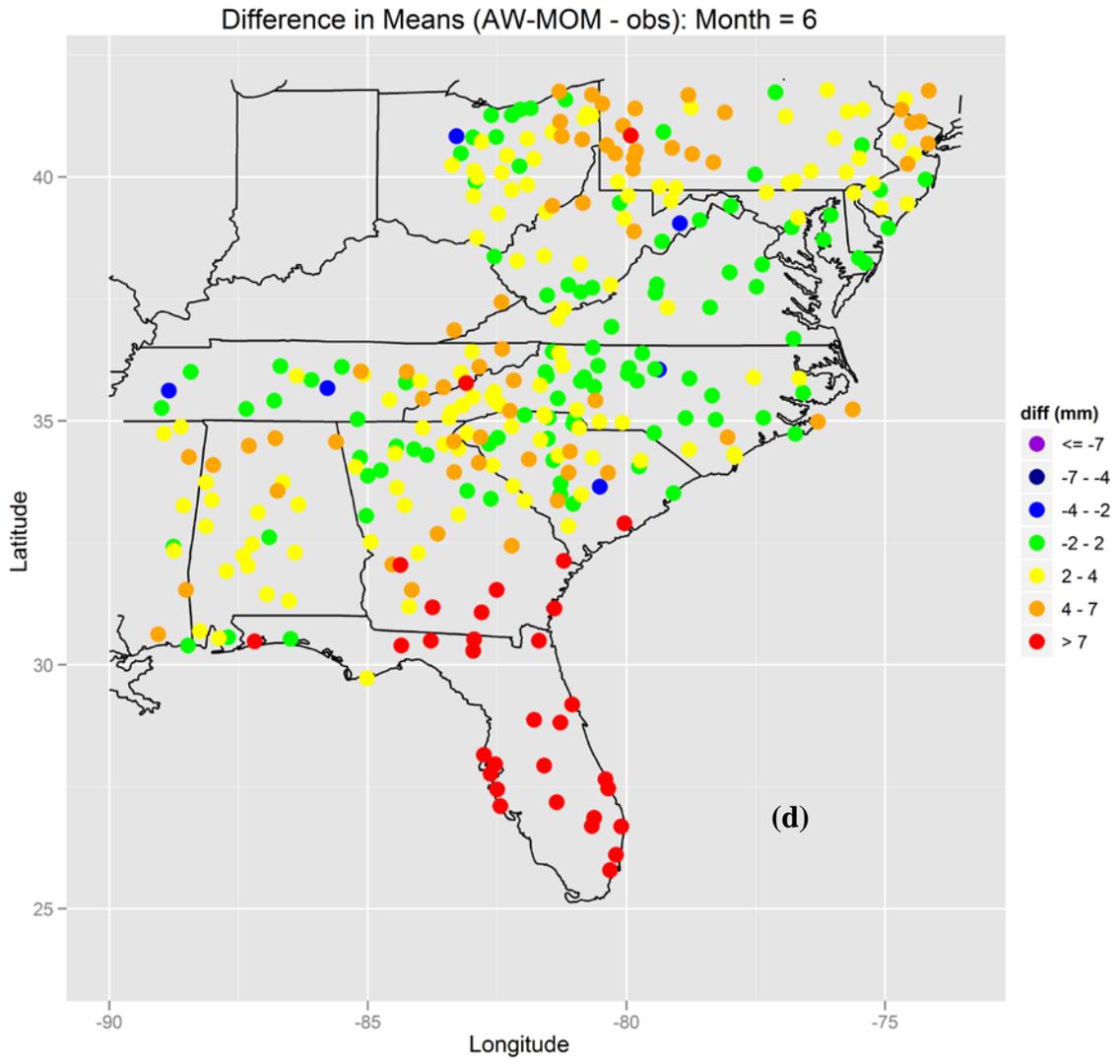
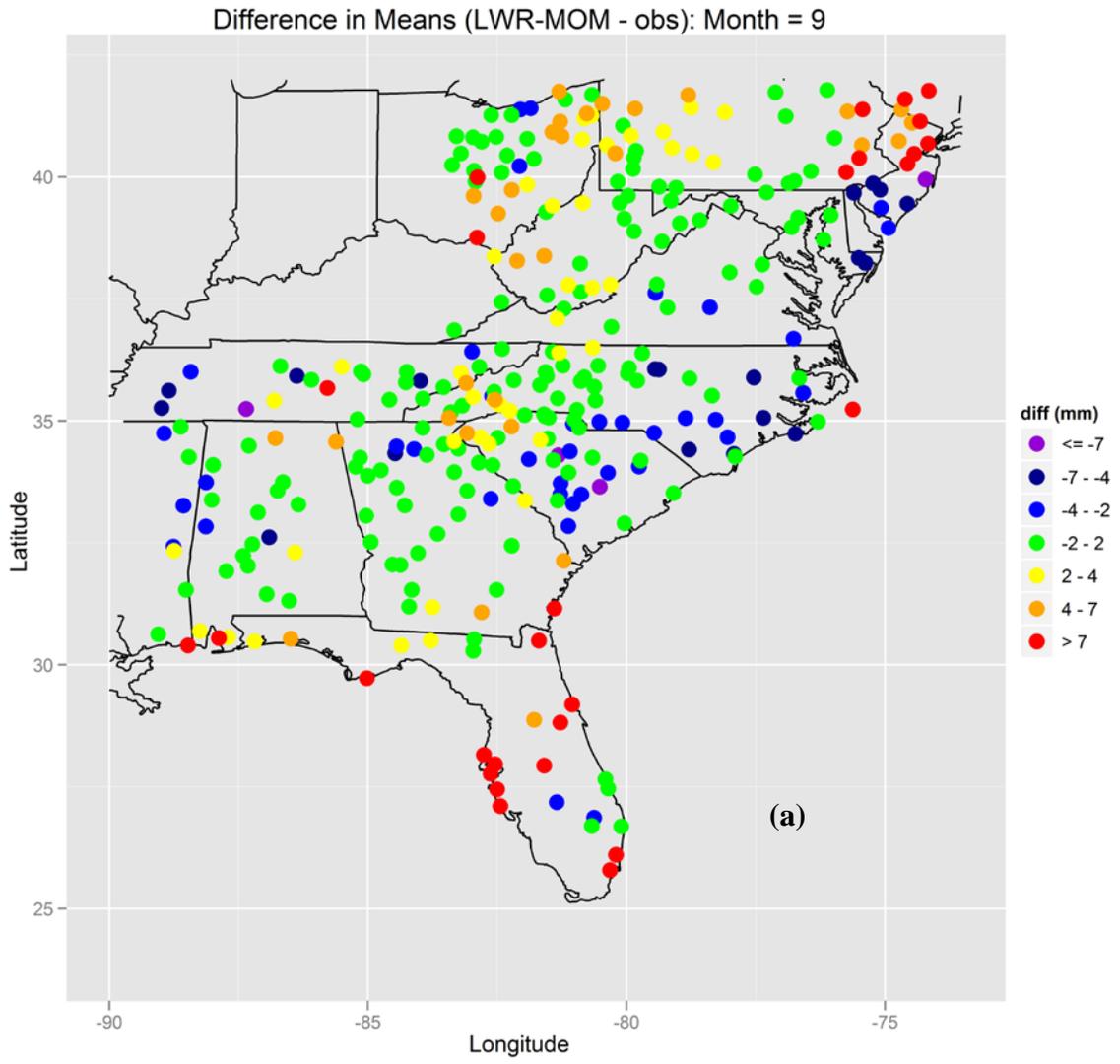
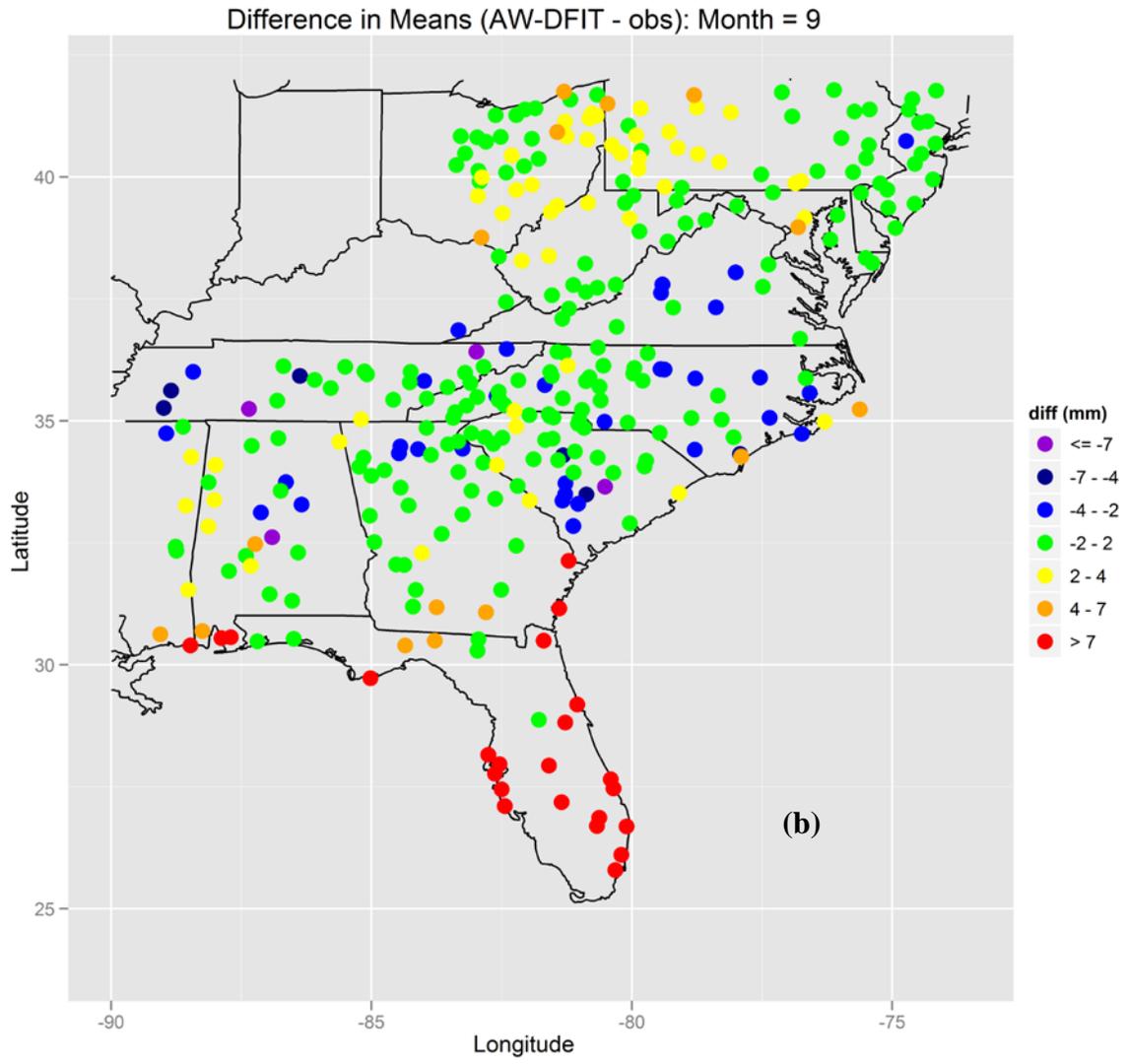
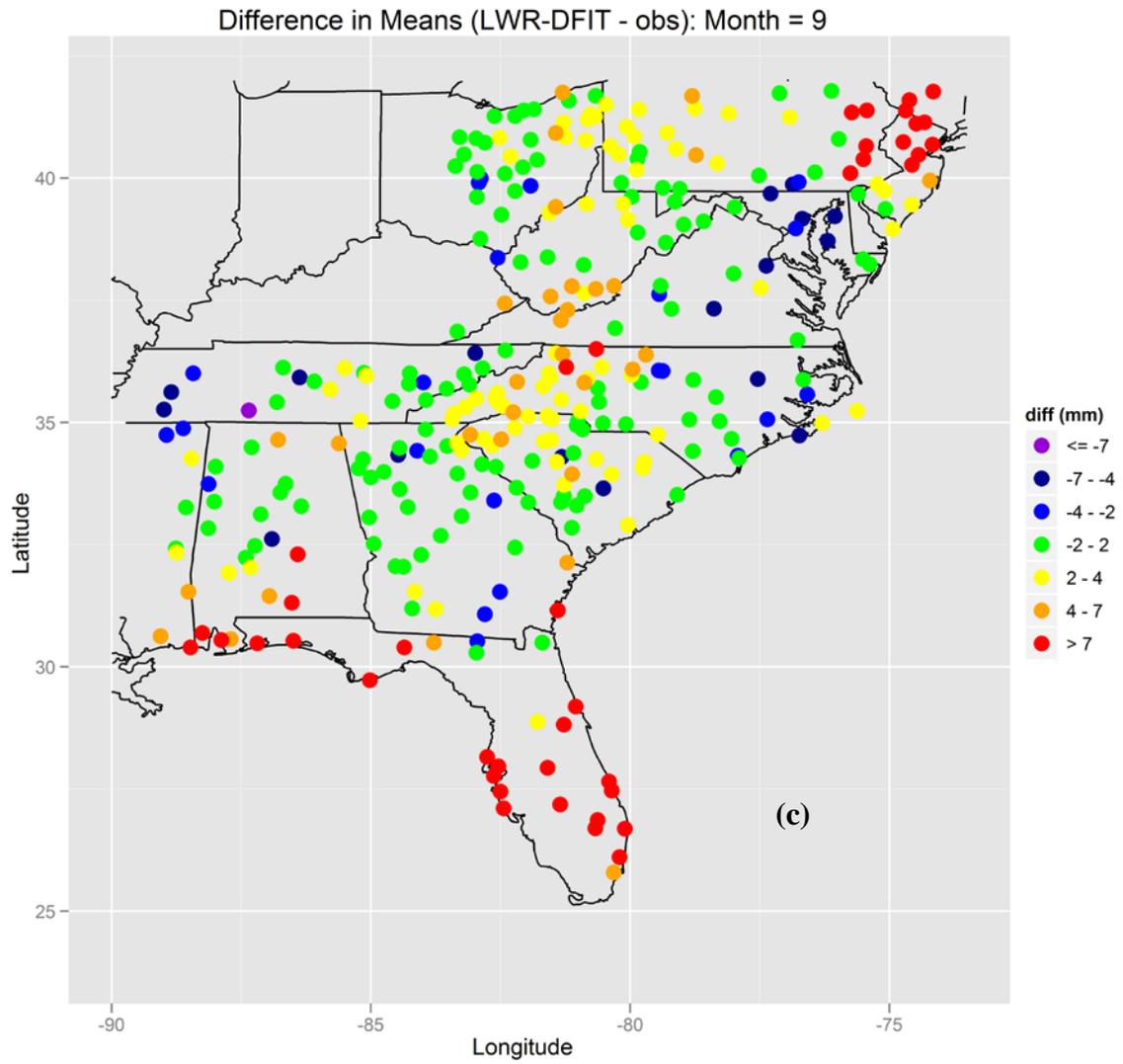
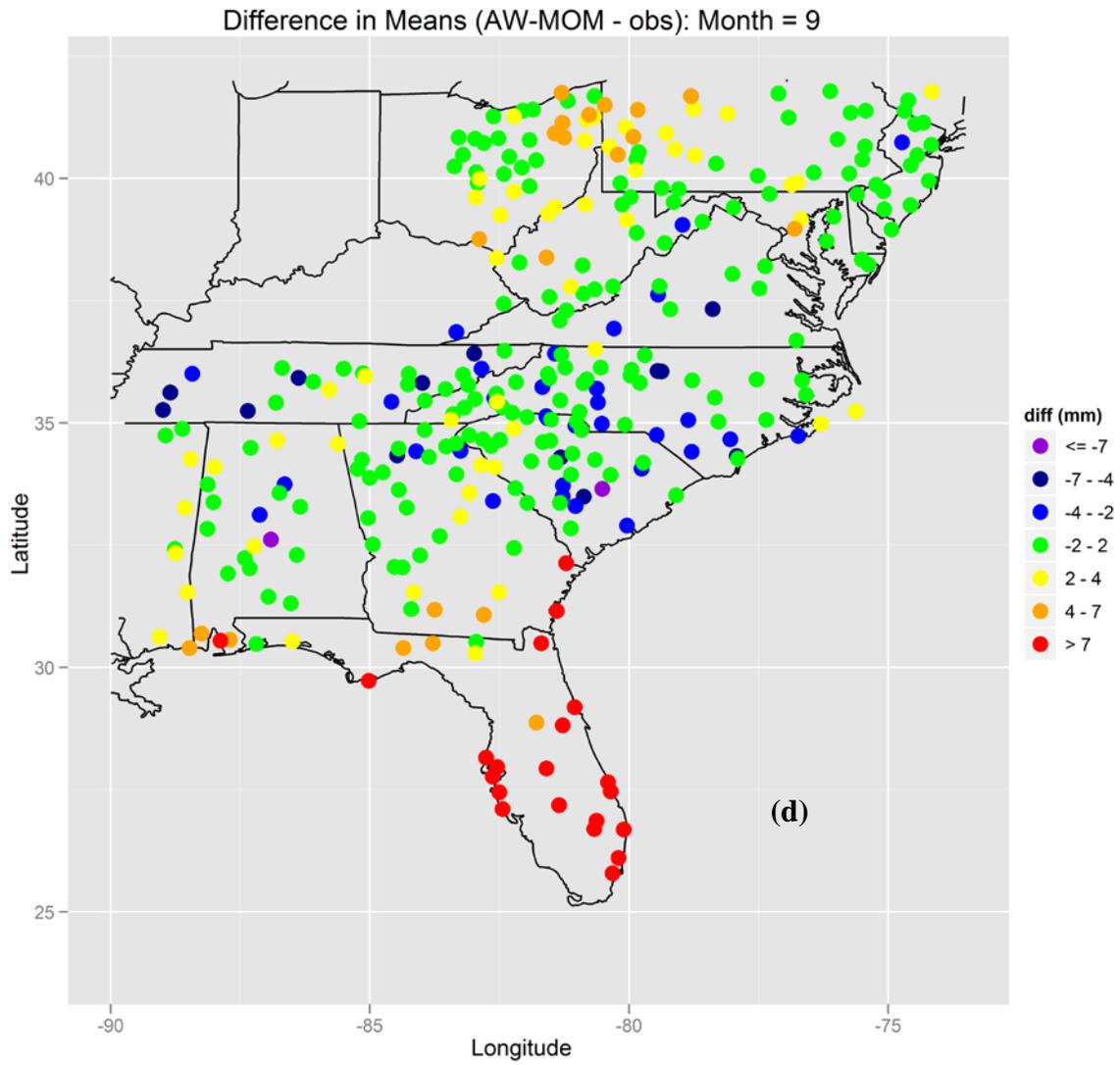


Figure 4.7. Difference between generated and observed values of mean daily nonzero precipitation (μ) for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for September.









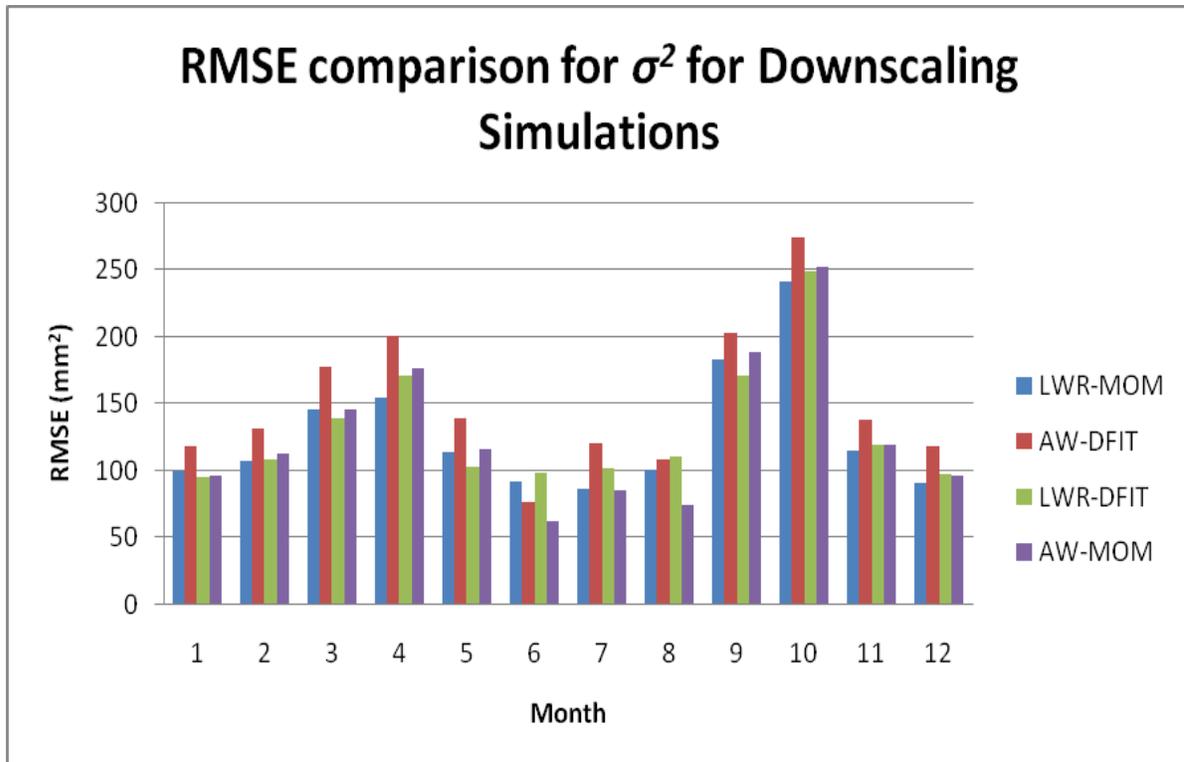
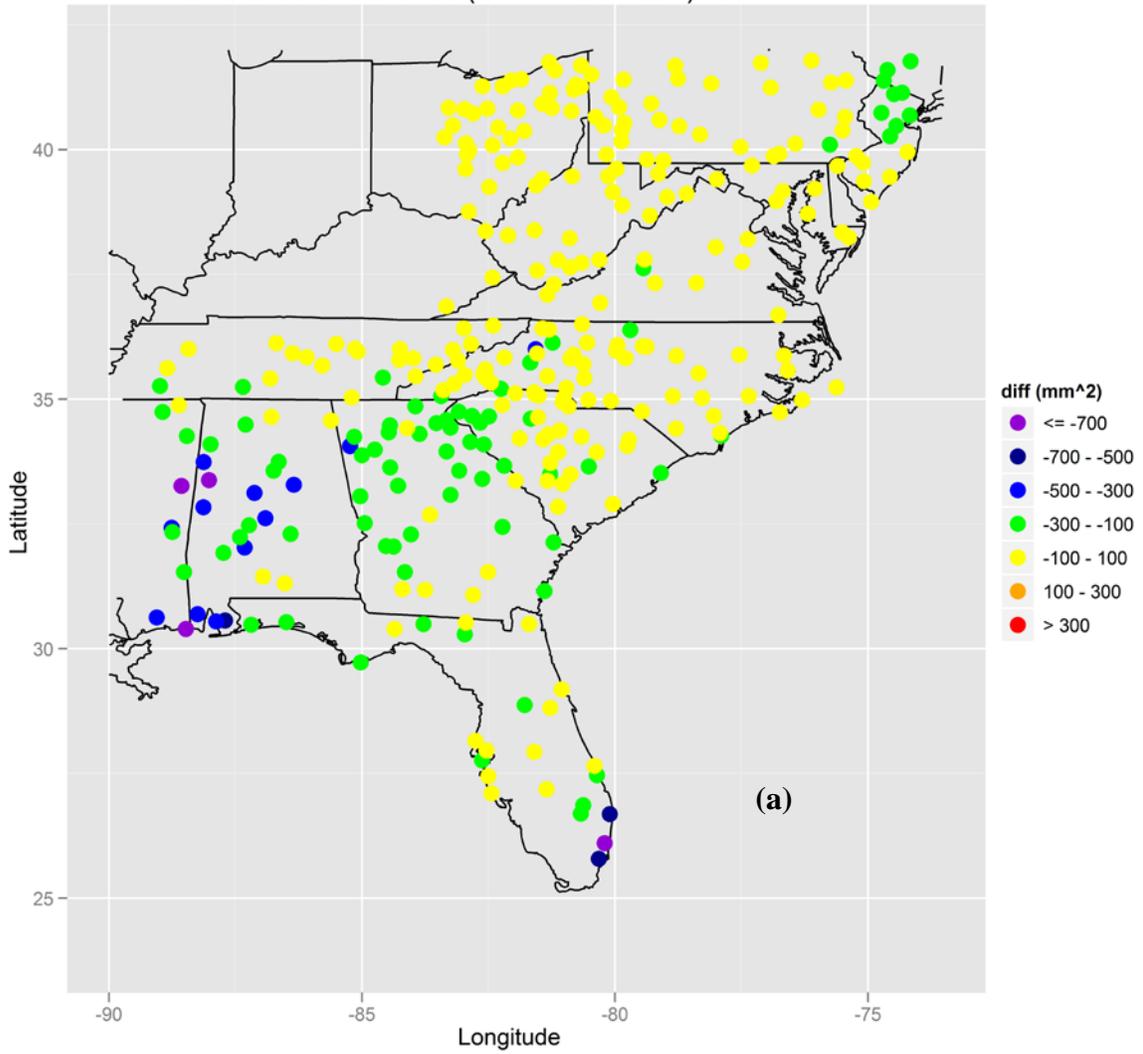


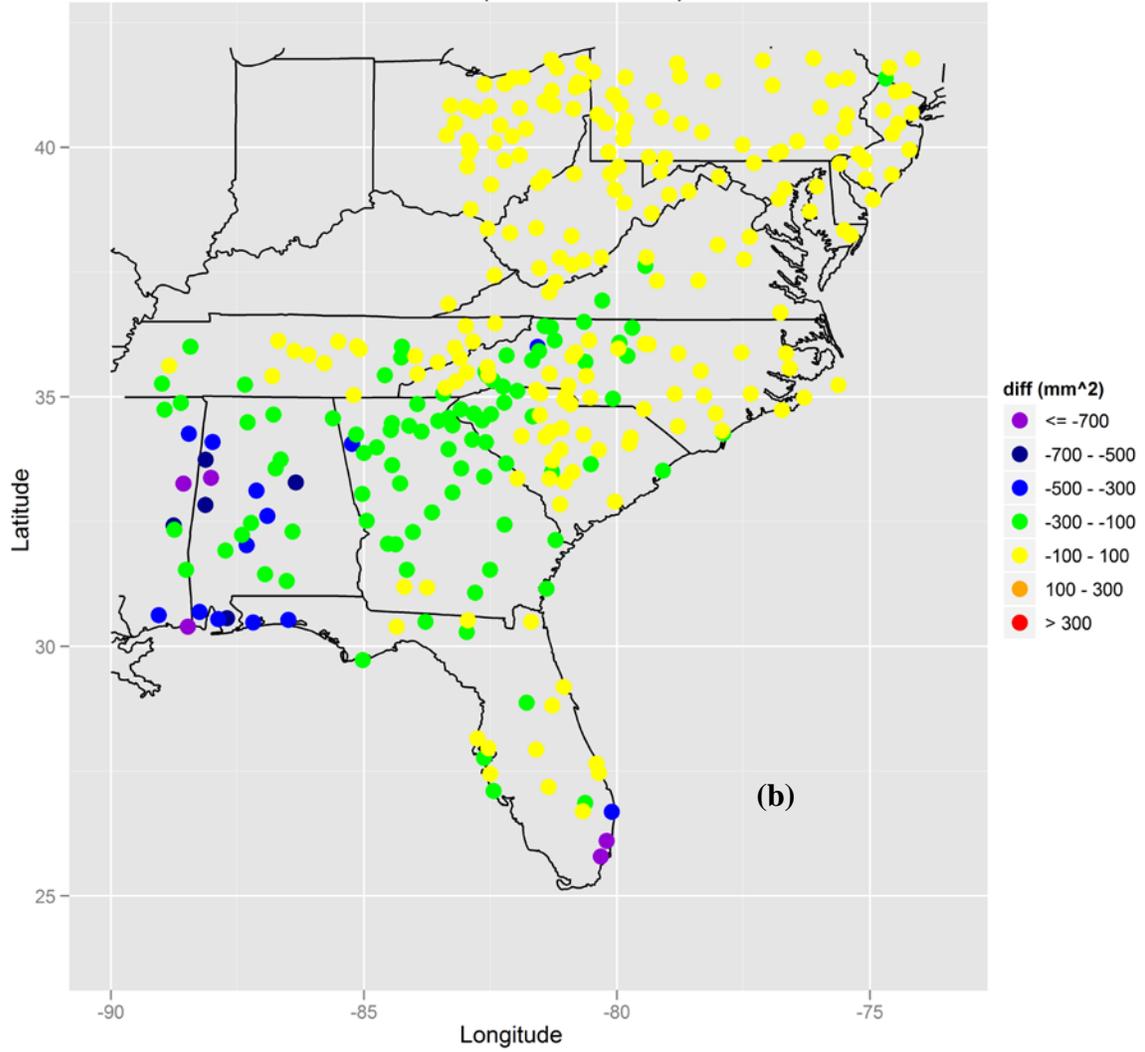
Figure 4.8. RMSE Comparison between downscaling simulations for the variance of daily nonzero precipitation (σ^2).

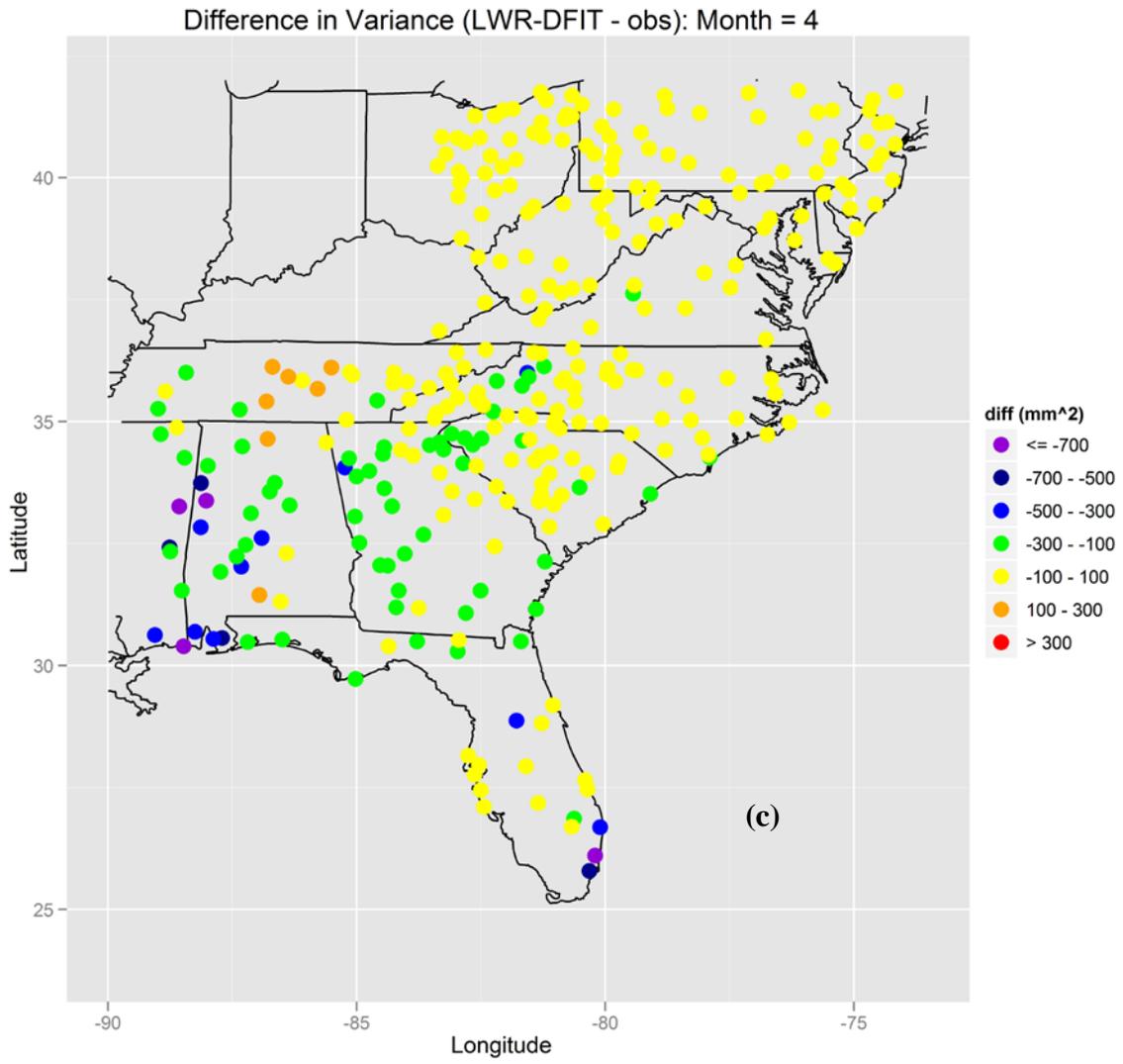
Figure 4.9. Difference between generated and observed values of the variance of daily nonzero precipitation (σ^2) for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for April.

Difference in Variance (LWR-MOM - obs): Month = 4



Difference in Variance (AW-DFIT - obs): Month = 4





Difference in Variance (AW-MOM - obs): Month = 4

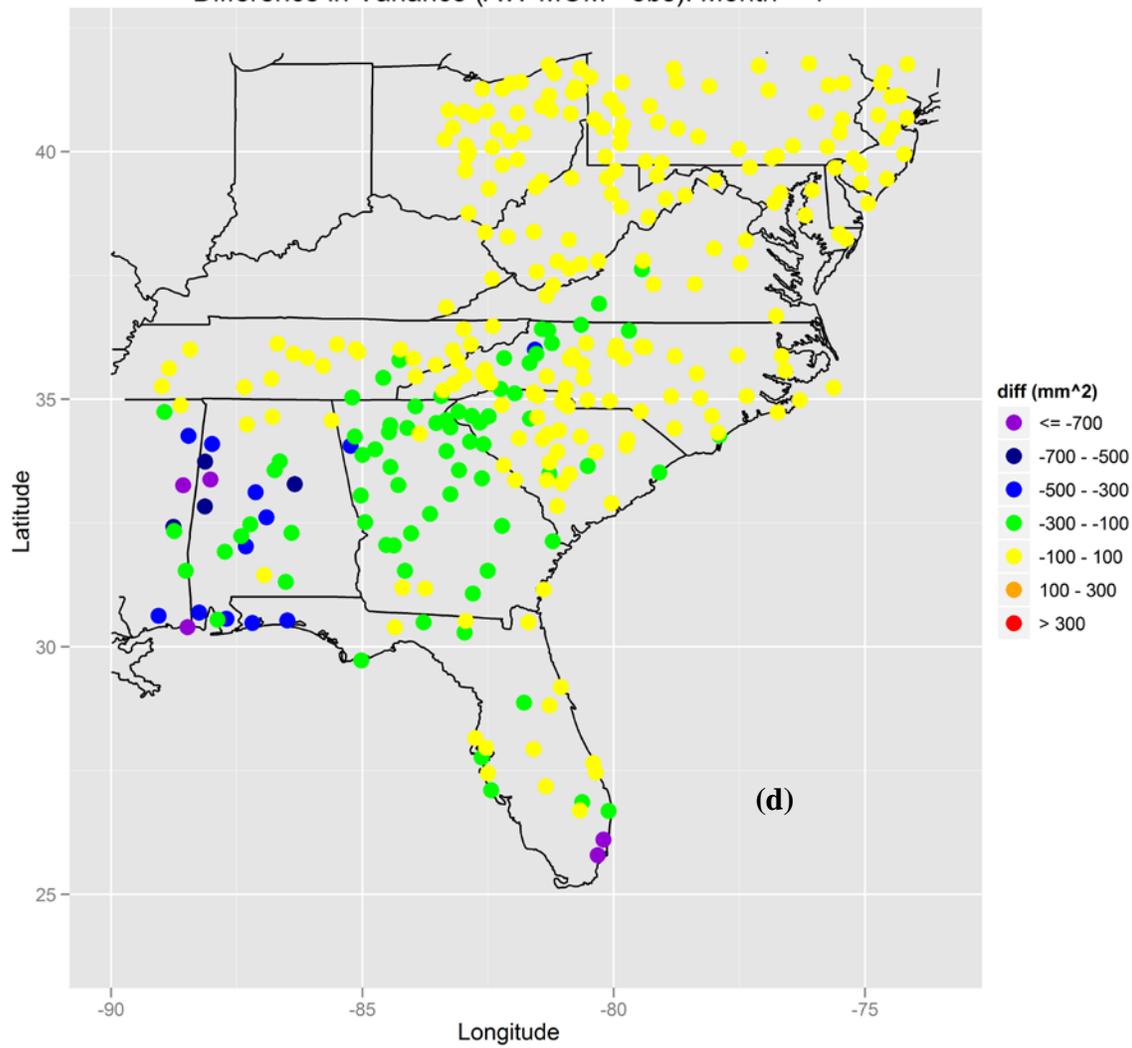
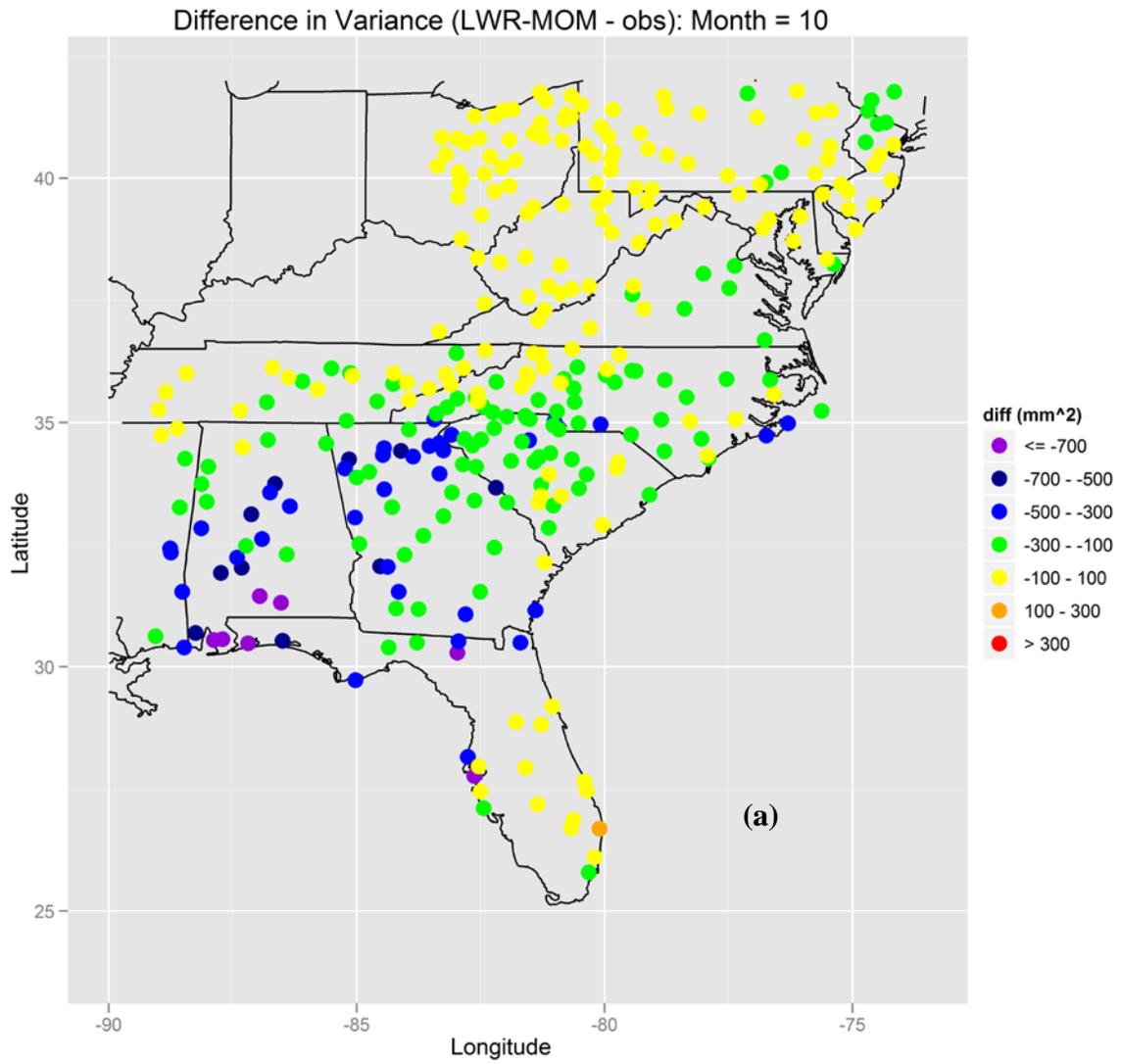
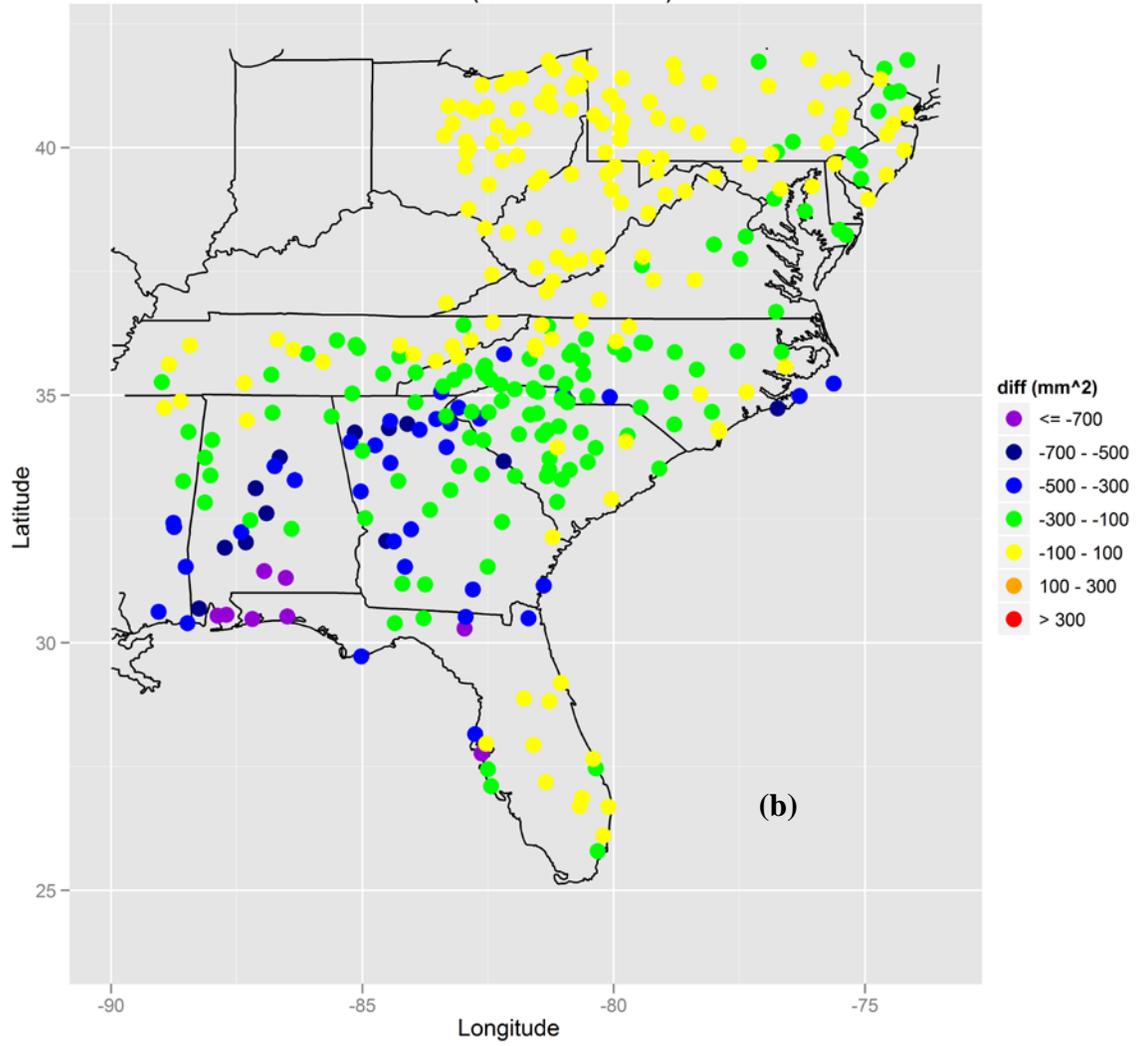
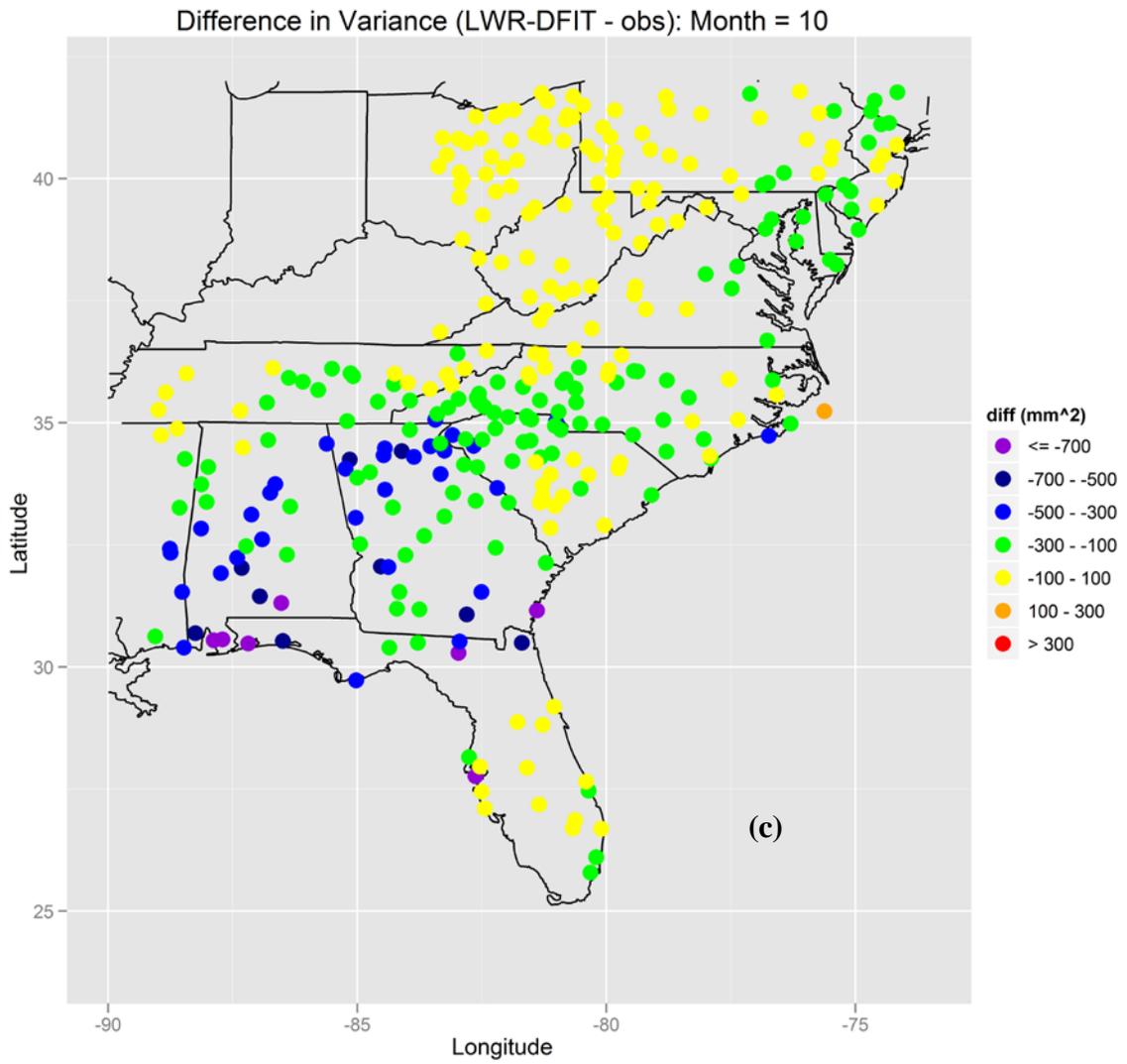


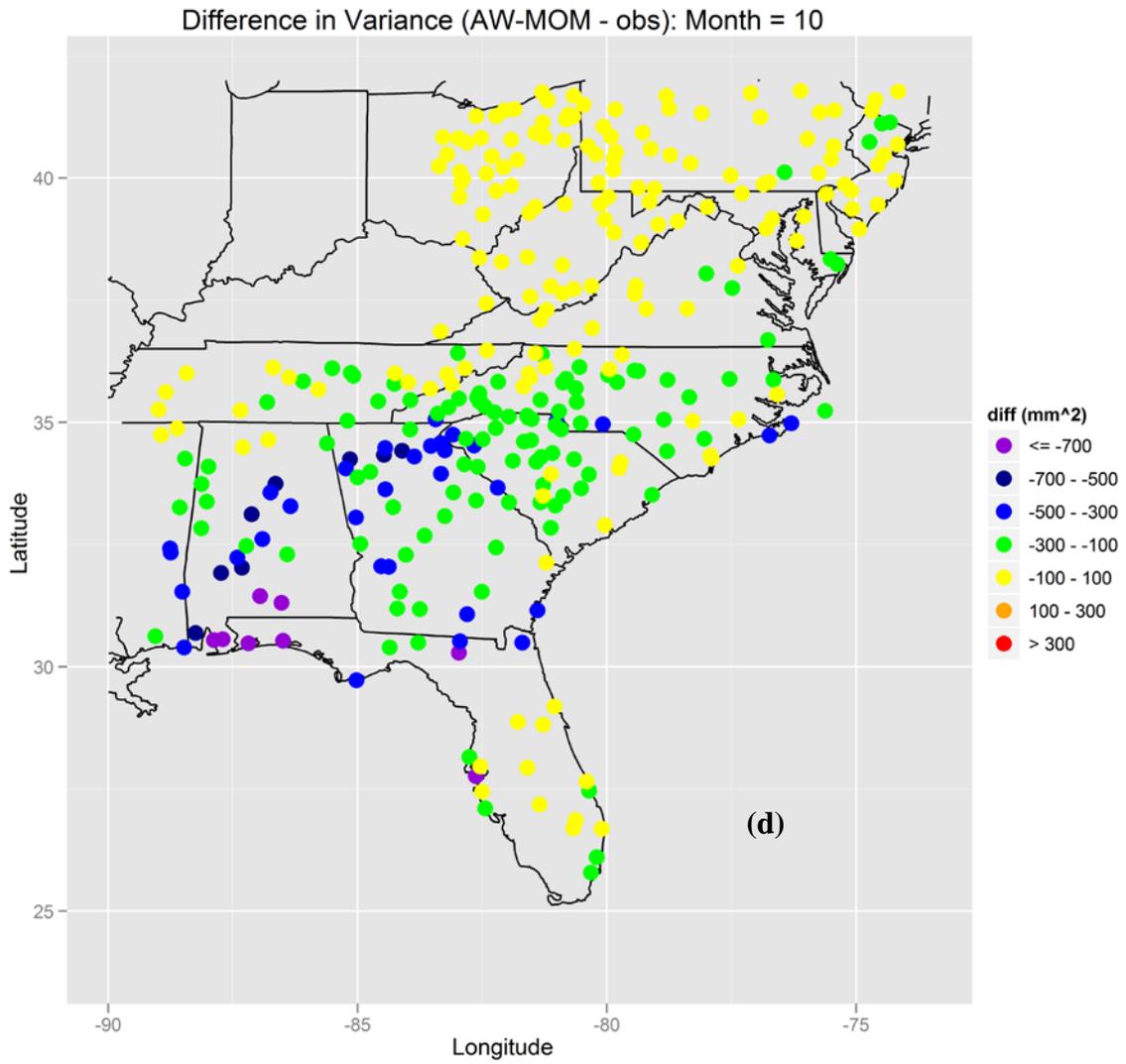
Figure 4.10. Difference between generated and observed values of the variance of daily nonzero precipitation (σ^2) for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for October.



Difference in Variance (AW-DFIT - obs): Month = 10







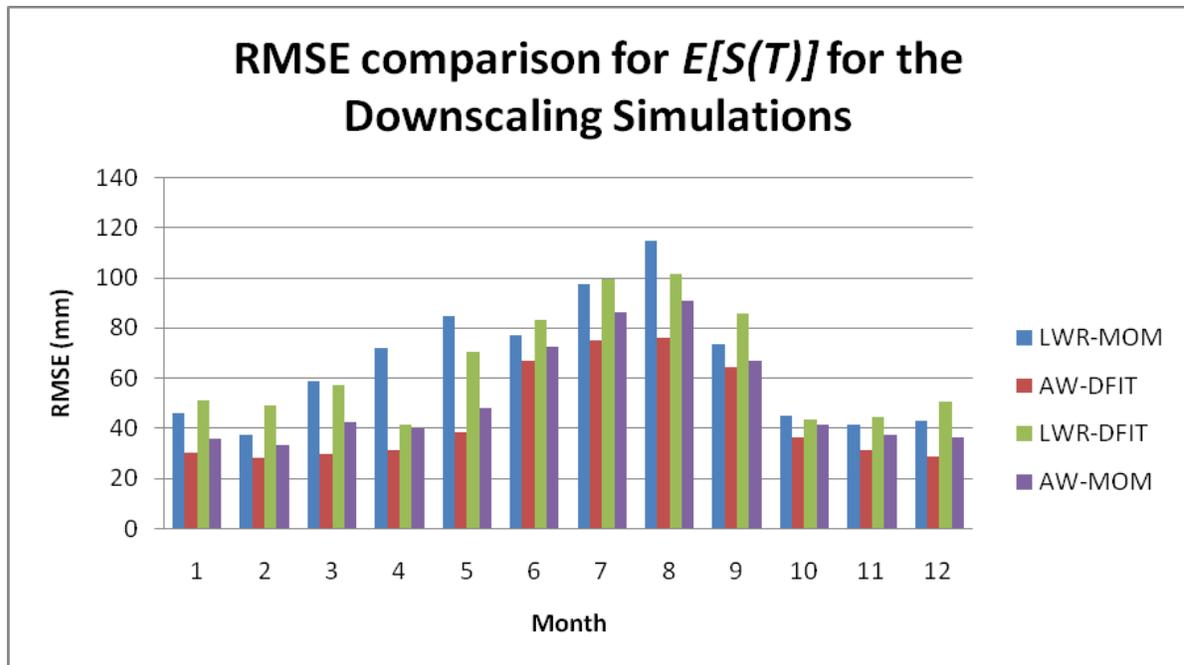


Figure 4.11. RMSE Comparison between downscaling simulations for the RMSE of the average total precipitation ($E[S(T)]$).

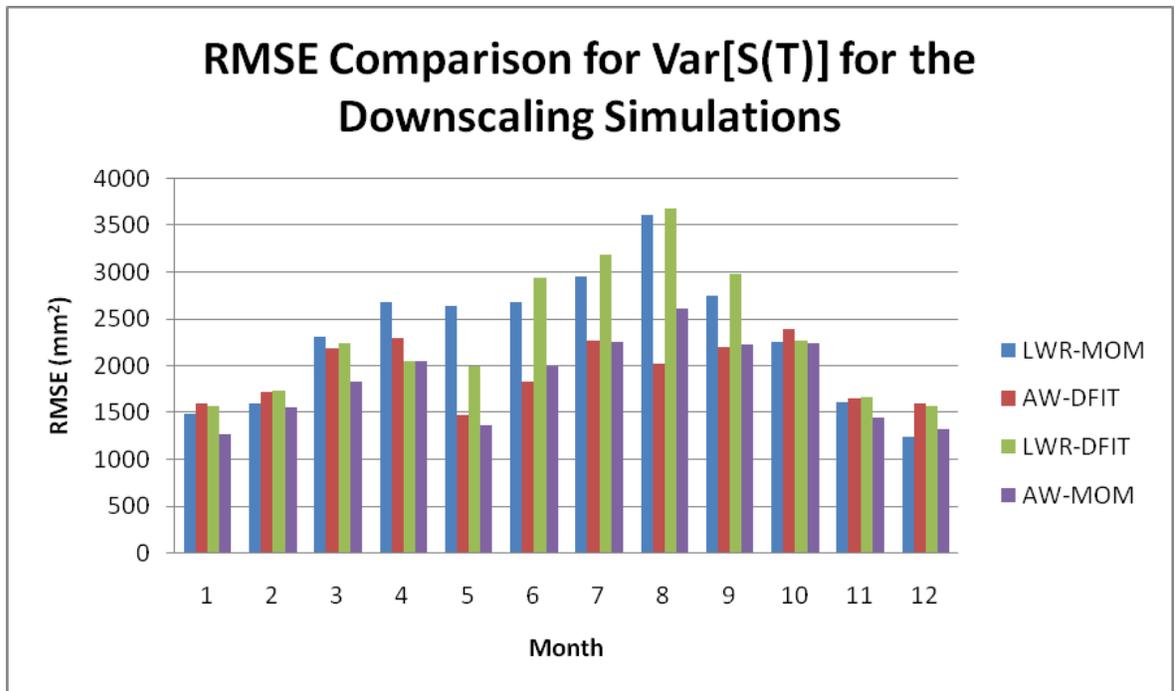
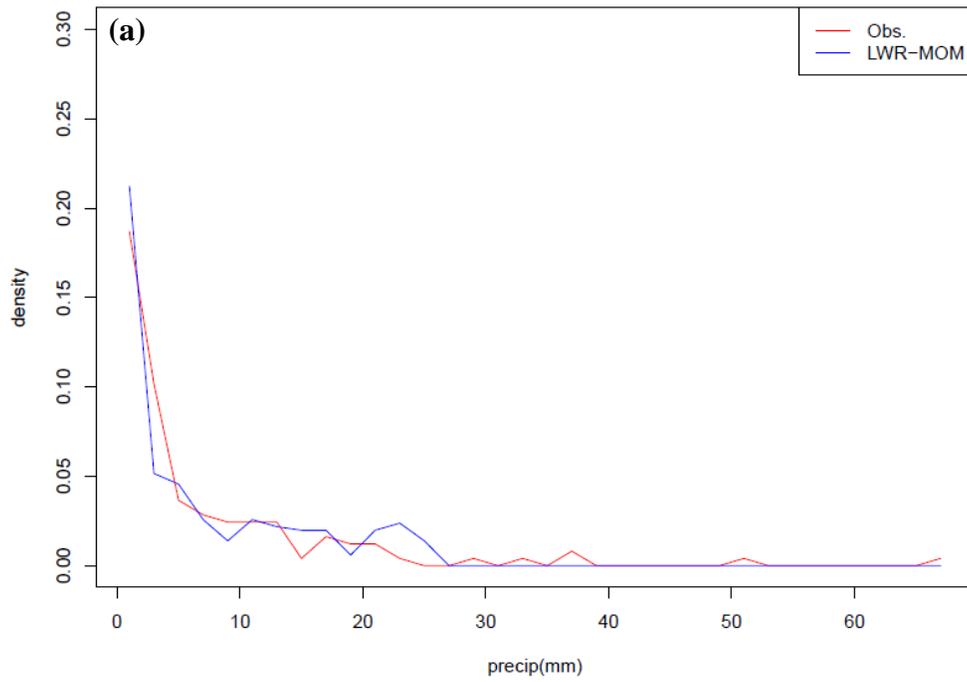


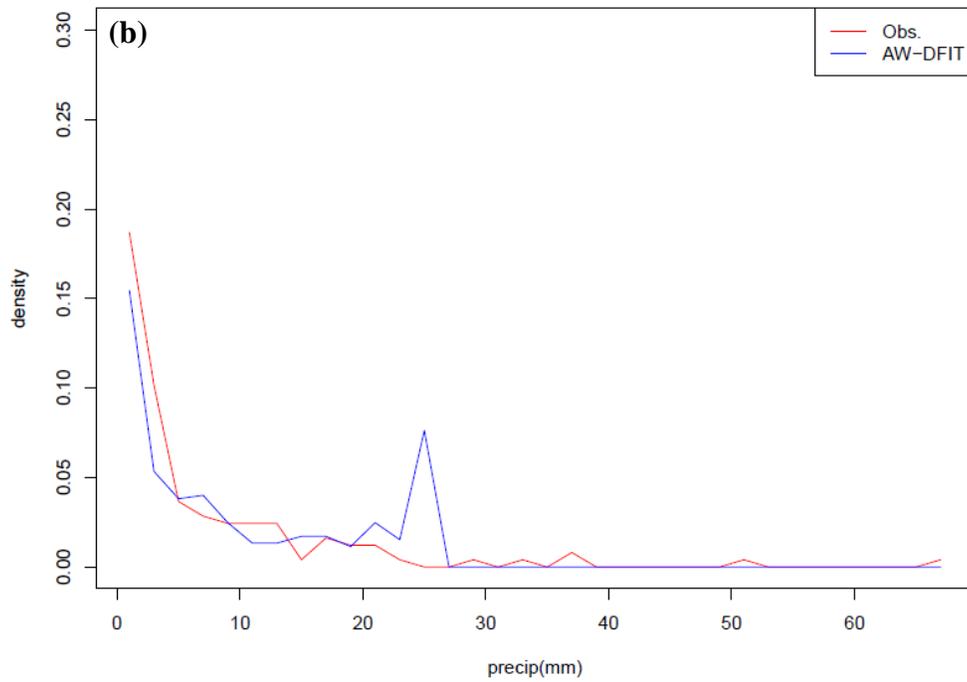
Figure 4.12. Comparison between downscaling simulations for the RMSE of the inter-annual variability ($\text{Var}[S(T)]$).

Figure 4.13. Generated vs. Observed PDF examples for a station 083163 in Ft. Lauderdale, FL for January for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).

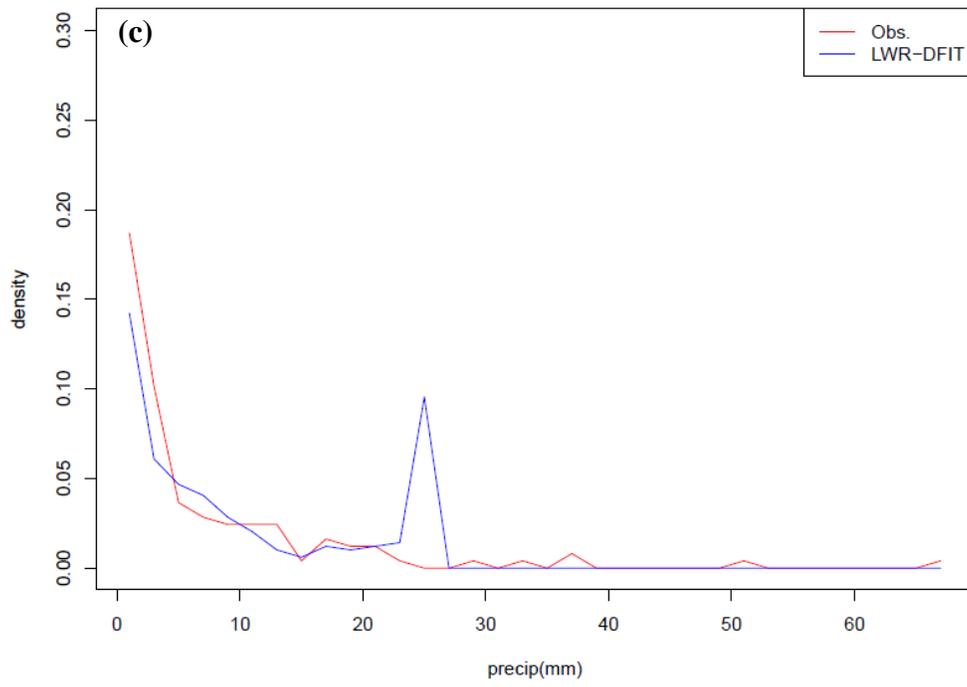
Comparison of Nonzero Rainfall PDF
Station: 083163 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 083163 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 083163 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 083163 Month: 1

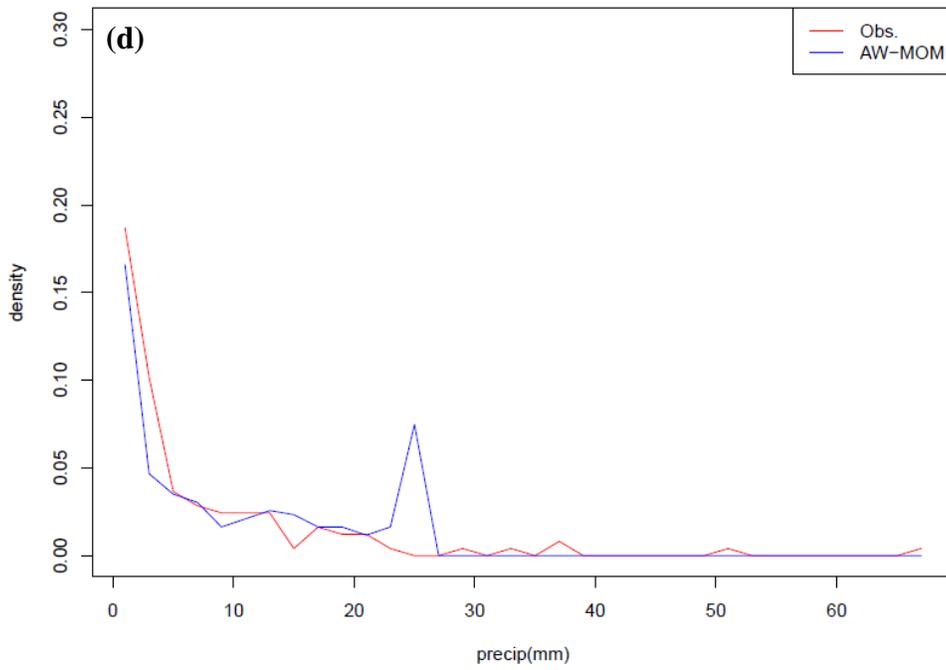
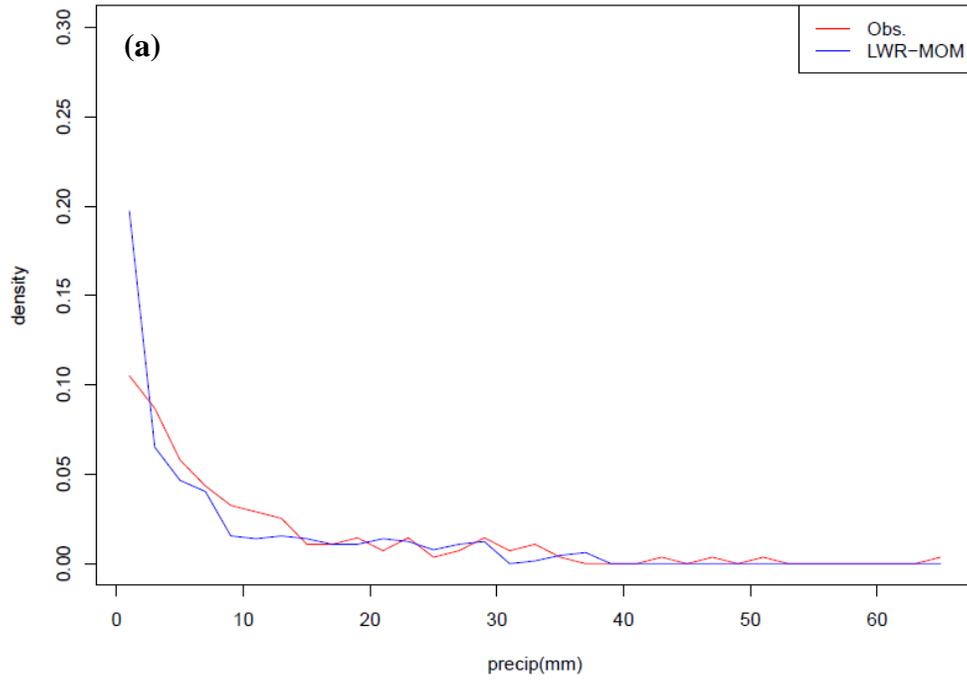
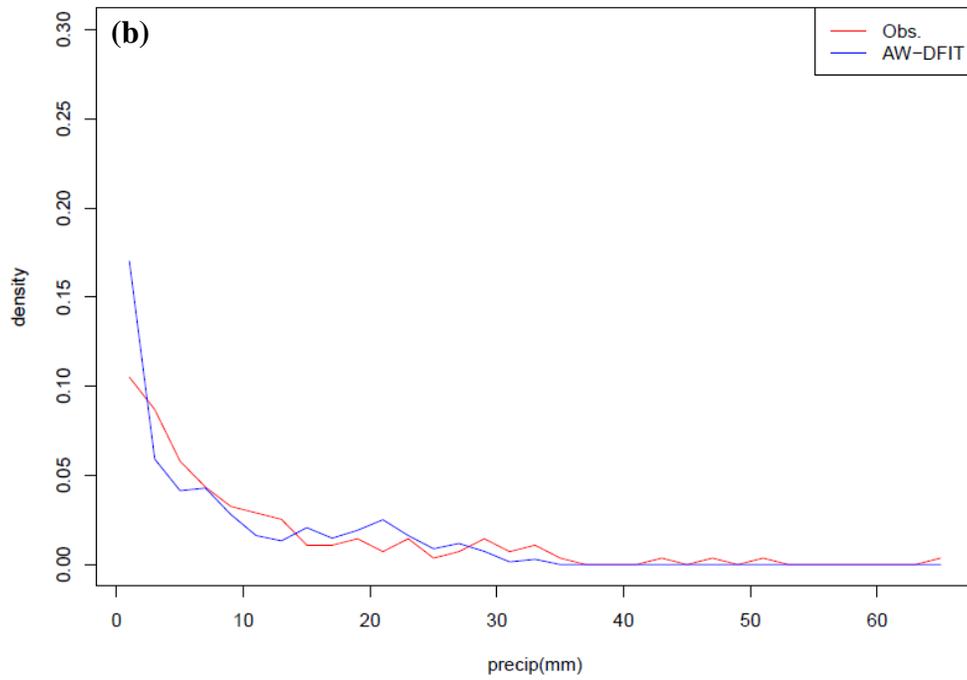


Figure 4.14. Generated vs. Observed PDF examples for a station 311975 in Concord, NC for January for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).

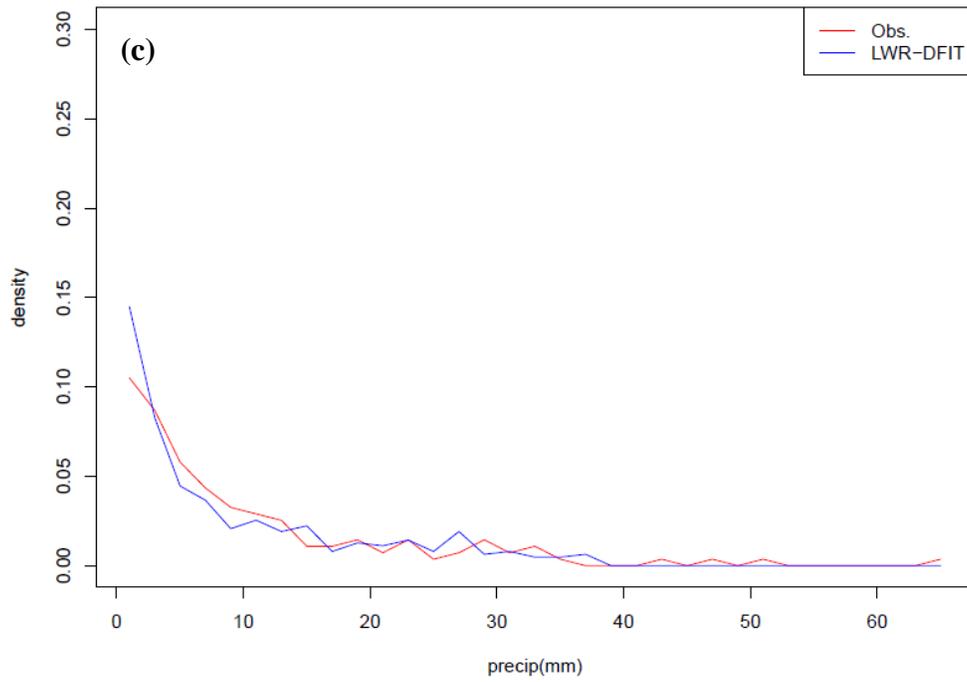
Comparison of Nonzero Rainfall PDF
Station: 311975 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 311975 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 311975 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 311975 Month: 1

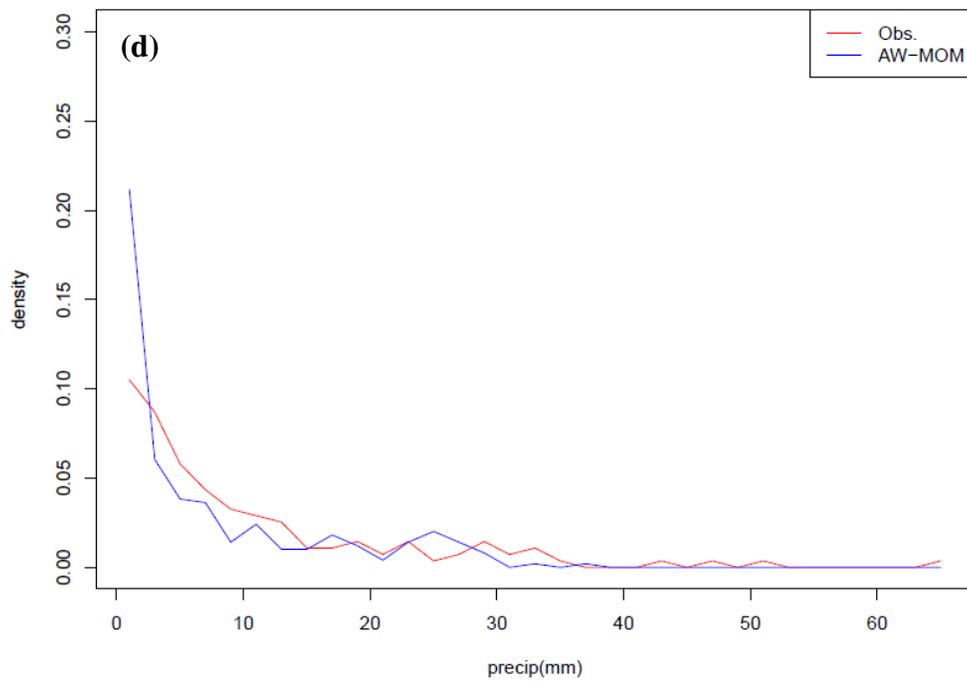
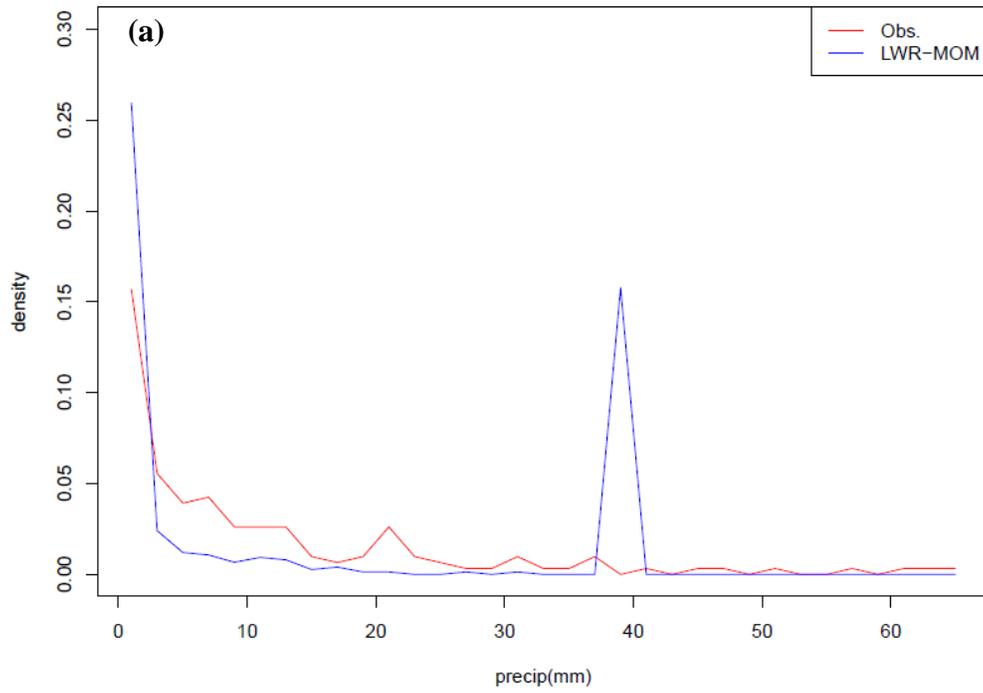
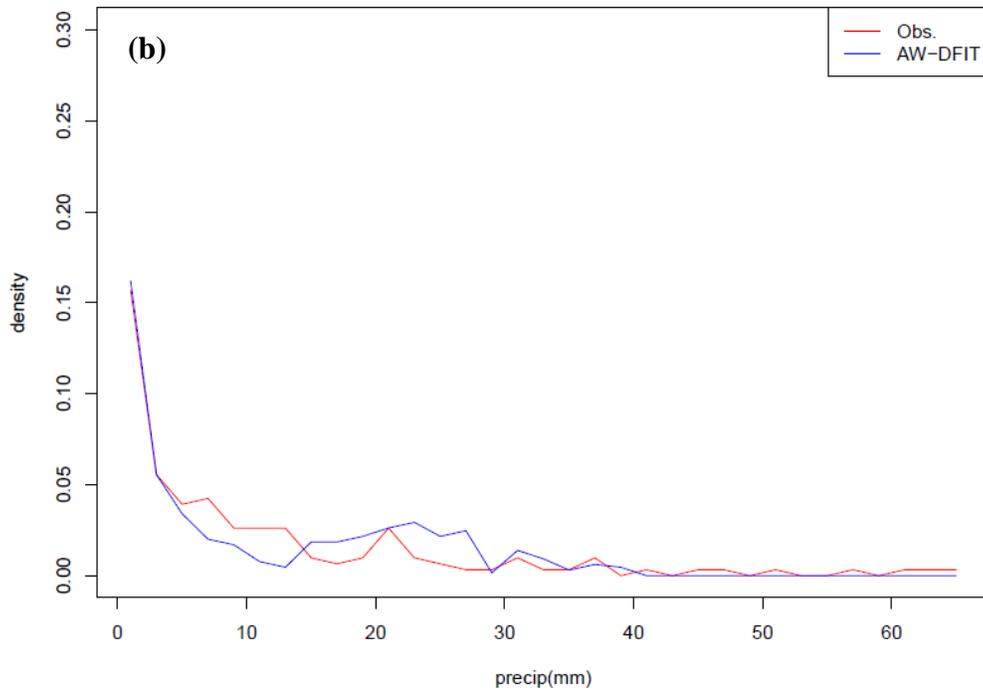


Figure 4.15. Generated vs. Observed PDF examples for a station 283516 in Greenwood Lake, NJ for January for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).

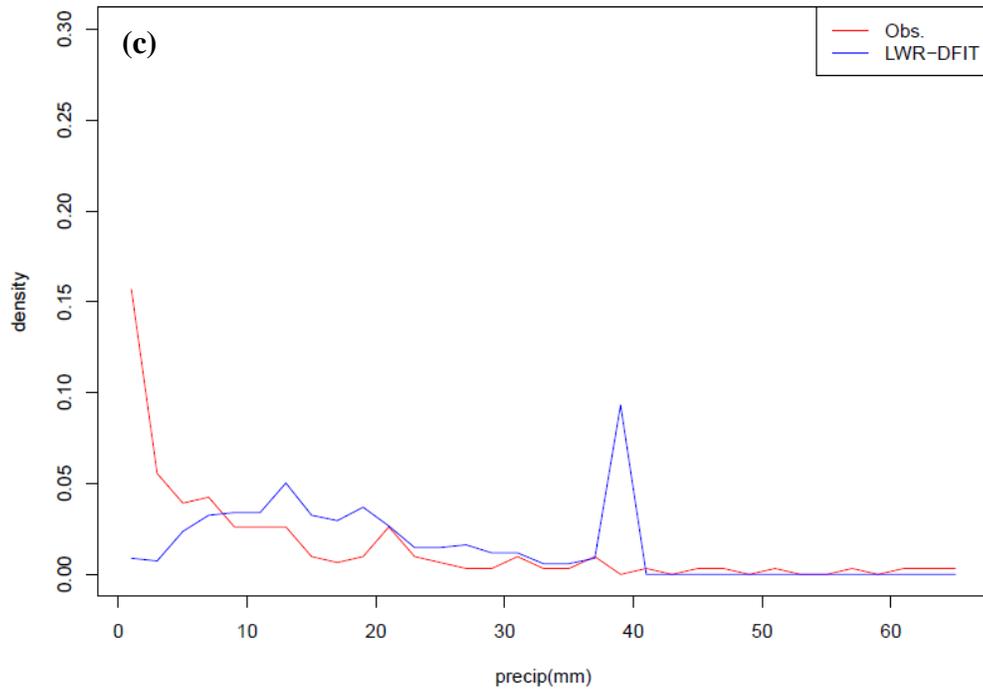
Comparison of Nonzero Rainfall PDF
Station: 283516 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 283516 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 283516 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 283516 Month: 1

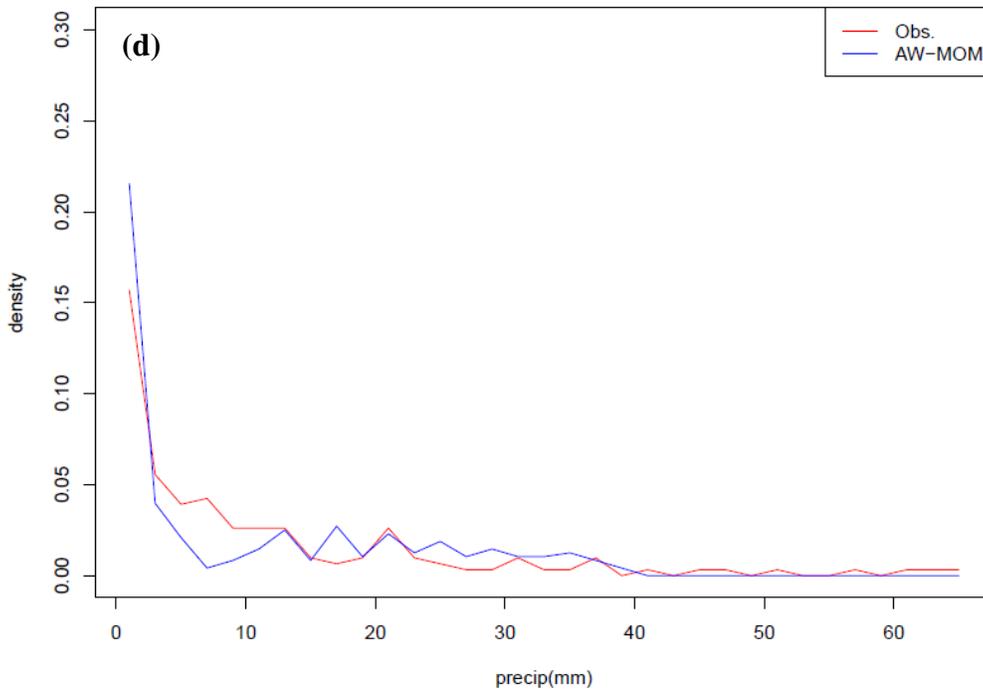
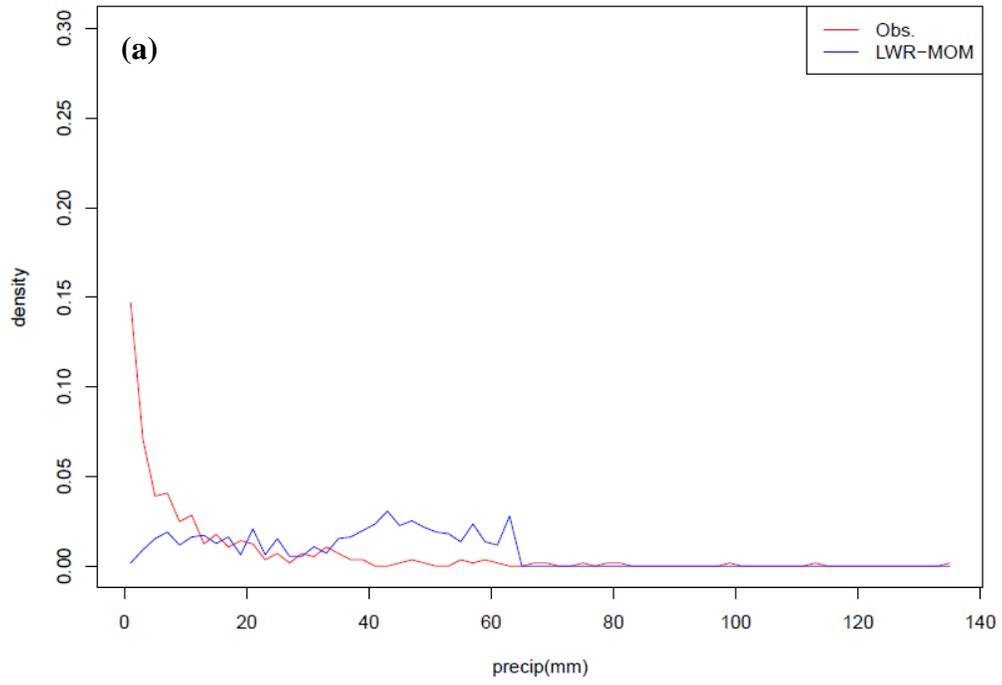
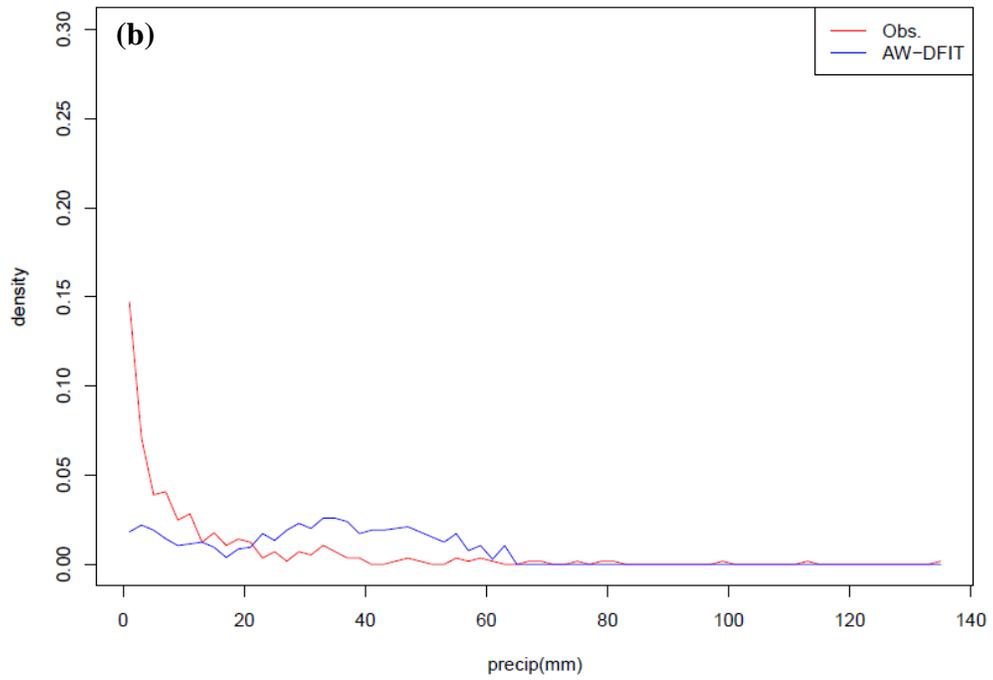


Figure 4.16. Generated vs. Observed PDF examples for a station 083163 in Ft. Lauderdale, FL for September for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).

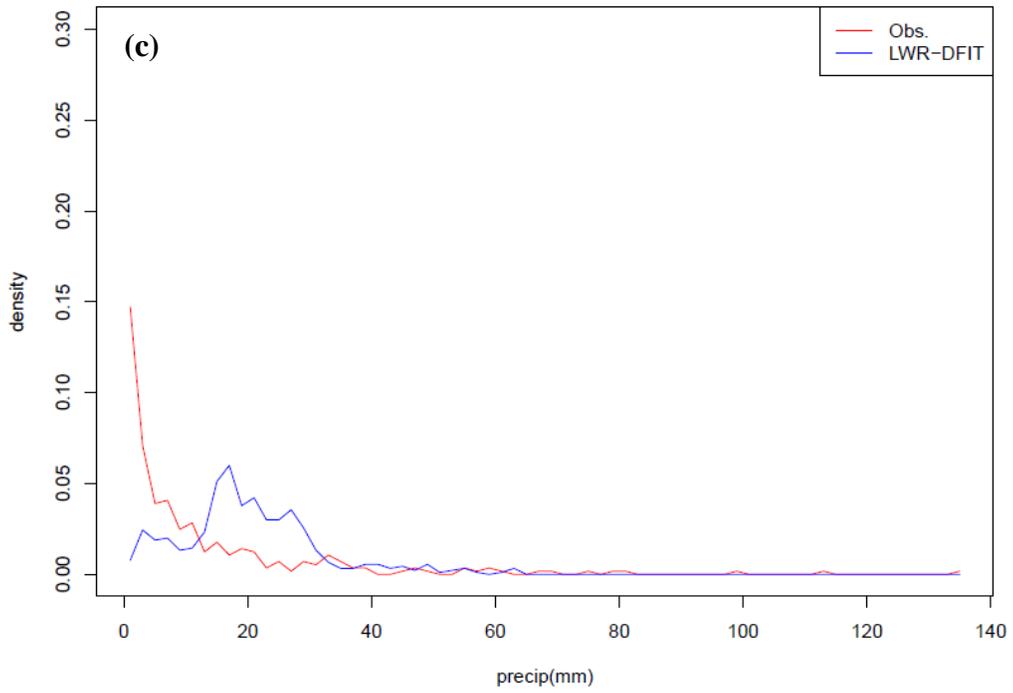
Comparison of Nonzero Rainfall PDF
Station: 083163 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 083163 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 083163 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 083163 Month: 9

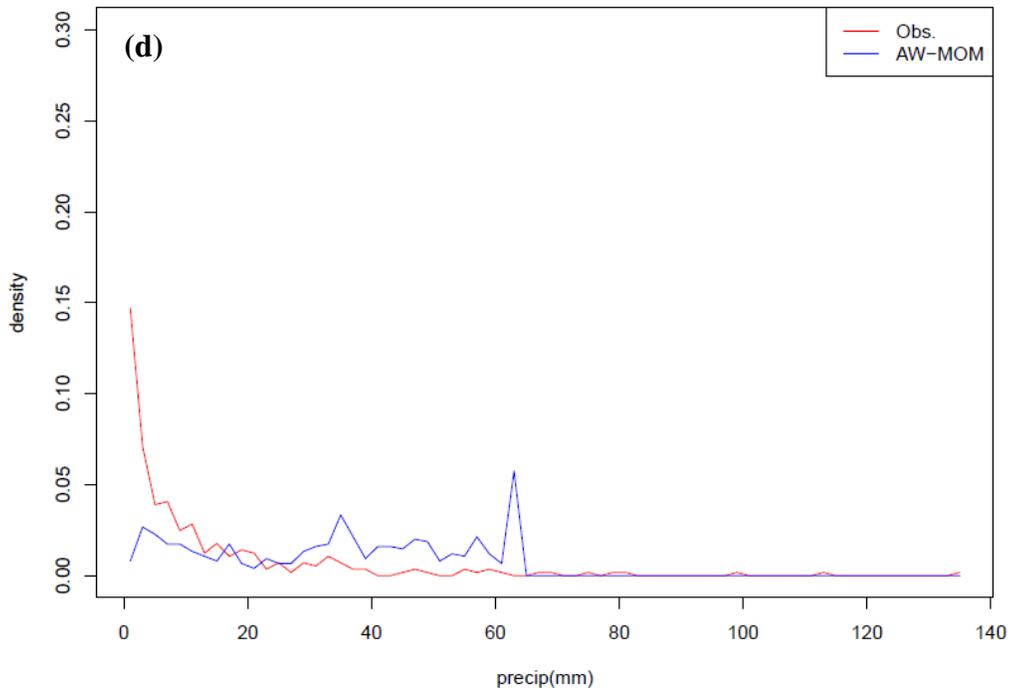
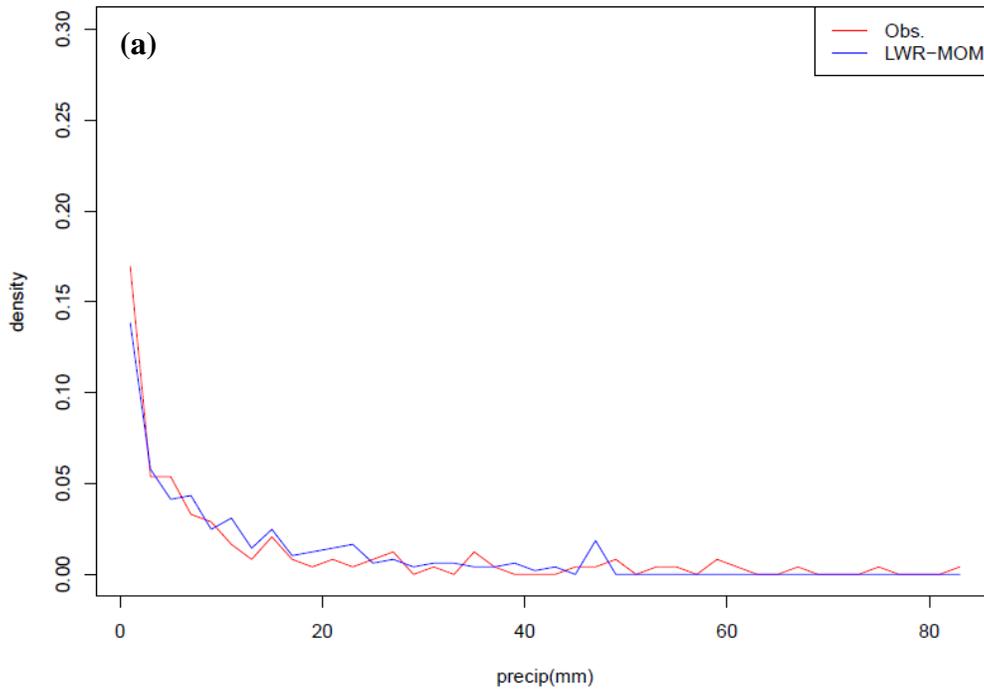
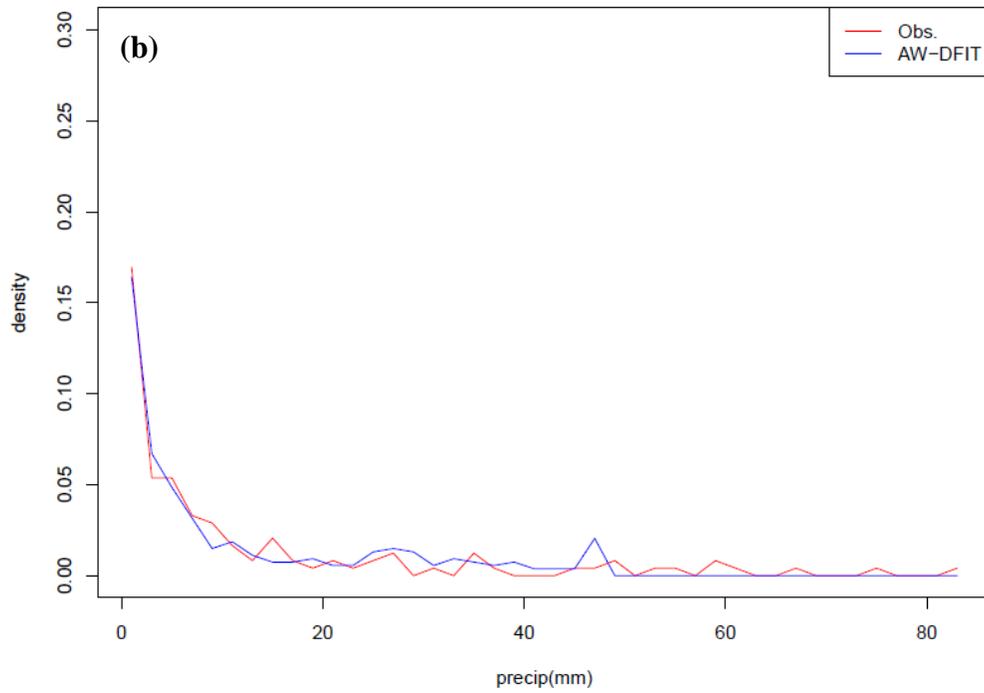


Figure 4.17. Generated vs. Observed PDF examples for a station 311975 in Concord, NC for September for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).

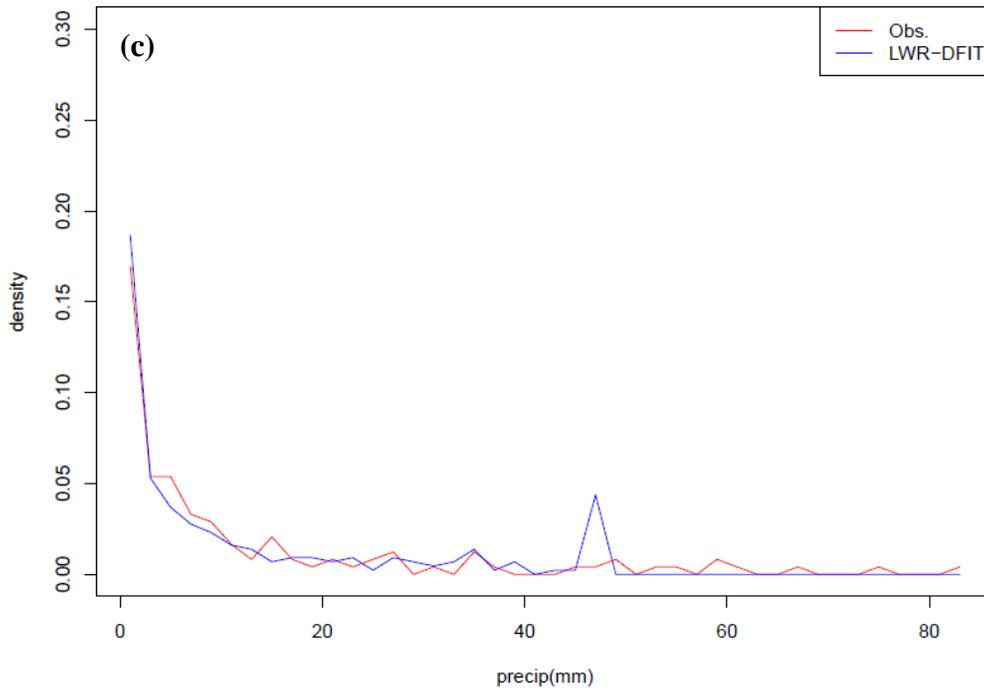
Comparison of Nonzero Rainfall PDF
Station: 311975 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 311975 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 311975 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 311975 Month: 9

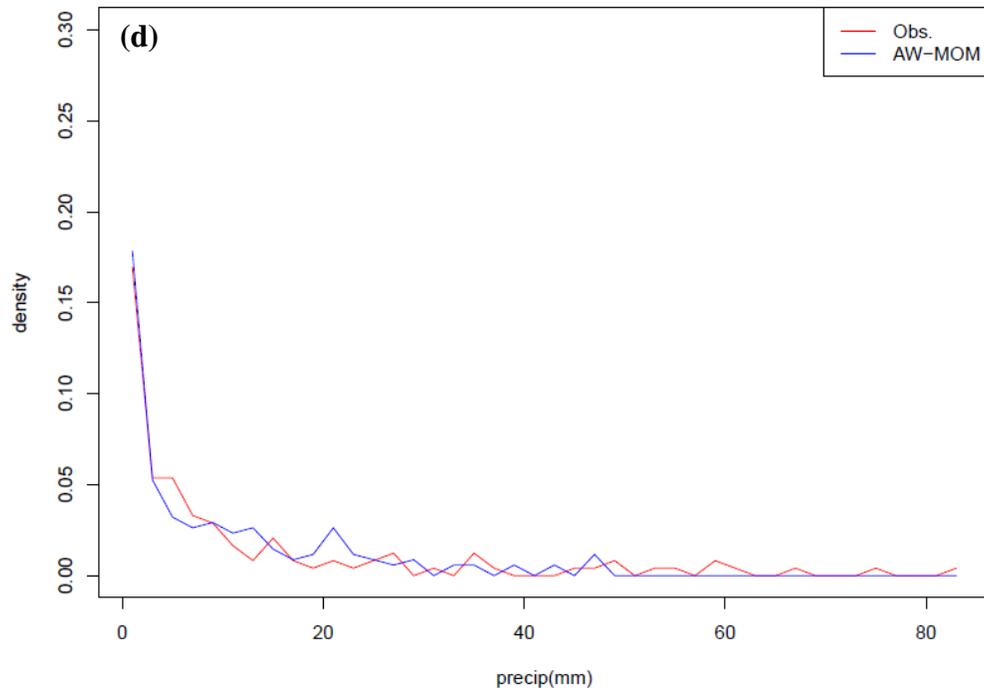
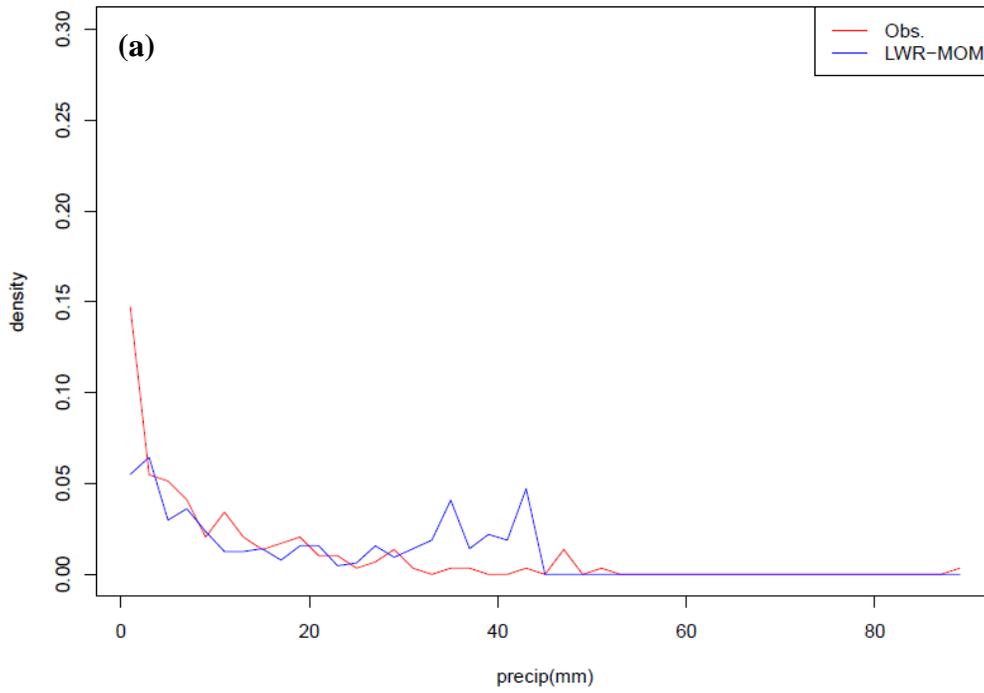
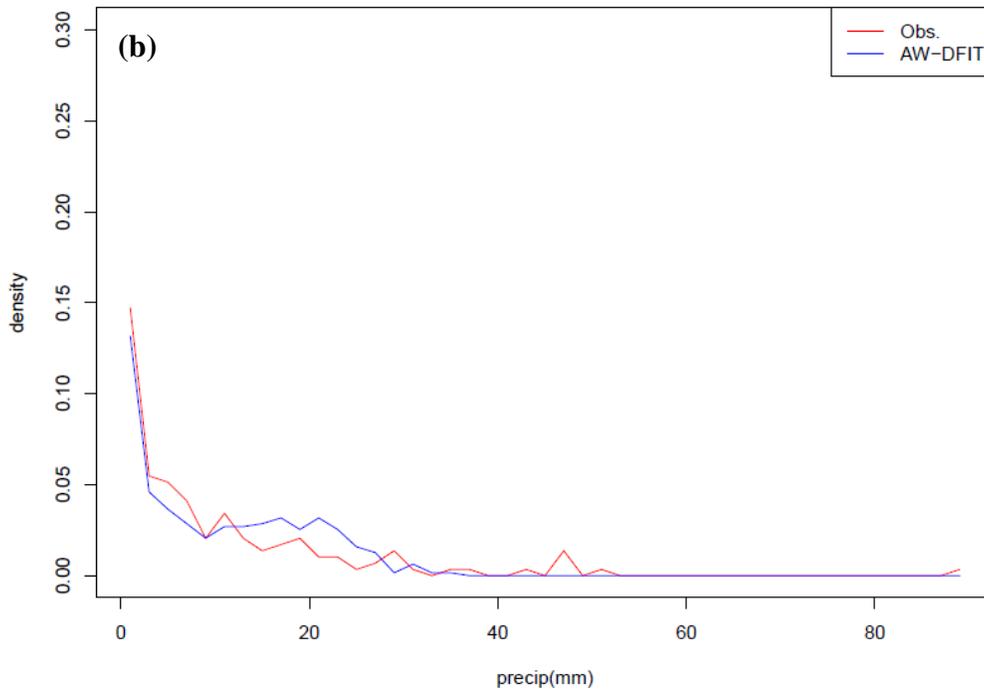


Figure 4.18. Generated vs. Observed PDF examples for a station 283516 in Greenwood Lake, NJ for September for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).

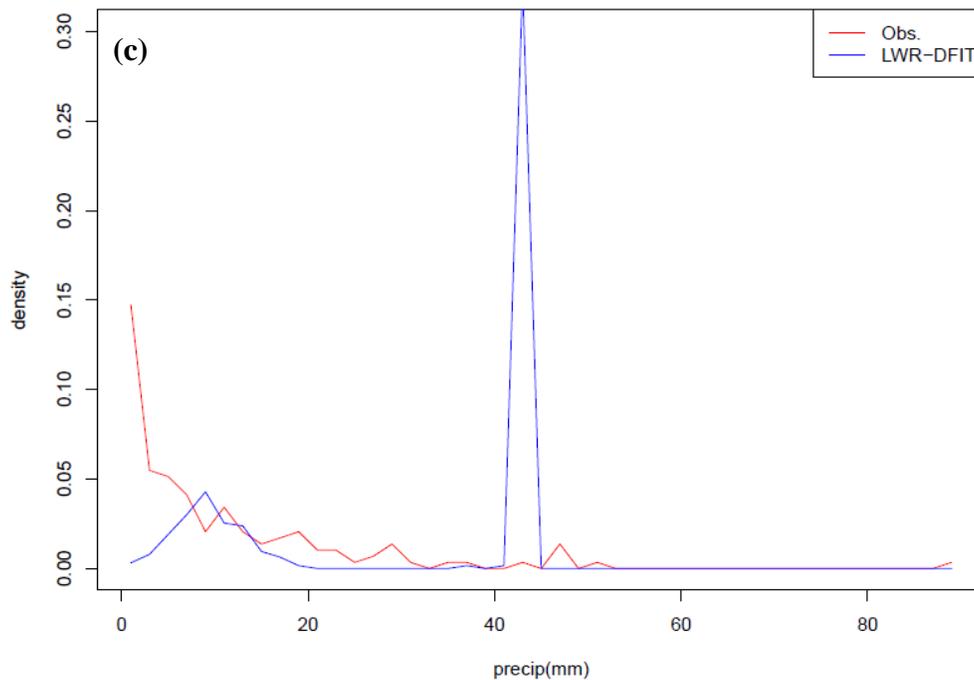
Comparison of Nonzero Rainfall PDF
Station: 283516 Month: 9



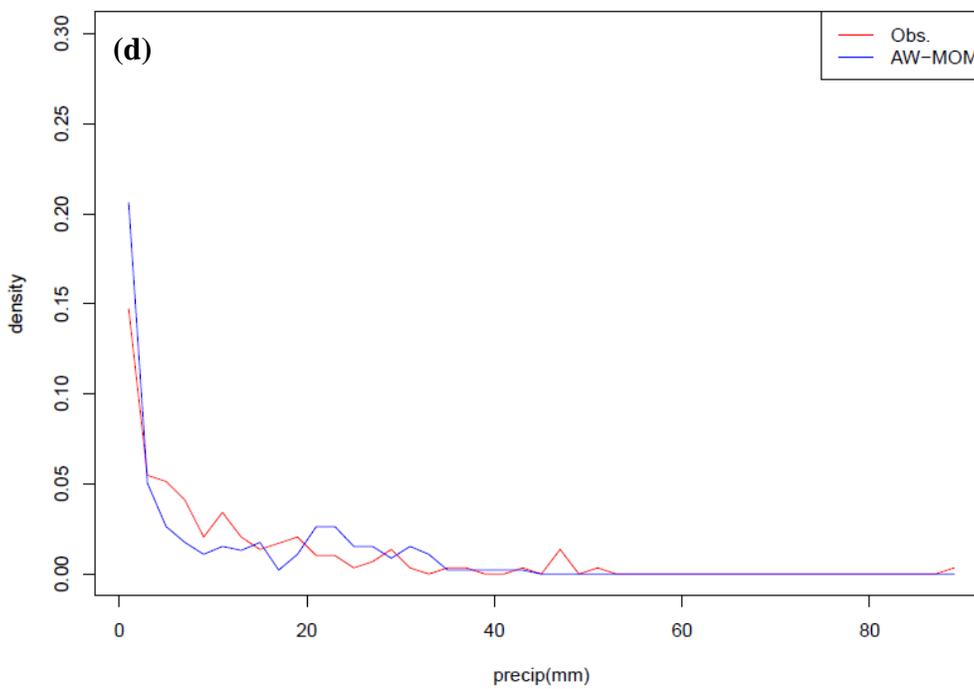
Comparison of Nonzero Rainfall PDF
Station: 283516 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 283516 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 283516 Month: 9



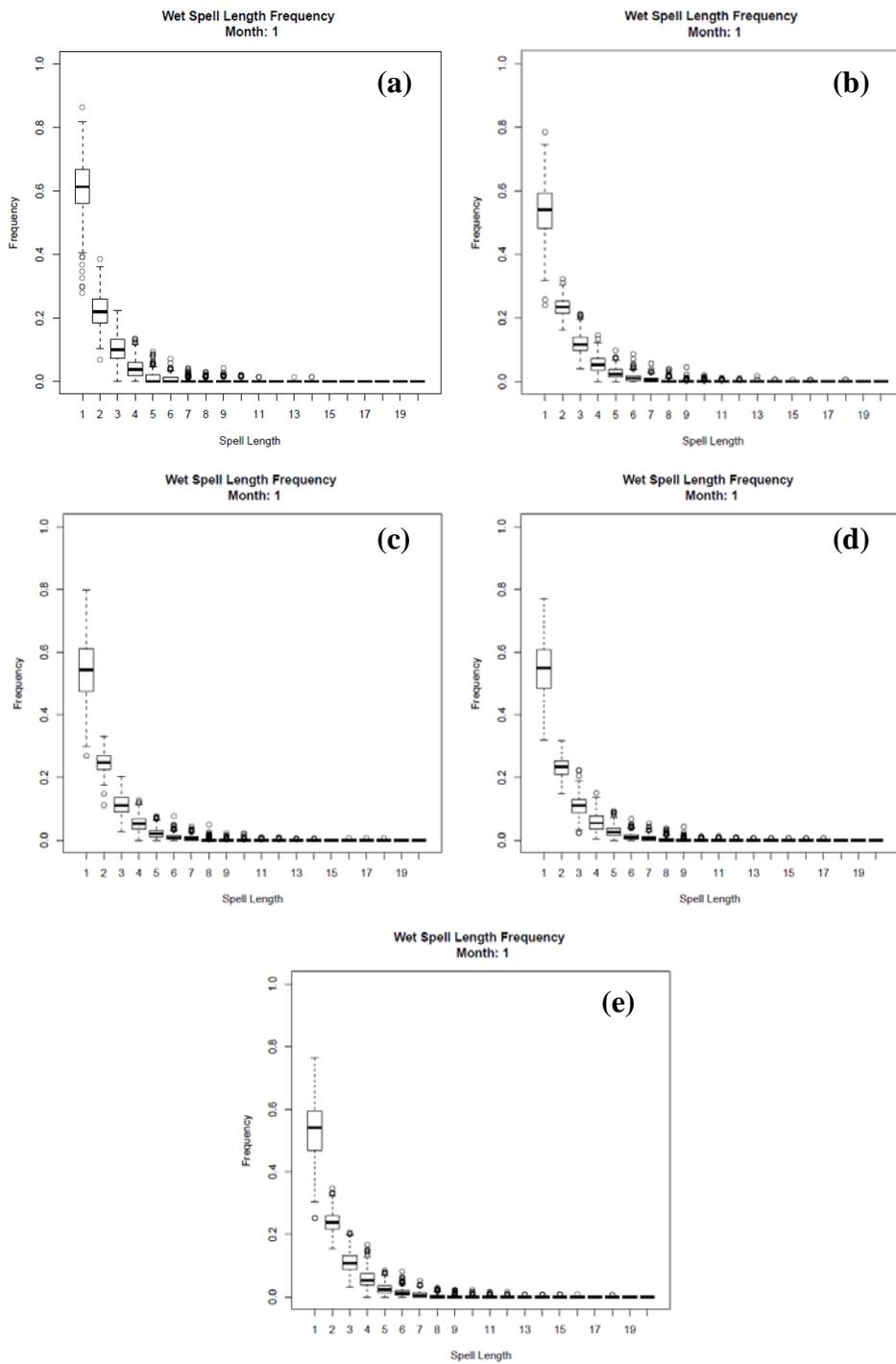


Figure 4.19. Boxplots of January Wet Spell Frequencies for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).

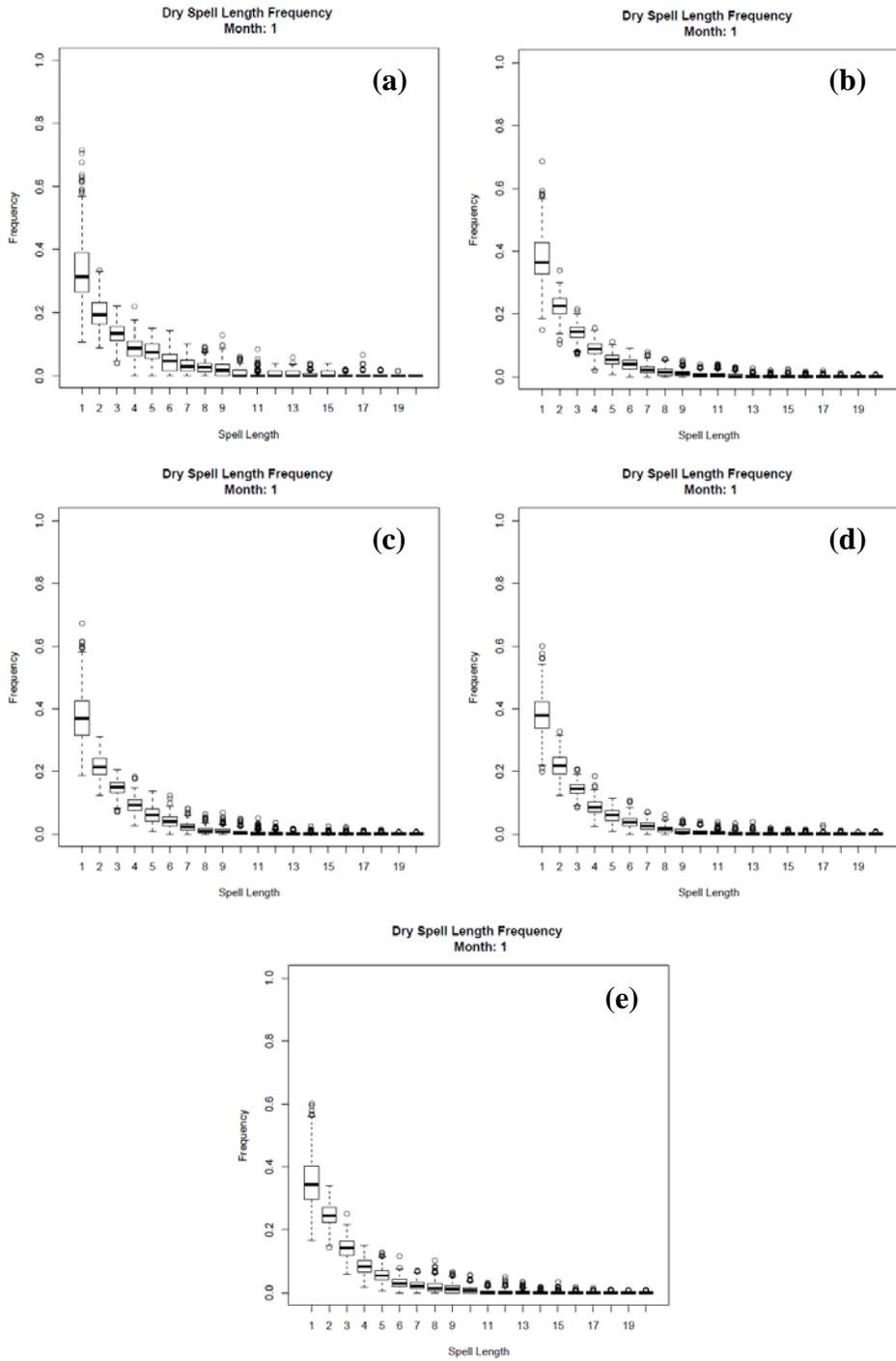


Figure 4.20. Boxplots of January Dry Spell Frequencies for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).

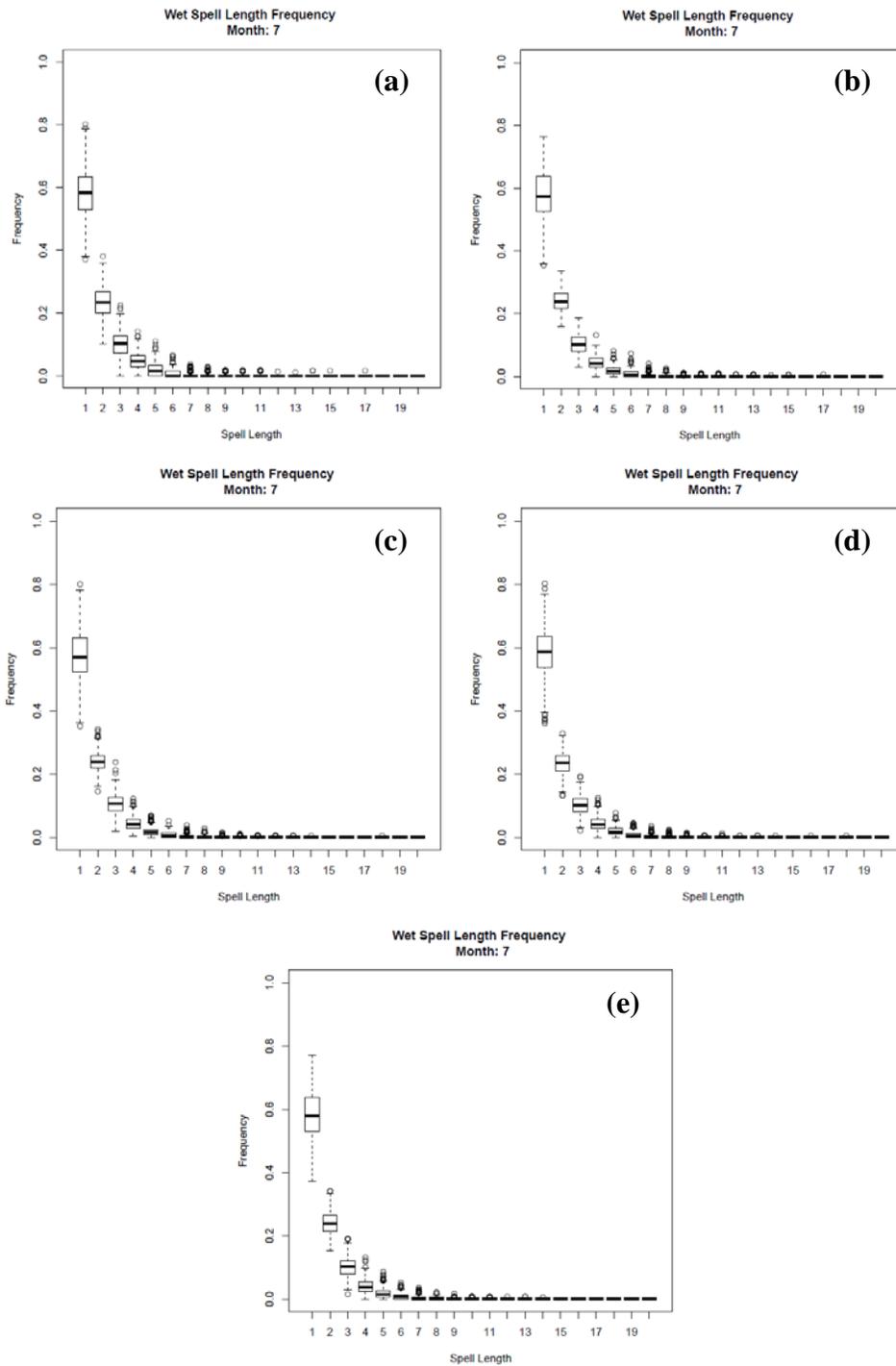


Figure 4.21. Boxplots of July Wet Spell Frequencies for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).

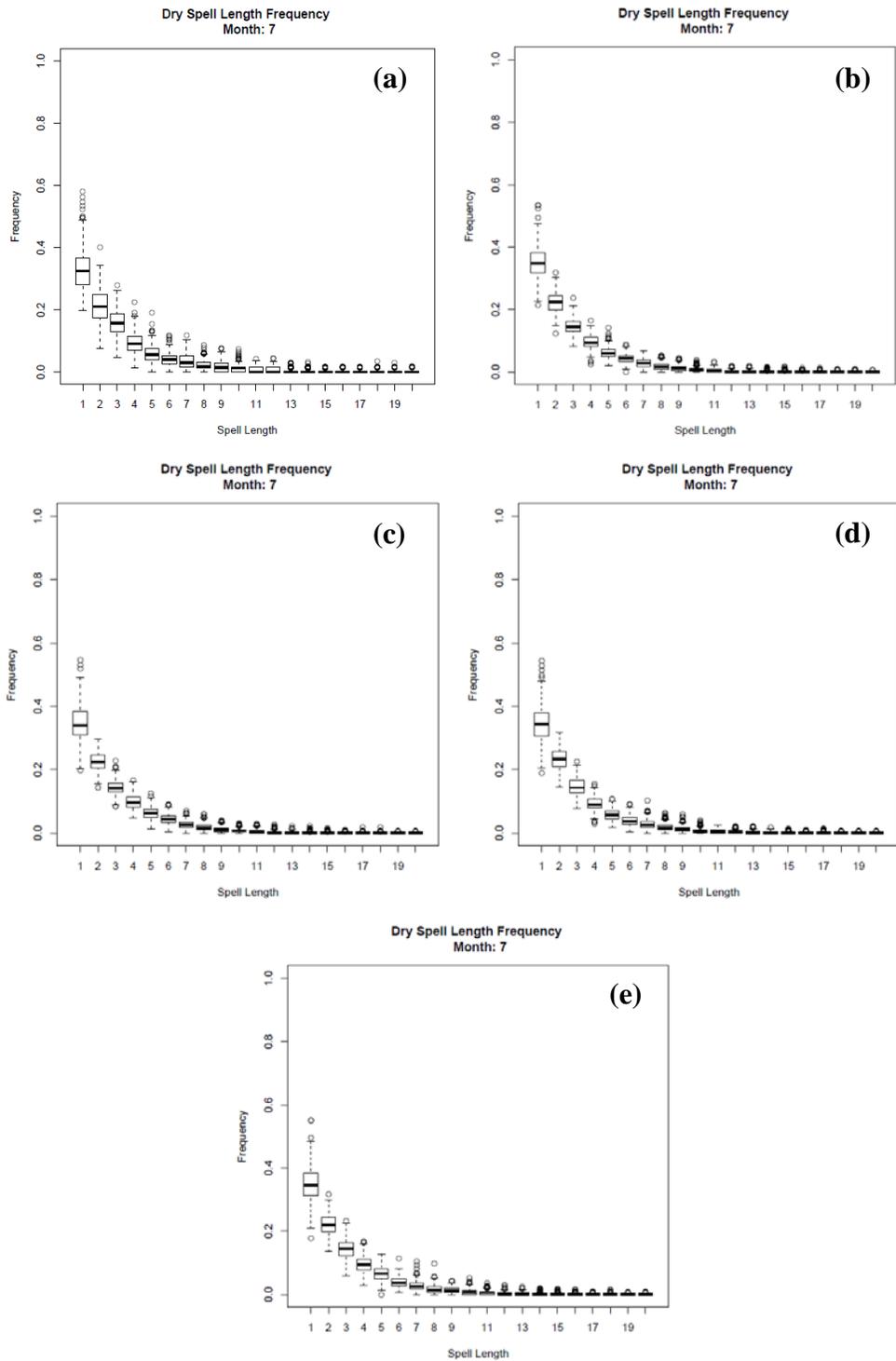


Figure 4.22. Boxplots of July Dry Spell Frequencies for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).

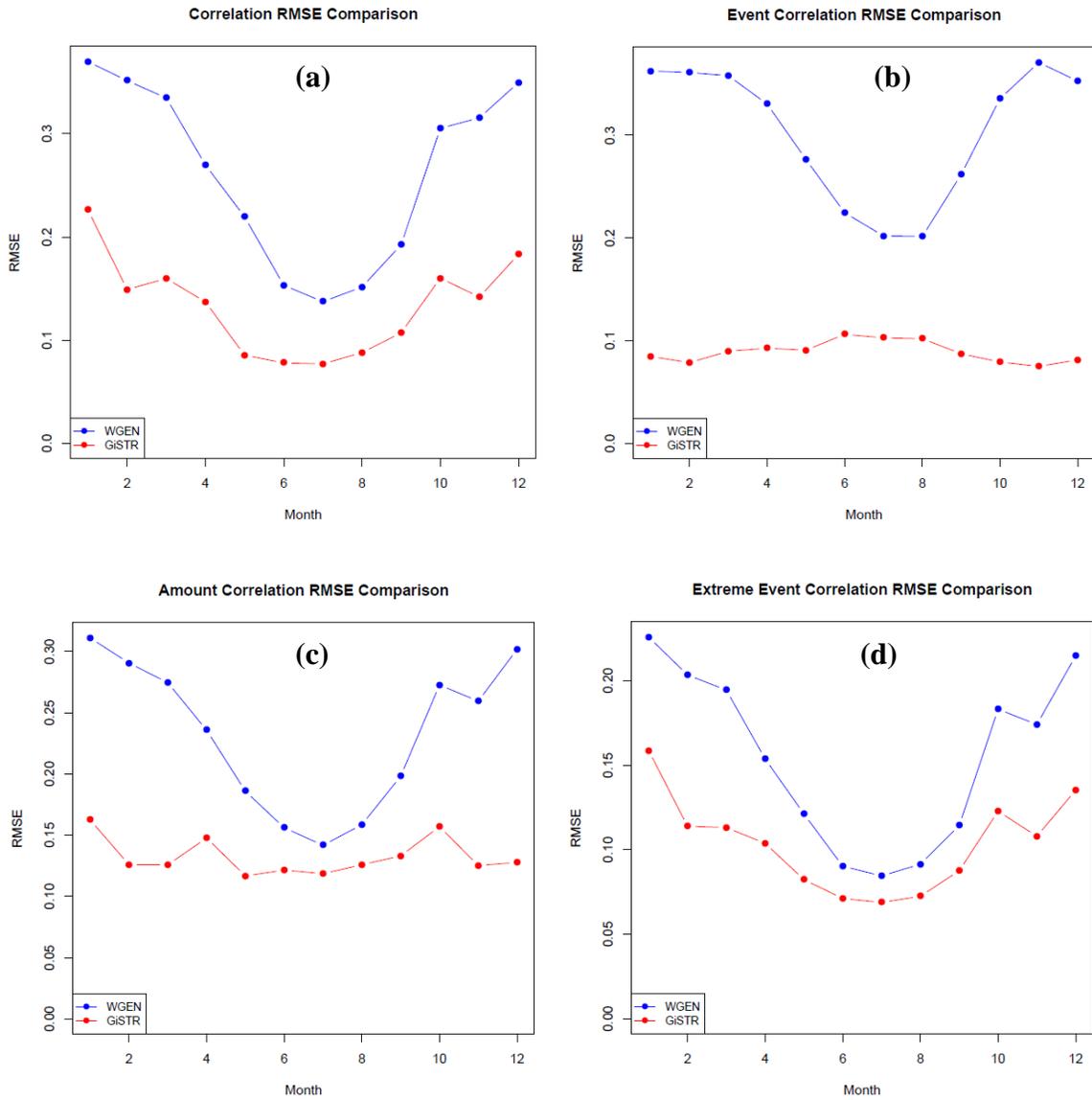


Figure 4.23. RMSE of the WGEN and GiSTR simulations by month for the correlation matrix of precipitation (ρ) (a), the correlation matrix of precipitation events (ρ_{ev}) (b), the correlation matrix of precipitation amounts (ρ_{am}) (c), and the correlation matrix of precipitation extreme events (ρ_{ex}) (d).

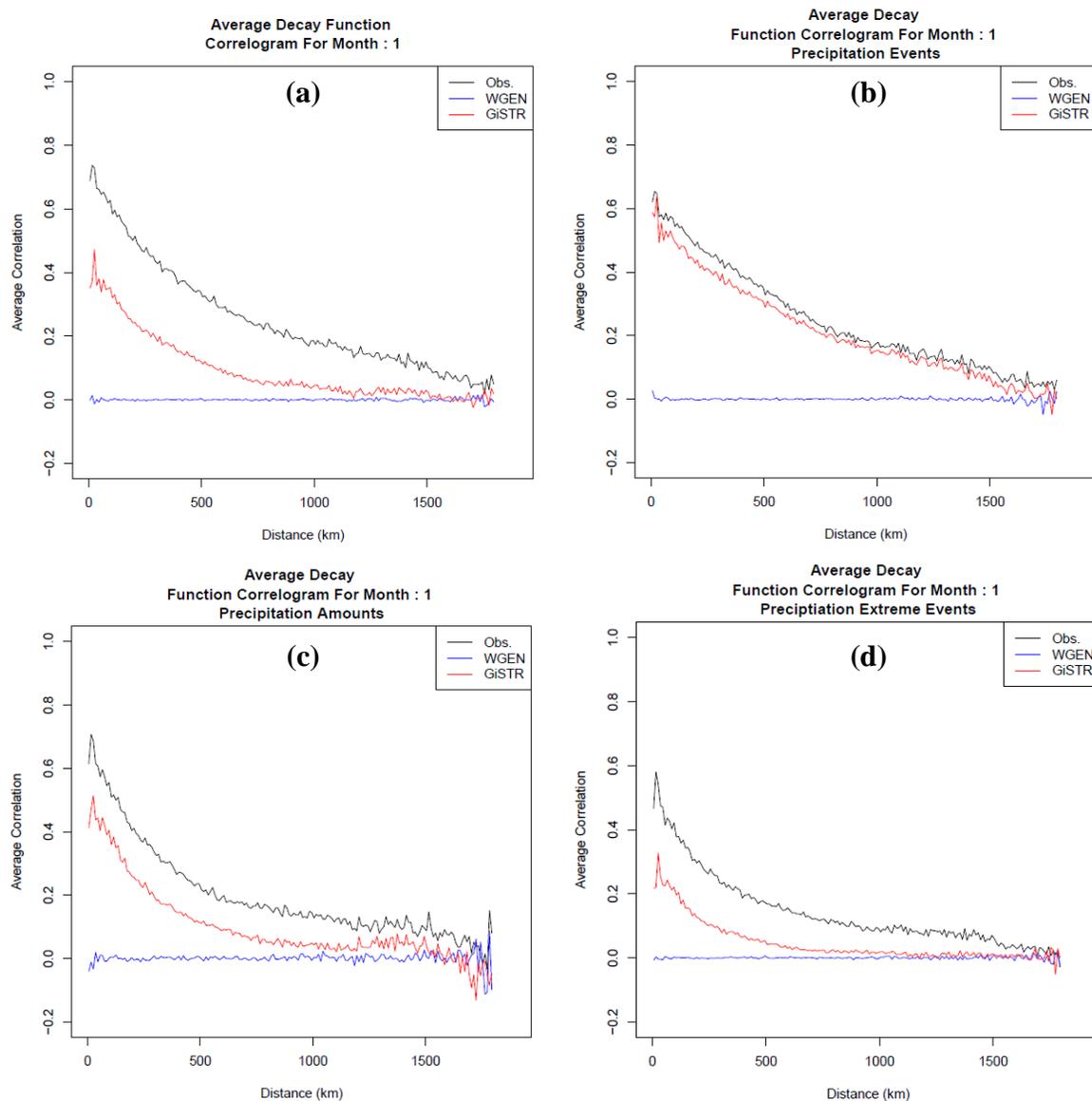


Figure 4.24. Average Decay Function Correlogram for WGEN and GiSTR compared to observations for January for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 1979-2000 time period.

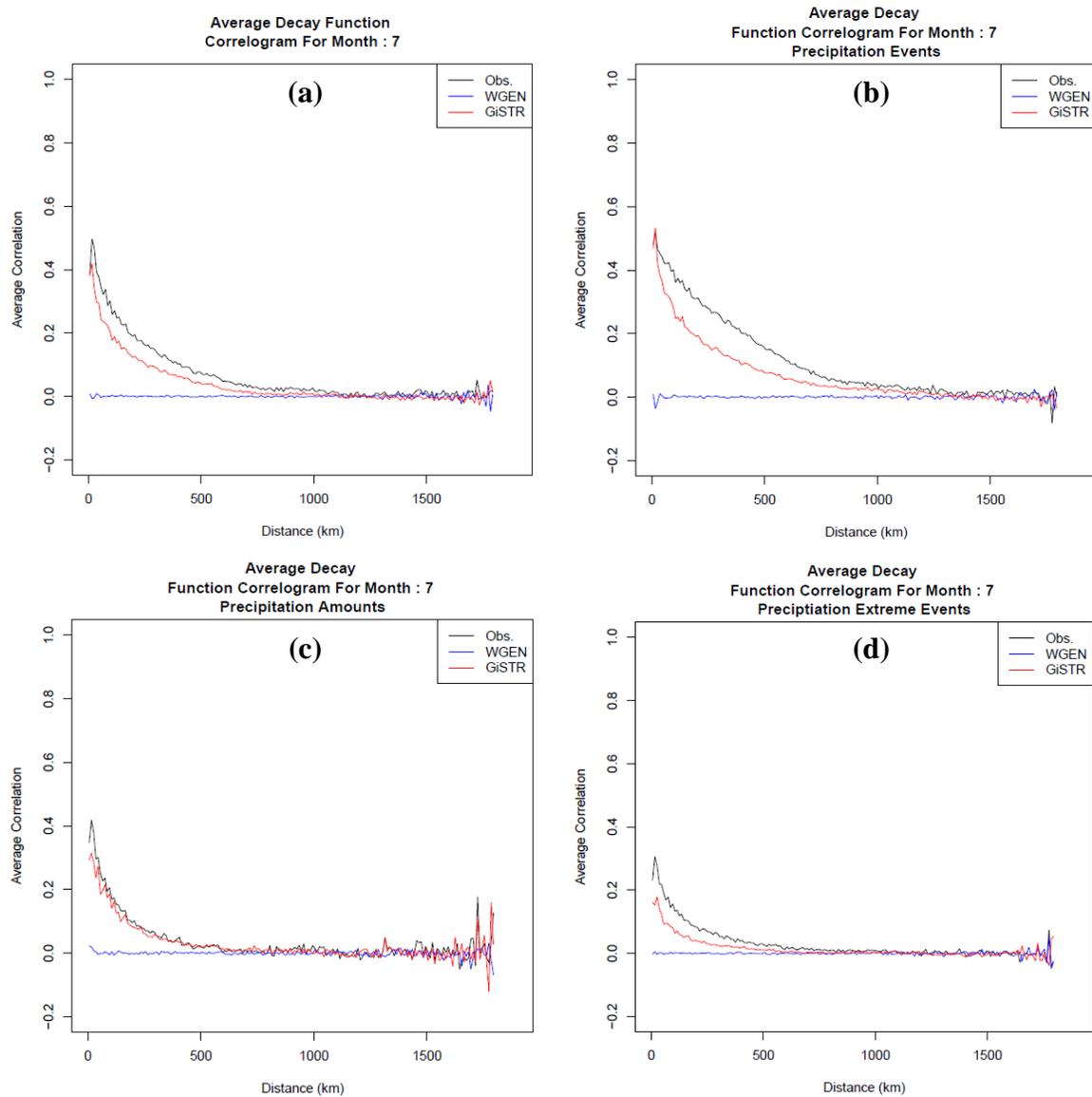


Figure 4.25. Average Decay Function Correlogram for WGEN and GiSTR compared to observations for July for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 1979-2000 time period.

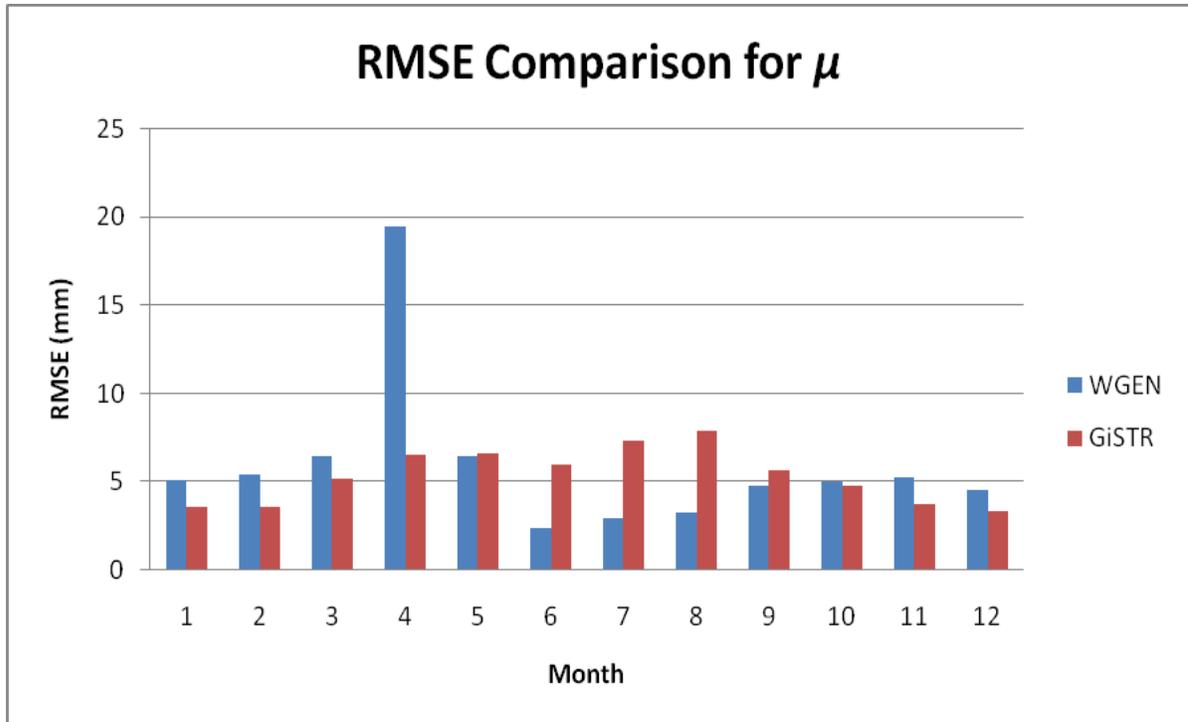


Figure 4.26. RMSE Comparison between downscaling simulations for the mean daily nonzero precipitation (μ).

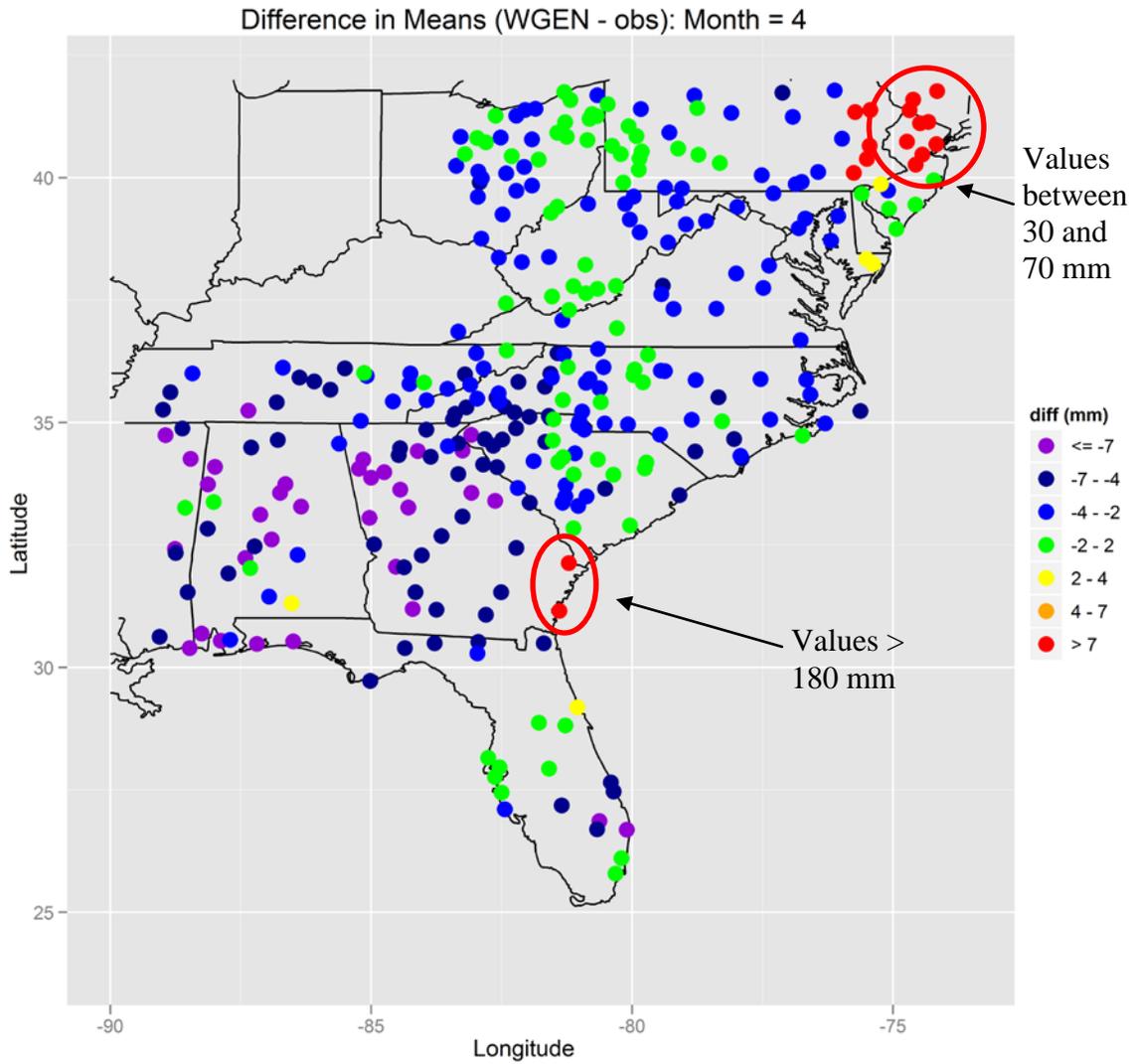


Figure 4.27. Difference between generated and observed values of the mean daily nonzero precipitation (μ) for the WGEN simulation in April.

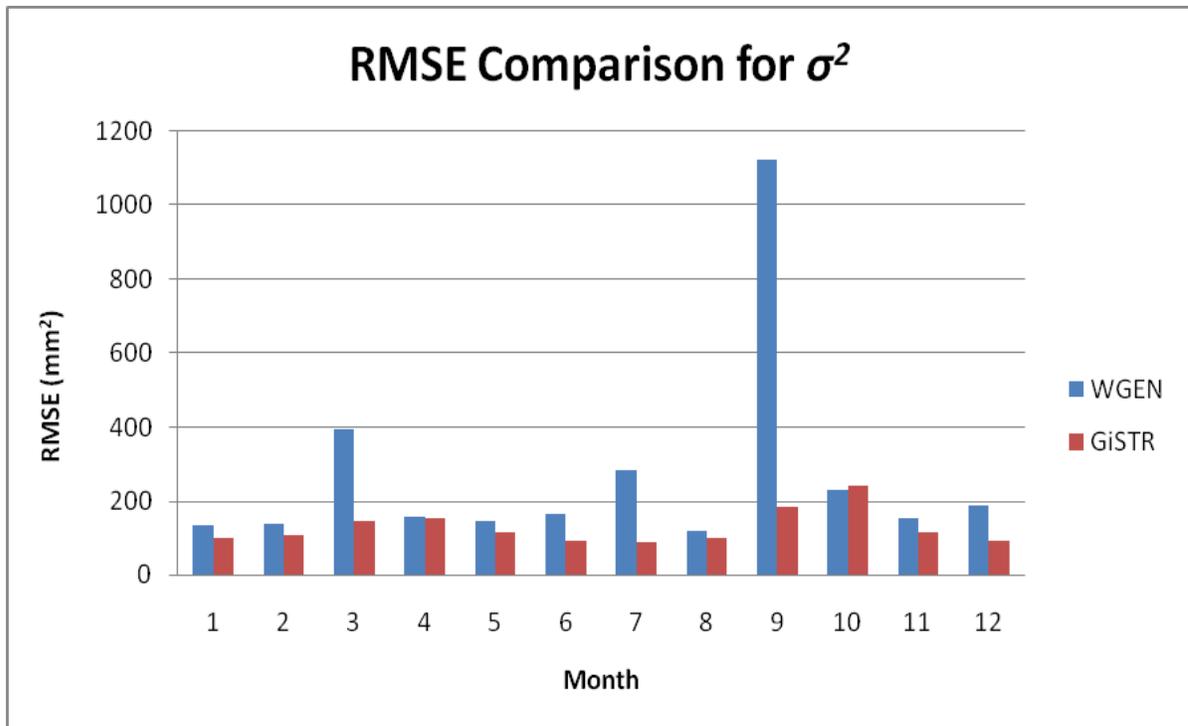


Figure 4.28. RMSE Comparison between downscaling simulations for the variance of the daily nonzero precipitation (σ^2).

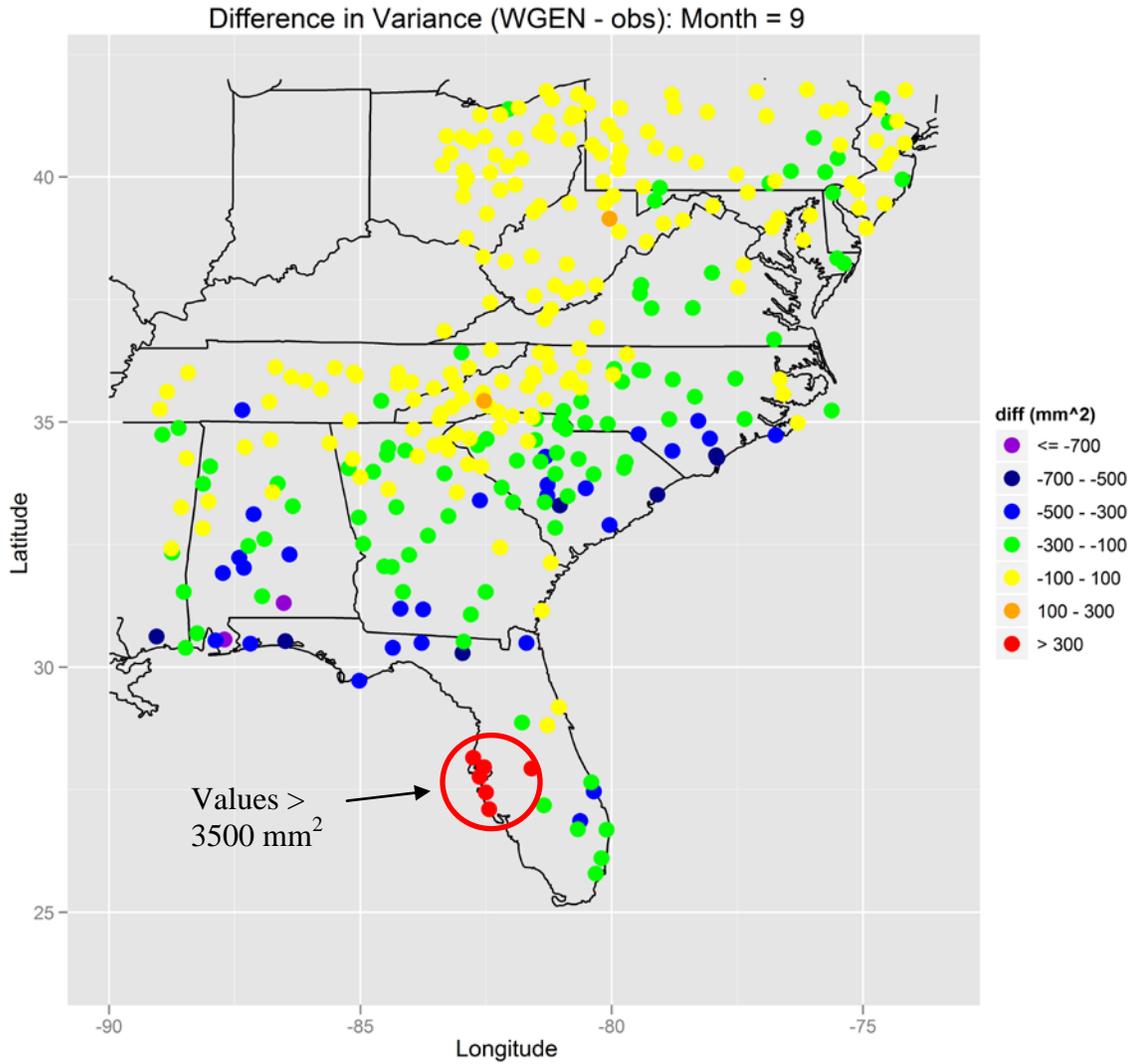


Figure 4.29. Difference between generated and observed values of the variance of the daily nonzero precipitation (σ^2) for the WGEN simulation in September.

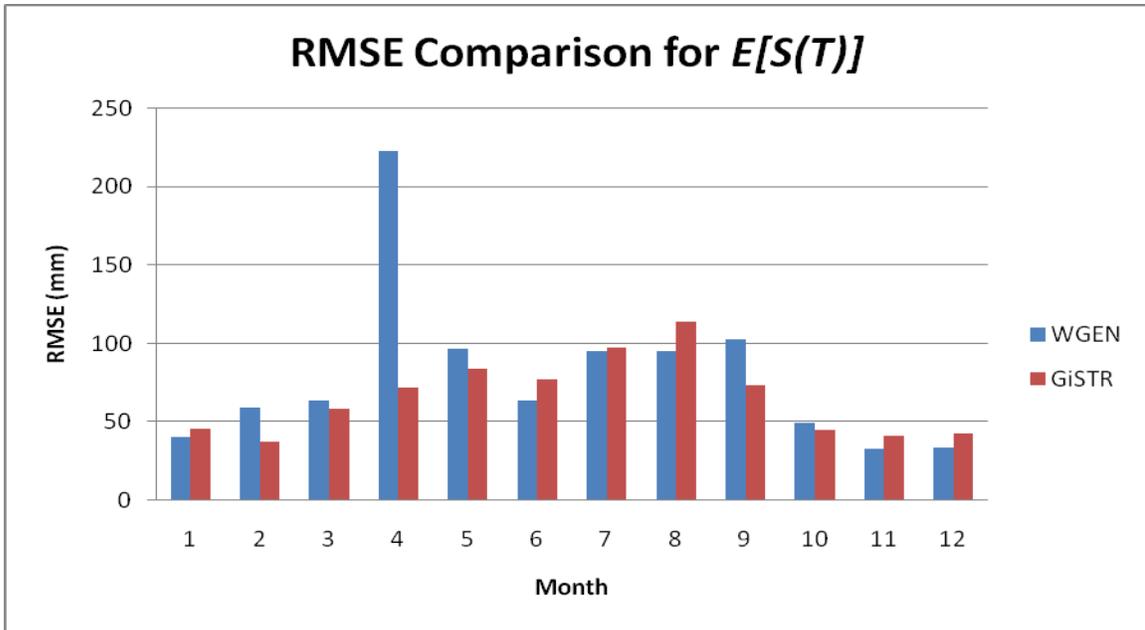


Figure 4.30. RMSE Comparison between downscaling simulations for the average total precipitation ($E[S(T)]$).

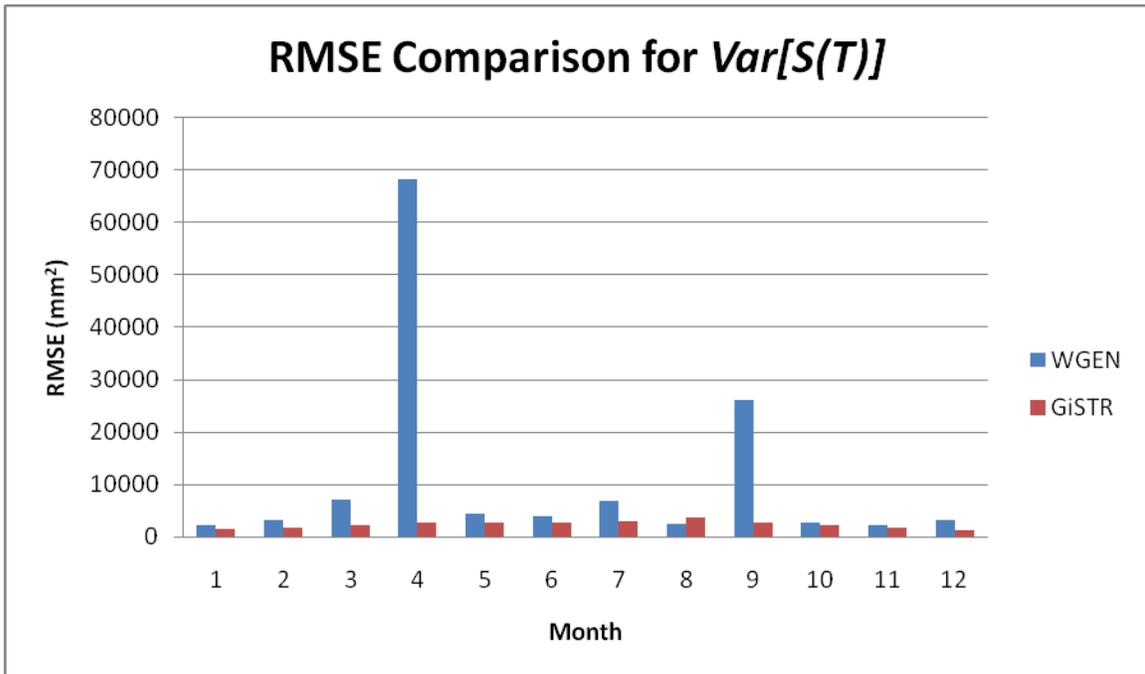


Figure 4.31. RMSE Comparison between downscaling simulations for the inter-annual variability ($Var[S(T)]$).

**Comparison of Nonzero Rainfall PDF
Station: 097847 Month: 4**

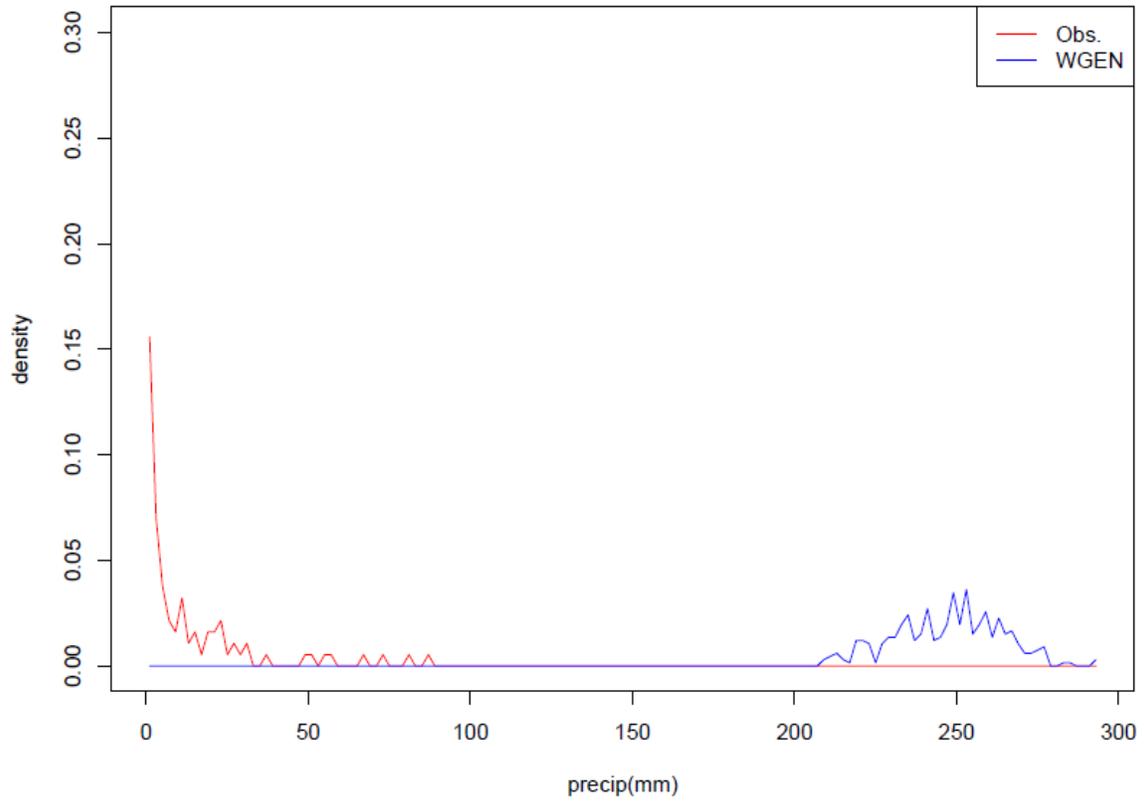


Figure 4.32. Difference between generated and observed PDF of nonzero precipitation for station 097847 in Savannah, GA for the WGEN simulation in April.

**Comparison of Nonzero Rainfall PDF
Station: 080945 Month: 9**

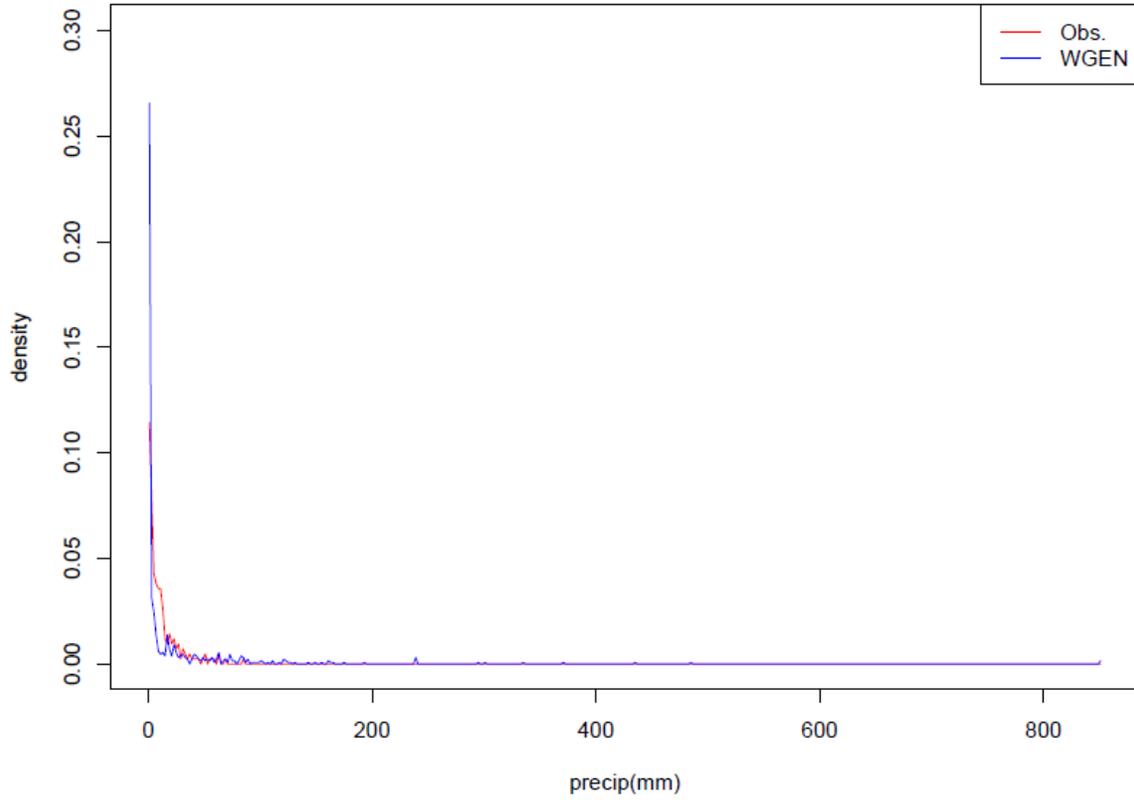


Figure 4.33. Difference between generated and observed PDF of nonzero precipitation for a station 080945 in Bradenton, FL for the WGEN simulation in September.

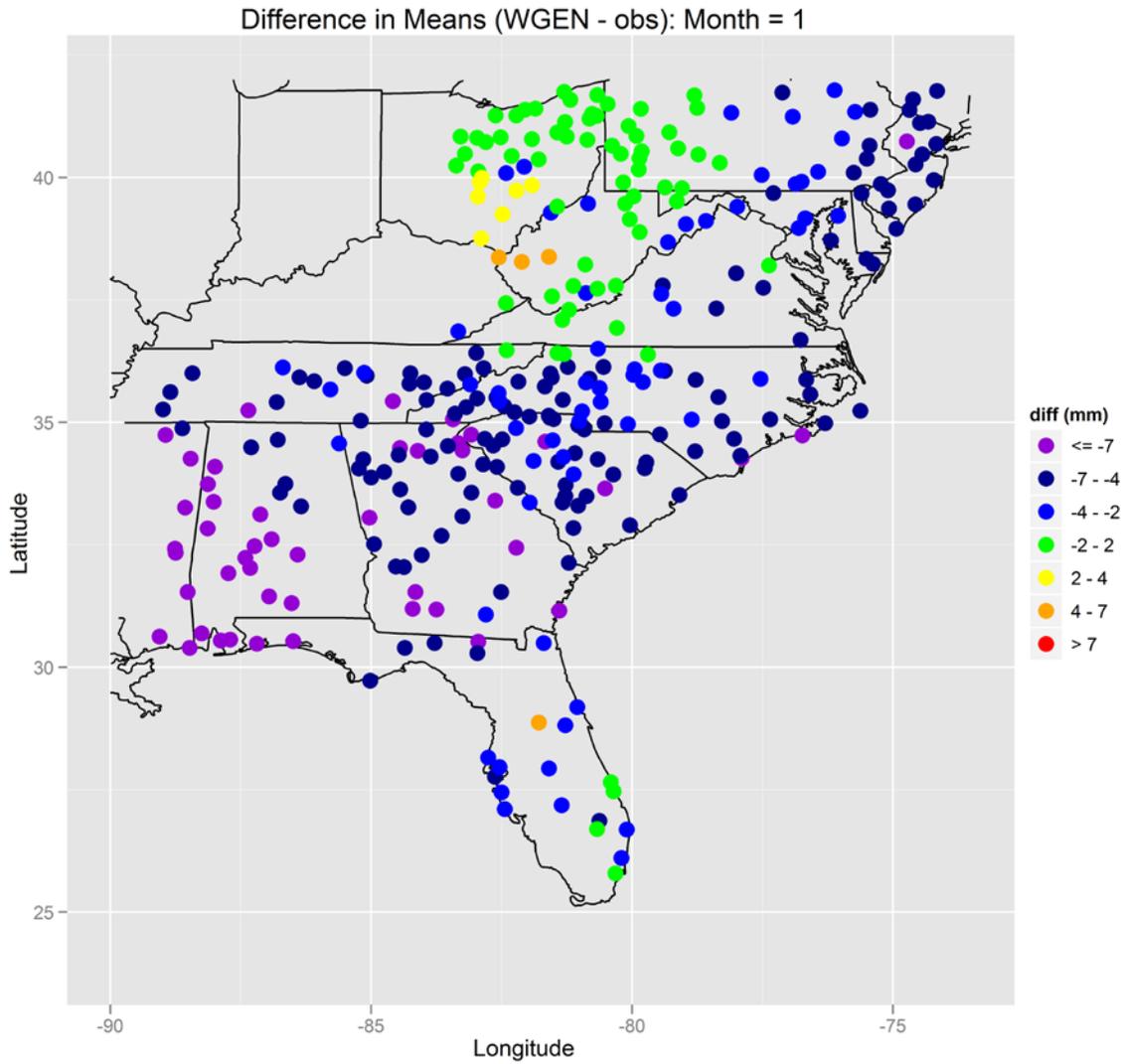


Figure 4.34. Difference between generated and observed values of the mean daily nonzero precipitation (μ) for the WGEN downscale simulation in January.

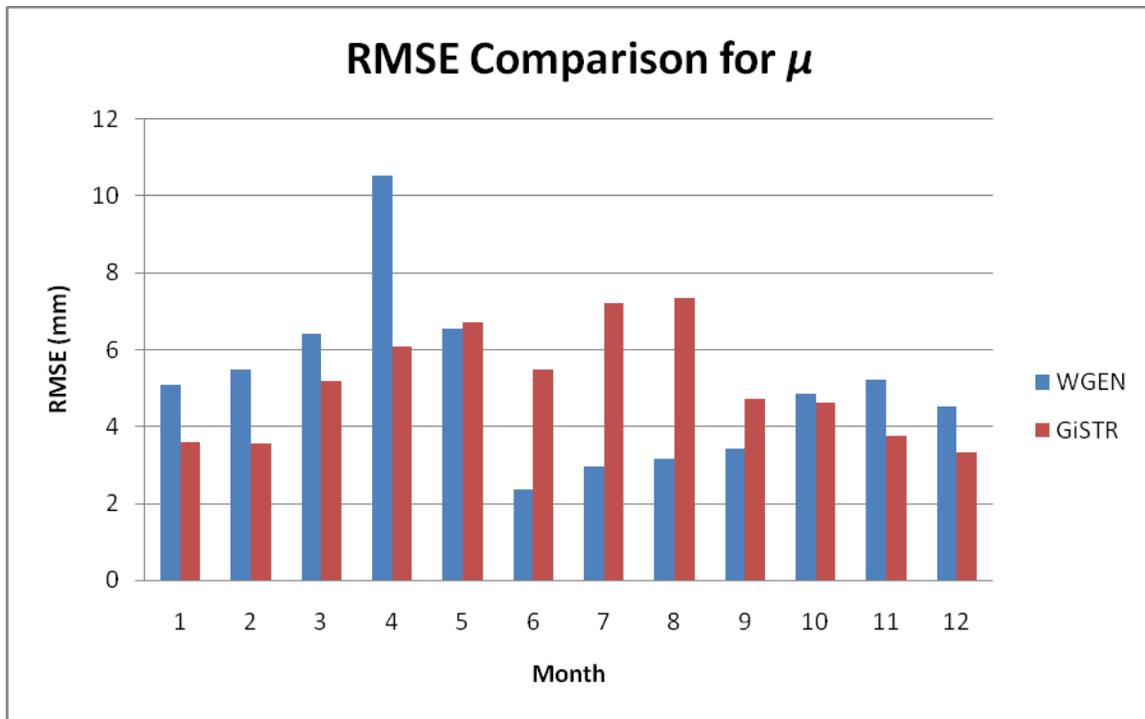


Figure 4.35. RMSE Comparison between downscaling simulations for the mean daily nonzero precipitation (μ) with outliers removed.

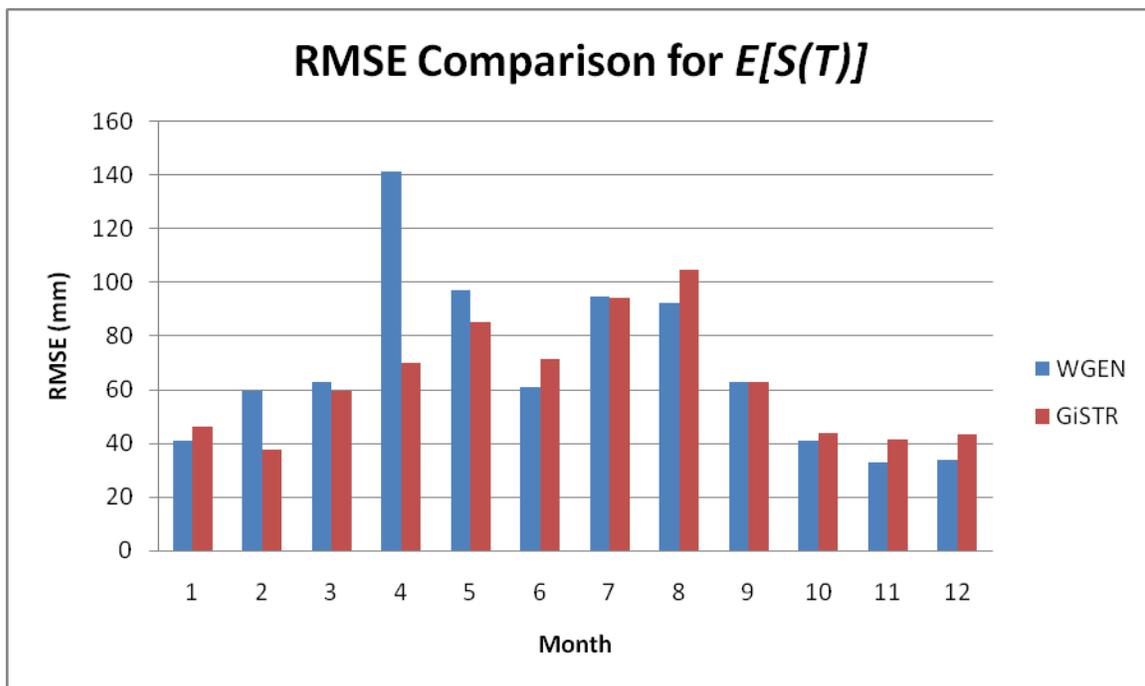


Figure 4.36. RMSE Comparison between downscaling simulations for the average total precipitation ($E[S(T)]$) with outliers removed.

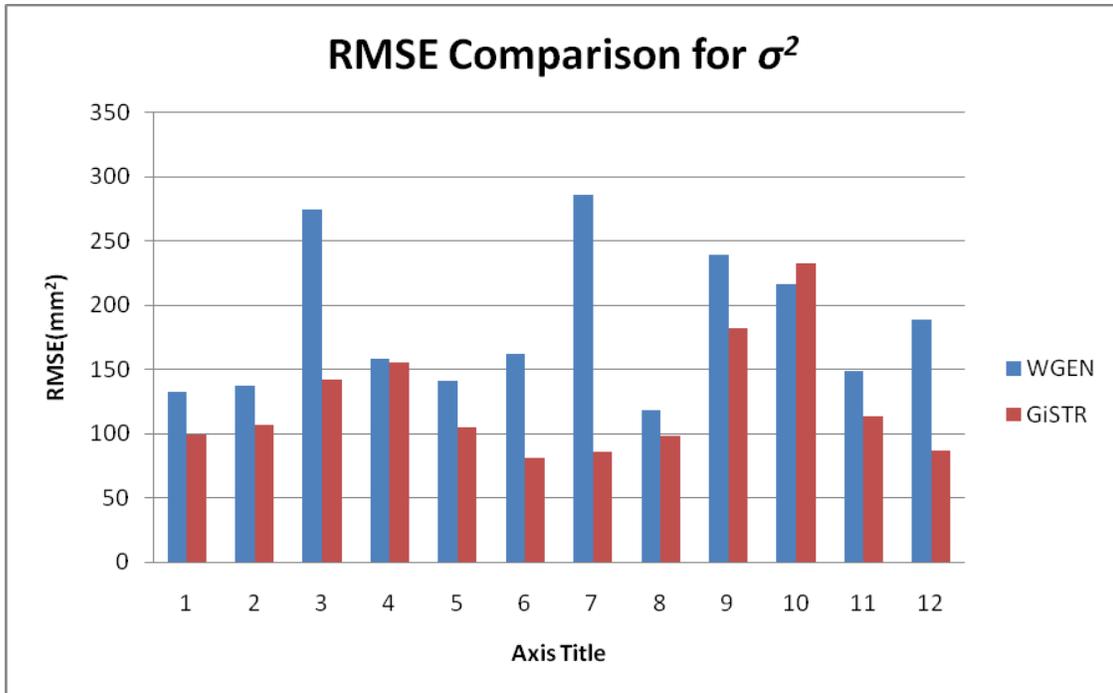


Figure 4.37. RMSE Comparison between downscaling simulations for the variance of the daily nonzero precipitation (σ^2) with outliers removed.

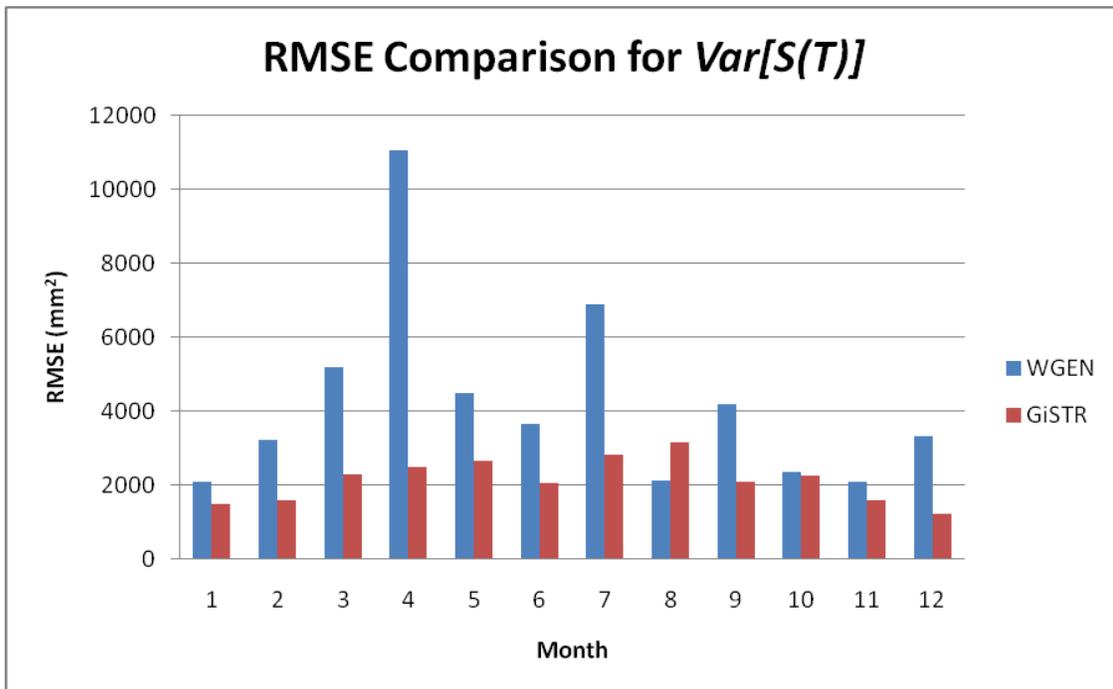


Figure 4.38. RMSE Comparison between downscaling simulations for the inter-annual variability ($Var[S(T)]$) with outliers removed.

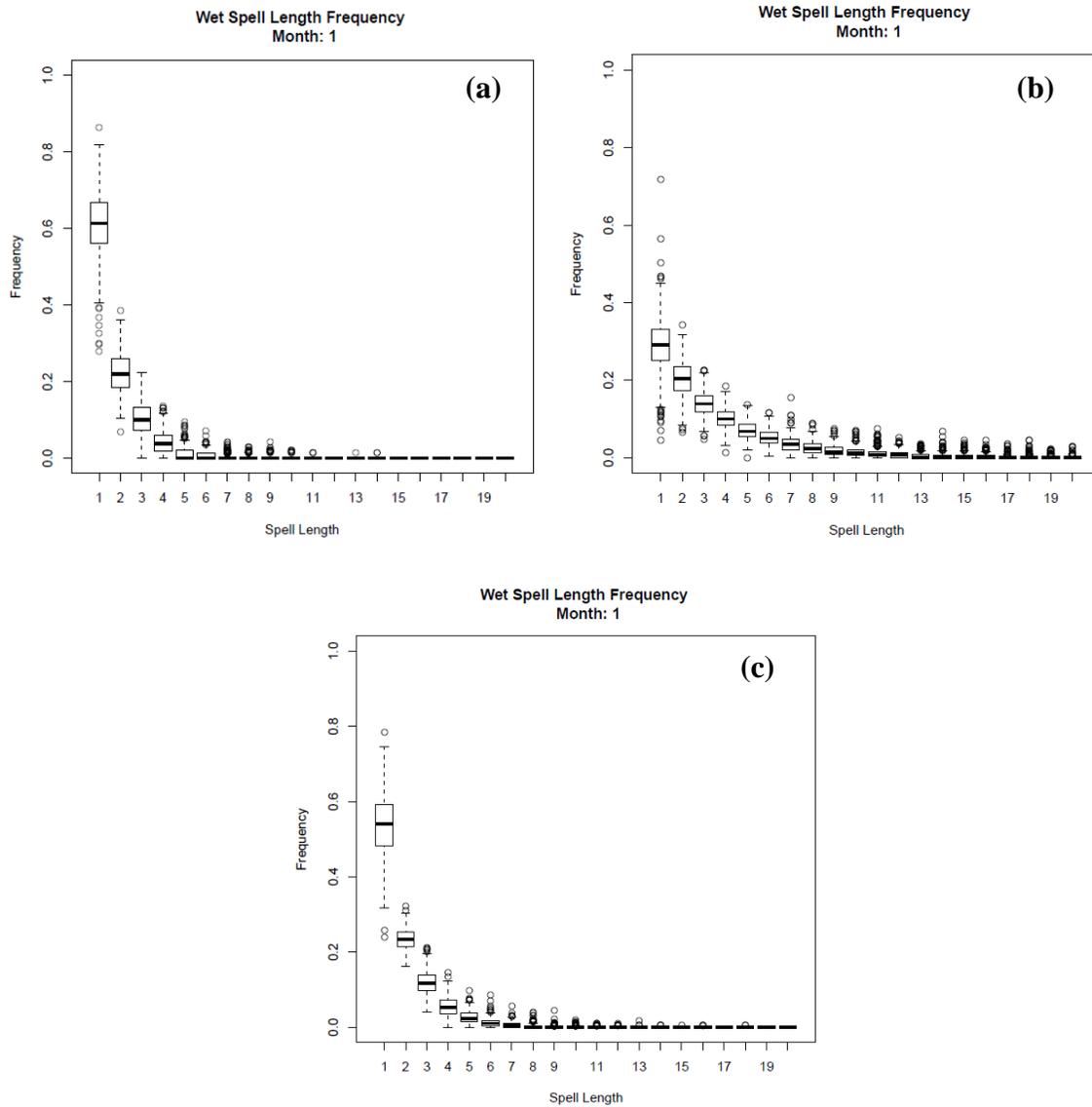


Figure 4.39. Boxplots of the frequency of wet spells of various lengths for January for observations (a), WGEN simulation (b), and GiSTR simulation (c).

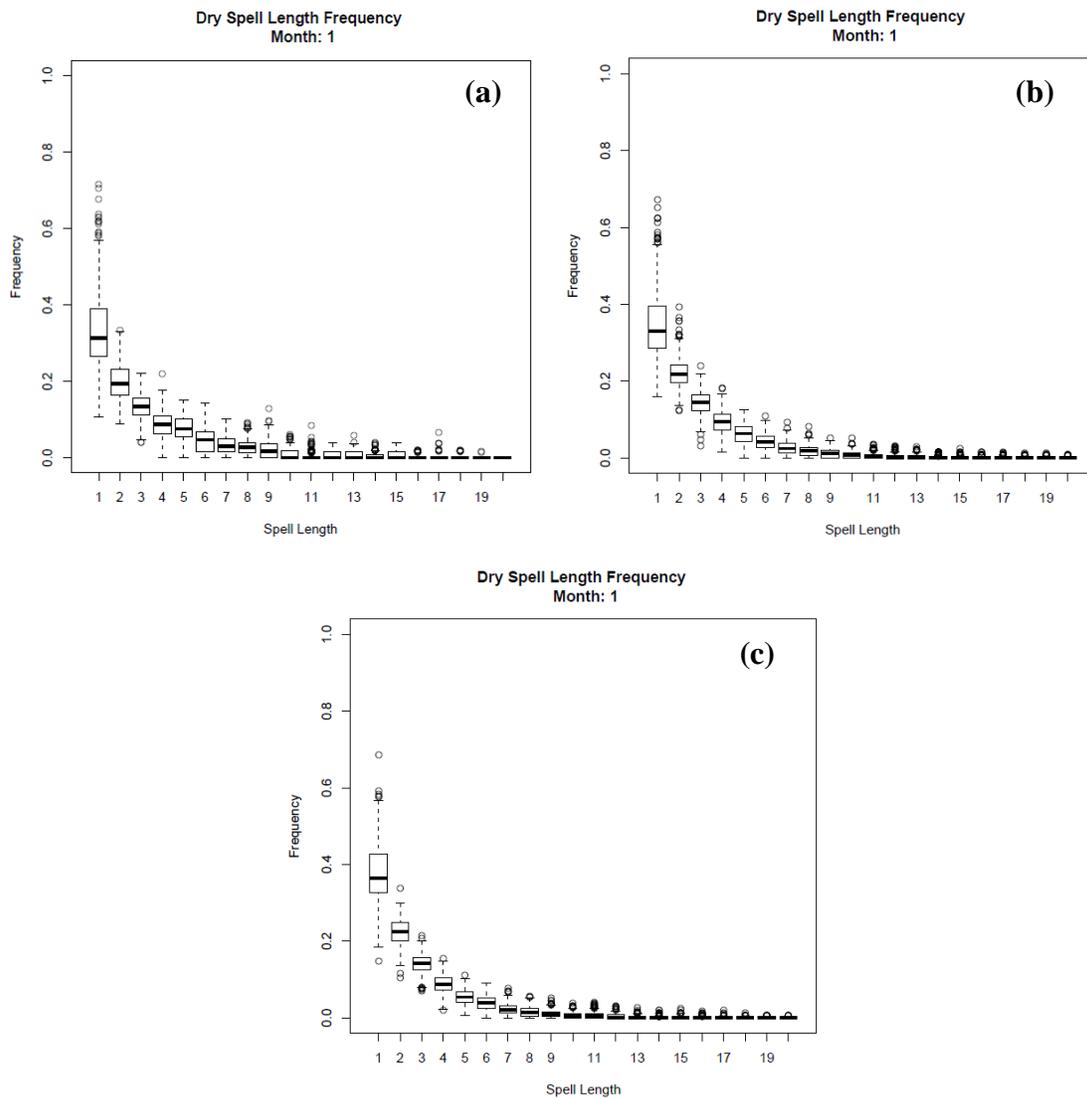


Figure 4.40. Boxplots of the frequency of dry spells of various lengths for January for observations (a), WGEN simulation (b), and GiSTR simulation (c).

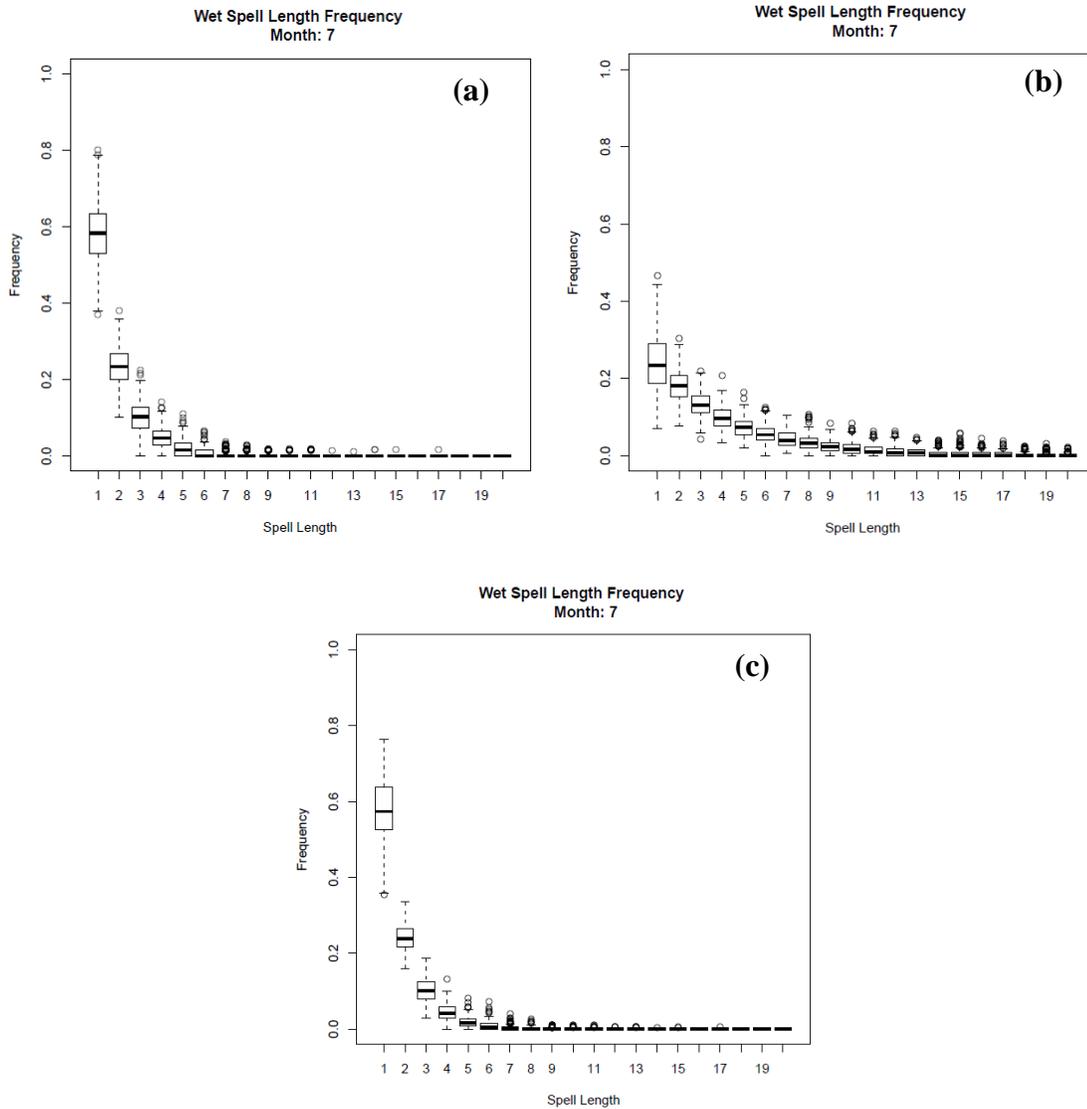


Figure 4.41. Boxplots of the frequency of wet spells of various lengths for July for observations (a), WGEN simulation (b), and GiSTR simulation (c).

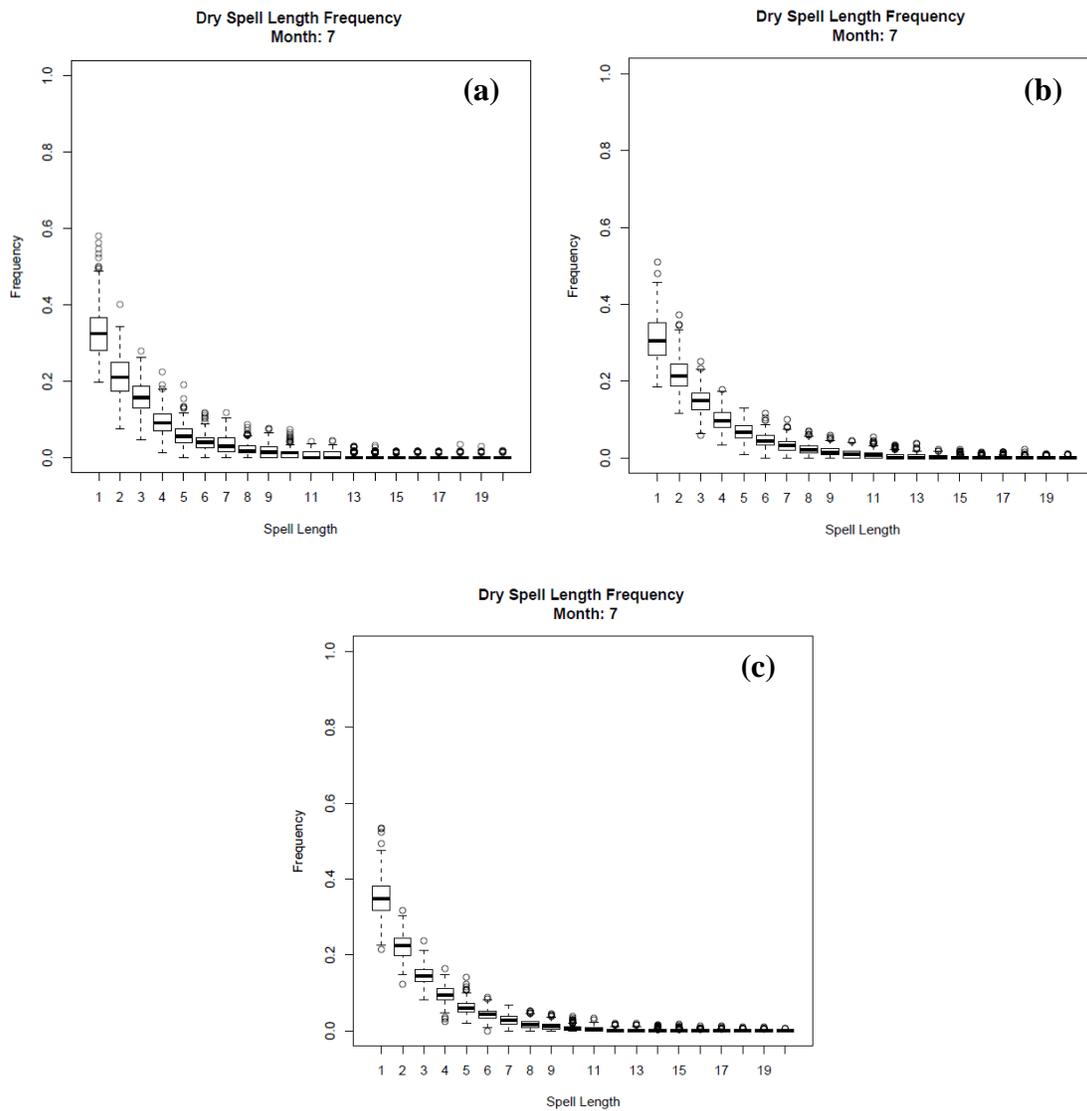


Figure 4.42. Boxplots of the frequency of dry spells of various lengths for July for observations (a), WGEN simulation (b), and GiSTR simulation (c).

Chapter 5. Evaluation of Downscaling Variations: 2001-2009

The previous chapter focuses on evaluating the difference between each of the variations of the two key assumptions of downscaling with weather generators, and the difference between two different generators used in downscaling with a fixed set of assumptions. However, that analysis is in the same time period as the observations used. The analysis presented here focuses on 2001-2009 as a forecast period.

The downscaling simulations in this period do not make use of the observations from the different stations in this period. Therefore, the analysis in this chapter shows the potential error associated with downscaling in a “forecast mode” rather than during a calibration period. For all analyses in this time period, the CFSR data was again used as the GCM, since the CFS Reforecast data was not available at the time of the analysis. Therefore, while this chapter will provide insight into the potential error associated with using any of the techniques described for seasonal forecast downscaling, it is recommended that the analysis in this chapter be repeated with whichever GCM is used for seasonal forecasting over the region.

Finally, while the 319 stations used in the previous analysis could be used for this evaluation, not all of these stations have sufficient data in 2001-2009 to be used for cross-validation. For all the evaluations in this chapter, 216 stations of the original 319 used in the previous chapter are used (as shown in Figure 5.1). These stations have less than 10% of the observations missing in 2001-2009 and are again checked for quality. These stations are used in an evaluation of the different downscaling variations and the difference between weather generators when used for downscaling. For analysis of downscaling variations, GiSTR is the generator used for the downscaling. For analysis of generator differences, the downscaling assumptions use dare held constant to compare GiSTR and WGEN. As with chapter 4, WGEN is used because it is used as a standard for building Markov chain generators, and GiSTR is used since it better replicates the variability associated with extremes (which is a historical problem for weather generators, Wilks and Wilby, 1999) and better replicates the spatial structure as compared to WGEN.

5.1. Downscaling Simulation Comparison

While many of these results are consistent with results in the previous chapters the evaluation in this chapter shows the difference between time periods as well as exaggerated differences between downscaling simulations.

5.1.1. Temporal Parameters

The temporal parameters, which influence precipitation occurrence, were produced similarly by each downscaling variation for the 2001-2009 time period. For the Markov transition probabilities (P_{01} and P_{11}), the difference in RMSE between simulations is less than 0.004 on average. For the RMSE of P_{01} , AW-DFIT and LWR-MOM have smaller RMSE than LWR-DFIT and AW-MOM across the domain. The comparison of the RMSE of P_{01} for each simulation and all months, shown in Table 5.1, shows that there is no seasonal trend for each simulation and that the LWR-MOM and AW-DFIT have smaller RMSE than the remaining simulations for most months. Similarly, for the RMSE of P_{11} , AW-DFIT also has the smallest RMSE on average. Table 5.2 shows a comparison of the RMSE of P_{11} for each simulation and all months. The results of this comparison show that there is also no seasonal trend in the RMSE of P_{11} for each simulation. This comparison also shows that the AW-DFIT and LWR-MOM have smaller RMSE than the remaining simulations for most months. The two-sample T-test results for the RMSE of both parameters between simulations, summarized in Table 5.3, shows that every test has a p-value greater than 0.183 and a difference less than 0.004 for most tests. These additional two-sample T-tests for the RMSE of each parameter between simulations show that there is no significant difference between simulations for a level of significance of 0.05. Each of these results are consistent with the results for the Markov Transition Probabilities in Section 4.1.1 which also indicates that the LWR-MOM and AW-DFIT simulations have smaller error than the remaining simulations for these parameters.

Given that errors in this time period are representative of potential errors associated with use in a forecasting simulation, the error of each simulation in this time period are tested against the error of the matching simulations from the previous chapter. For instance, the RMSE of each parameter for the LWR-MOM variation in the 2001-2009 time period are tested against the RMSE of the same variation for the 1979-2000 time period. The results of these two-sample T-tests are summarized in Table 5.4. While there is no significant difference between simulations for this time period, comparing the individual 2001-2009 simulations to their counterparts from Chapter 4 shows that the RMSE of each parameter for downscaling simulations in the 2001-2009 time period is significantly larger than in the 1979-2000 time period. This is expected given that the observations used in the scaling relationship are for the 1971-2000 time period, not the 2001-2009 time period. No station observations in any analysis are used during the 2001-2009 time period.

Similar results are shown for unconditional probability of rain (π) and persistence (γ). The RMSE comparison between simulations for each month for values of π , shown in Table 5.5, indicates that there is no seasonal trend in the RMSE of π and that AW-DFIT and LWR-MOM have smaller RMSE than the remaining simulations for most months and on average. This comparison also shows that the difference between simulations is less than 0.004 for all months. Similarly, the AW-DFIT and LWR-MOM have smaller RMSE of γ on average than the LWR-DFIT and AW-MOM simulations. The RMSE comparison between simulations for each month for values of γ , shown in Table 5.6, indicates that there is no seasonal trend in the RMSE of γ for any simulation. This comparison also indicates the difference between simulations is less than 0.004 for all months and that the LWR-MOM and AW-DFIT has the smaller RMSE for most months and on average than the LWR-DFIT and AW-MOM simulations. Given that the difference between simulations for both parameters is less than 0.004, the two-sample T-test for π and γ , summarized in Table 5.7, is performed to determine significant difference. Similarly to the results for the P_{01} and P_{11} , the two-sample T-test shows that there is no significant difference in the RMSE of π or γ between the 2001-2009 downscaling simulations. The results are also consistent with the results shown in Chapter 4,

which indicates that the LWR-MOM and AW-DFIT have lower error than the remaining simulations for both π and γ .

The RMSE of all four parameters for each 2001-2009 simulation are also tested using the two-sample T-test against the RMSE of the corresponding 1979-2000 simulations in Chapter 4. The results of these tests, summarized in Table 5.8, indicates that the RMSE for each parameter for each of the 2001-2009 simulations is significantly larger than the RMSE of the same parameters for their corresponding simulations from Chapter 4. These results make sense given the direct relationship between P_{01} , P_{11} , π , and γ as described by Equations 2.5 and 2.6. However, the fact that the RMSE of all the temporal parameters increases in this “forecast” time period, indicates the potential error associated with forecasting these parameters. Since these parameters reflect the occurrence of precipitation events, it is likely that this increased error would impact the forecasted number of precipitation events in the region. Regardless, the AW-DFIT and LWR-MOM have lower error for each of the temporal parameters in the Southeast U.S than the LWR-DFIT and AW-MOM simulations.

5.1.2. Spatial Structure Parameters

Following the procedure from Chapter 4, the spatial structure parameters are not altered by CFSR. However, given that the affect of CFSR on other parameters (specifically P_{01} and P_{11}) and that how GiSTR generates precipitation events is influenced by those same parameters, the spatial structure produced by GiSTR is also affected. Therefore, each of the simulations may represent the spatial structure of precipitation differently. For each of the simulations, the spatial structure parameters were held fixed to represent the observed spatial structure from the period of 1971-2000. This assumes that the spatial structure between individual stations is not very different between the two periods. The RMSE for each simulation for the correlation matrix of precipitation (a, ρ), the correlation matrix of precipitation events (b, ρ_{ev}), the correlation matrix of precipitation amounts (c, ρ_{am}), and the correlation matrix of precipitation extreme events (d, ρ_{ex}) for all months is shown in Figure 5.2. For all the correlation matrices there is very little difference in the RMSE between each

simulation. This is also consistent with the average decay function correlogram produced by each simulation. Figure 5.3 show the observed average decay function correlogram compared to those created by each simulation for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d) for January. Each simulation shows a tendency to underestimate the observed correlogram at several distances, though the LWR-MOM simulation is slightly more accurate for precipitation amounts (Figure 5.3c) and precipitation extreme events (Figure 5.3d). This pattern is not shown by the correlograms created by each simulation for summer months. Figure 5.4 shows the observed average decay function correlogram compared to those created by each simulation for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d) for July. For all four parameters, each simulation produces a similar correlogram to every other simulation. However each simulation also shows a tendency to underestimate the observed correlogram for precipitation (Figure 5.4a), precipitation events (Figure 5.4b), and precipitation extreme events (Figure 5.4d) for stations less than 1000km apart. Similarly to results from Section 4.1.2, each simulation provides a slightly worse estimate of the observed correlogram of precipitation events compared to observed correlogram produced by GiSTR outside of downscaling throughout the year. This is likely related to the indirect influence of the downscaling variation on the spatial structure output by the simulations. These results are also consistent with the RMSE of the different correlation matrices and testing may reveal significant differences in the RMSE between each simulation.

The average RMSE for each correlation matrix is shown in Table 5.9. Testing the RMSE for each correlation matrix for each downscaling simulation shows that there is no significant difference in the RMSE between most downscaling simulations, which is consistent with the small differences in the average RMSE between simulations for each parameter indicated by Table 5.9. However, for specific downscaling simulations and correlation matrices there are instances where the RMSE is significantly larger or smaller than others. Using the two-sample T-test, AW-MOM is shown to have a significantly smaller RMSE for values of the correlation matrix of precipitation events (ρ_{ev}) than all other simulations (p-values of 0.02 to 0.03). This result of this analysis suggests a contrast to the

same comparison in the previous chapter. Section 4.1.2 suggests that the relationship between the errors of the temporal parameters might be used to determine the impact of these parameters on values of ρ_{ev} . However, the fact that the RMSE of AW-MOM is significantly smaller for this parameter while the RMSE of the temporal parameters is larger is in contrast to the earlier conclusion in Chapter 4. In addition, the testing also shows that LWR-DFIT has a significantly larger RMSE for values of the correlation matrix of precipitation amounts (ρ_{am}) than all other simulations (p-values of 0.003 to 0.03). These results are contradictory to the results presented in Chapter 4 for the same comparison. For all other parameters, the results of Chapter 4 show that there is no significant difference in the RMSE of any of the spatial structure parameters between simulations.

However, while the majority of the tests are consistent with the results presented in Chapter 4, the average RMSE is also similar between the 2001-2009 and 1979-2000 simulations. The only exception is RMSE for values of ρ_{am} , where the testing shows that the RMSE of all the 2001-2009 simulations for this parameter are significantly larger than for the 1979-2000 simulations (p-values of $2E-9$ to $2E-5$). This increase in the RMSE of ρ_{am} from the 1979-2000 simulations to the 2001-2009 simulations provide the largest change reflecting the indirect influence of CFSR on the spatial structure. For downscaling with GiSTR, the spatial structure of precipitation amounts is used in conjunction with the values of μ and σ^2 . Therefore, although the spatial structure itself is not affected directly by CFSR, the spatial structure is indirectly affected by CFSR through the impact on the values of μ and σ^2 . Given that the RMSE of ρ , that is the spatial structure with no distinction between events, extreme events, or amounts, is not significantly different between simulations and the error for most spatial structure parameters is consistent between domains, there are two simulations which are recommended for potential use in forecast downscaling. Given that the error is smaller and comparable between the AW-DFIT and LWR-MOM simulations, these two simulations provide the most accurate simulation of the spatial structure and is consistent with results in Chapter 4.

5.1.3. Precipitation Amount Parameters

Given the impact of CFSR on spatial structure through values of μ and σ^2 , the RMSE of each downscaling simulation for μ and σ^2 is also considered.

5.1.3.1. Mean Daily Nonzero Precipitation (μ)

The error of each simulation for the mean daily nonzero precipitation (μ) reflects the pattern between simulations described in Section 4.1.3.1. On average, the RMSE is smallest for the AW-DFIT and largest for the LWR-MOM. The average RMSE of μ for LWR-MOM, AW-DFIT, LWR-DFIT, and AW-MOM is 6.10, 3.99, 5.95, and 4.50 mm respectively. The RMSE comparison between simulations for each month, shown in Figure 5.5, shows that the RMSE is smallest for AW-DFIT and larger for LWR-MOM and LWR-DFIT for most months than AW-MOM and AW-DFIT. The RMSE comparison also indicates that the peak RMSE for most simulations occurs in August or September. The LWR-MOM also has the largest RMSE in April and May. These errors are similar in magnitude to the previous time period, and the exact sources of error between the two domains are also similar between the two time periods. These patterns are consistent with the results shown in Chapter 4, which indicate that the AW-DFIT simulation has the lowest error and the peak RMSE is in August or September for most simulations.

Figure 5.6 shows the error of μ for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for all stations in the domain in January. For January, the error for values of μ is roughly the same across much of the domain, with some improvement in error across the domain compared to the 1979-2000 time period. However, the tendency for error to be concentrated in particular areas remains common persist in the 2001-2009 time period. Particularly for the LWR-DFIT (Figure 5.6c) the value of μ is overestimated in the northeast corner of the domain by 10 to 22 mm in January. In addition, there is little difference between LWR-MOM (Figure 5.6a), AW-DFIT (Figure 5.6b), and AW-MOM (Figure 5.6d) simulations in January. Each of these simulations shows a tendency to underestimate values

of μ in the western Carolinas and northern Georgia, which is common to the spatial pattern of error shown in Chapter 4 for these simulations..

Figure 5.7 shows the error of μ for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for all stations in the domain in June. For this month each simulation shows a similar spatial pattern to the error shown in the 1979-2000 domain with some changes in the magnitude. The LWR-MOM (Figure 5.7a) shows only slightly less or slightly more error than for the corresponding 1979-2000 LWR-MOM simulation. The AW-DFIT simulation in Chapter 4 shows a broad overestimate of μ in the northern and southern portions of the domain. However, the AW-DFIT for the 2001-2009 time period (Figure 5.7b) overestimates by a smaller magnitude, and underestimates by a larger magnitude in the central portion of the domain than in the 1979-2000 time period. The LWR-DFIT simulation in 2001-2009 (Figure 5.7c) has lower error for estimates of μ in Alabama and a smaller overestimation for the northern portion of the domain than the 1979-2000 simulation. The AW-MOM simulation in 2001-2009 (Figure 5.7d) has smaller overestimates in the southern portions of the domain, and also underestimates more in the central portion of the domain than the AW-MOM simulation for the 1979-2000 time period. The spatial pattern is similar between simulations, but the LWR-MOM and LWR-DFIT show a tendency to underestimate μ in the western Carolinas and northern Georgia. In addition, the LWR-DFIT also has a tendency to underestimate in the northeast corner of the domain. Both of these tendencies are consistent with the results shown in Chapter 4, but are between 2 and 4 mm larger, which indicates the same pattern for both of these simulations. The error associated with these regions contributes to the increased RMSE of the LWR-MOM and LWR-DFIT compared to the AW-MOM and AW-DFIT in June and on average.

Figure 5.8 shows the error of μ for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), AW-MOM (d) in September. In this month, each simulation shows similar spatial patterns to those for the error of μ in the 1979-2000 period, with some notable exceptions. The results of Chapter 4 indicate that each 1979-2000 simulation shows a tendency to underestimate values of μ in the central and coastal Carolinas. For the 2001-2009 time period, each simulation shows a tendency to underestimate values of μ throughout much of the Carolinas,

with the largest magnitudes in the western Carolinas and northern Georgia. The underestimation in this area is the larger for the LWR-MOM (Figure 5.8a) and AW-DFIT (Figure 5.8b) and smallest for the LWR-DFIT (Figure 5.8c). In addition, the LWR-MOM (Figure 5.8a) and LWR-DFIT (Figure 5.8c) show a tendency to overestimate values of μ in the northeast corner of the domain which is also shown by the results in Chapter 4. The combination of the overestimation in the northeast corner of the domain and underestimation in the western Carolinas causes the RMSE of μ for LWR-MOM and LWR-DFIT to be larger than the remaining simulations.

One pattern that is most evident in the error of μ for all four simulations is that much of the error during this time period is in the mountains of the Carolinas and Georgia. There are primarily underestimates in this region, which is also evident in other warm months, which is not consistent with the results of Chapter 4. For the earlier time period, the error of μ in the summer months was primarily an overestimation in Florida and the northeast corner of the domain. The overestimation in Florida is shown to be a consistent problem with GiSTR in Chapter 3 and shown to increase when CFSR data are used in Chapter 4. In addition, the error in the northeast corner was shown to be a source of error primarily for LWR-MOM and LWR-DFIT in Chapter 4, which use the LWR interpolation. However, while the LWR-MOM and LWR-DFIT exhibit the same tendency to overestimate μ in the northeast corner, each simulation overestimates μ with a smaller magnitude in Florida during the 2001-2009 time period than during the 1979-2000 period. While there are several spatial patterns, the AW-DFIT simulation has the smallest error for μ for this domain.

5.1.3.2. Variance of Daily Nonzero Precipitation (σ^2)

The error for the variance of daily nonzero precipitation (σ^2) between simulations follows a similar pattern to the results presented in Section 4.1.3.2. The average RMSE is smaller for the AW-MOM and LWR-MOM than the remaining simulations. The average RMSE of σ^2 for LWR-MOM, AW-DFIT, LWR-DFIT, and AW-MOM simulations is 182.89, 185.70, 187.52, and 177.19 mm^2 respectively. Figure 5.9 compares the RMSE for each

simulation across all months. This comparison indicates that there is little difference in the RMSE of each simulation for each month, and that the RMSE of each simulation peaks in September.

The sources of error of σ^2 change somewhat between the two time periods. First, the difference in the RMSE for each simulation between the two time periods in most months is less than 50 mm^2 . For example, for the month of April, results for the 1979-2000 time period show a peak in the RMSE. This same peak is not present for the 2001-2009 downscaling simulations. The difference in the RMSE of σ^2 for April between the two time periods for each simulation is approximately 1.06 mm^2 . Given that in most months each individual simulation shows little difference in the RMSE of σ^2 between the two time periods, there are a couple months which may cause the difference in the RMSE between time periods to be significant. For example, in September the average difference in the RMSE of σ^2 for each individual downscaling simulation is 255.51 mm^2 . The large difference between time periods in August and September is the likely cause of the increase in the RMSE of σ^2 between the two time periods to being significant. While the RMSE of σ^2 in 2001-2009 is largest in September for each simulation, the error pattern shown in Chapter 4 for this parameter is consistent in this time period.

Figure 5.10 shows the error of σ^2 for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for April. For this month, each 2001-2009 downscaling simulation has similar error for σ^2 across the entire domain. There is a general tendency to underestimate values of σ^2 on the northern Gulf Coast and in southern Alabama, which is consistent with the spatial patterns shown in Chapter 4. The LWR-DFIT simulation (Figure 5.10c) has the smallest number of underestimates in this area, but also has more overestimates in northern Alabama and central Tennessee. Of the three simulations remaining, the LWR-MOM (Figure 5.10a) and AW-MOM (Figure 5.10d) have a smaller number of underestimates on the northern Gulf Coast and southern Alabama than the remaining simulations. The results in Chapter 4 indicate that each simulation estimates values of σ^2 in most of southern Florida within 100 mm^2 . This is in contrast to Figure 5.10, which shows that each simulation underestimates values of σ^2 by between 100 and 500 mm^2 in southern Florida.

Figure 5.11 shows the error of σ^2 for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for September. For September, the error of σ^2 between each of the 2001-2009 downscaling simulations is again similar across the domain. Each downscaling simulation broadly underestimates values of σ^2 by up to 700 mm² across much of the domain. However, both the LWR-MOM (Figure 5.11a) and LWR-DFIT (Figure 5.11c) have smaller magnitudes for this underestimation in the Carolinas than the other simulations for September. Consistent sources of error include underestimates larger than 700 mm² on the east coast of Maryland, and overestimates on the coast of Georgia by between 100 and 300 mm².

Figure 5.12 shows the error of σ^2 for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for October. For this month, the error pattern for σ^2 between each 2001-2009 downscaling simulation is again consistent. Most of the error is less than 100mm², which is in contrast to the results shown in Chapter 4, which indicates that each simulation underestimates values σ^2 by up to 700 mm² in most of the center of the domain. Each simulation underestimates values of σ^2 1100 to 1300 mm² for several stations on the northern Gulf Coast and by 500 to 7000 mm² on the coast of North Carolina. These errors in σ^2 along the Gulf Coast and in North Carolina are consistent in summer months for both time periods. However, the error for the LWR-MOM (Figure 5.12a) and the LWR-DFIT (Figure 5.12c) is smaller for both of these regions, and the LWR-DFIT also overestimates the value of σ^2 on the Outer Banks of North Carolina by more than 300 mm².

Following the results shown in this section for σ^2 and the previous section for μ , the RMSE for both parameters between simulations is tested for significance. The two-sample T-test results for each parameter, summarized in Table 5.10, indicate that the RMSE of μ for LWR-MOM and LWR-DFIT is significantly larger than for AW-DFIT and AW-MOM, and that there is no significant difference in the RMSE of σ^2 between simulations, which is consistent with results for the same tests shown in Chapter 4. As mentioned in the previous section, there are several stations where the values of μ are overestimated by each simulation. In addition, results in the previous section indicate that the magnitude of error for each of these locations is larger for the LWR-DFIT and LWR-MOM than for AW-MOM and AW-

DFIT. This larger magnitude error causes the LWR-DFIT and LWR-MOM to have significantly larger RMSE for values of μ . Each downscaling simulation has similar error at all locations across the domain in every month, which is consistent with there being no significant difference in the RMSE of σ^2 between each of the 2001-2009 simulation. The results of the RMSE comparison between the 2001-2009 downscaling simulations for μ and σ^2 are consistent with the results presented in Chapter 4. However, comparing each of these 2001-2009 simulations to their counterparts for the period of 1979-2000 yields some interesting results. Table 5.11 summarizes the results of the two-sample T-test for the RMSE of both μ and σ^2 between time periods for each simulation. For values of μ there is no significant difference in the RMSE between time periods. However, for values of σ^2 , LWR-MOM, LWR-DFIT and AW-MOM have significantly larger RMSE in the 2001-2009 time period than in the 1979-2000 time period.

It is evident from Table 5.11 that the RMSE of μ is comparable between the two time periods for each simulation, with an average difference of less than 0.8 mm. This is consistent with the discussion in the previous section which indicated that while the sources of error changed, the overall magnitude did not change. However, the difference in the September and August RMSE of σ^2 for each simulation between time periods compared to the same difference for other months suggests that the difference in the RMSE of σ^2 between time periods is only significant for these months. In addition, the error for values of σ^2 is smaller for the LWR-MOM and AW-DFIT than the other simulations.

5.1.3.3. Average Total Precipitation ($E[S(T)]$)

The average total precipitation ($E[S(T)]$) is directly related to the mean daily nonzero precipitation (μ) and the error of each simulation for $E[S(T)]$ reflects the same pattern as the error for μ . For the 2001-2009 downscaling simulations, AW-DFIT has the smallest RMSE on average, while LWR-MOM and LWR-DFIT have larger RMSE on average across the domain than the other simulations. The average RMSE of $E[S(T)]$ for LWR-MOM, AW-DFIT, LWR-DFIT, and LWR-MOM is 79.61, 53.45, 75.74, and 62.66 mm respectively.

Figure 5.13 shows a RMSE comparison between simulations for all months. This comparison indicates that the AW-DFIT has the smallest RMSE for all months, while the LWR-MOM and LWR-DFIT have larger RMSE for all months than AW-DFIT and AW-MOM. In addition, the comparison also shows that the RMSE of each simulation has a peak in August and September. The pattern of error between each simulation for this parameter is consistent with the results presented for the values of μ . This is expected given the direct relationship between μ and $E[S(T)]$ described by Equation 2.7. The AW-DFIT simulation has the lowest error for values of $E[S(T)]$ in this “forecast” situation, which is consistent with the results for Section 4.1.3.1.

5.1.3.4. Inter-annual Variability ($Var[S(T)]$)

The inter-annual variability ($Var[S(T)]$) has a different pattern from the results of Section 1.3.4 in Chapter 4. For values of $Var[S(T)]$, AW-DFIT and AW-MOM have smaller RMSE on average across the domain than the remaining simulations. The average RMSE of $Var[S(T)]$ for LWR-MOM, AW-DFIT, LWR-DFIT, and AW-MOM is 3231.06, 2481.17, 3144.14, and 2515.36 mm² respectively. A comparison of the RMSE of each simulation by month, shown in Figure 5.14, indicates that the AW-MOM and AW-DFIT simulations have the smaller RMSE for all months than the remaining simulations. In addition, comparison also shows that each simulation has its peak RMSE in August and September. These results are consistent with the results shown in Chapter 4. Another interesting result is that the LWR-MOM simulation much larger RMSE than the other simulations in April and May. Recall that $Var[S(T)]$ is directly related to μ and σ^2 (Equation 2.8, Chapter 2). In addition, each simulation has a peak RMSE for μ and σ^2 in August and September and the LWR-MOM simulation has a second peak RMSE for μ in April and May, which did not exist in the same comparison in Chapter 4. The relationship between μ and σ^2 and $Var[S(T)]$ provide the link to the RMSE of $Var[S(T)]$ for each simulation. This relationship suggests that the same sources of error which impact both μ and σ^2 impacts $Var[S(T)]$. That is, the error for individual stations or groups of stations for μ and the broad underestimation of σ^2 , which is

largest during the peak of the Atlantic hurricane season, is the probable cause for the seasonal patterns in the error of $Var[S(T)]$.

Following the analysis of the Section 5.1.3.2, the RMSE of $E[S(T)]$ and $Var[S(T)]$ between each simulation was tested for significance. Table 5.12 summarizes the results of the two-sample T-tests for the RMSE between simulations for $E[S(T)]$ and $Var[S(T)]$. The T-tests for $E[S(T)]$ indicate that LWR-MOM and LWR-DFIT have significantly larger RMSE than AW-DFIT and AW-MOM. In addition, the tests for $Var[S(T)]$ also shows that there is no significant difference between simulations. These patterns are consistent with patterns shown for μ and σ^2 , and also consistent with results shown in Chapter 4, which indicate that the LWR-MOM and LWR-DFIT simulations have significantly larger RMSE than AW-DFIT and AW-MOM for $E[S(T)]$ and no significant difference between simulations for the RMSE of $Var[S(T)]$.

Given that the $E[S(T)]$ and $Var[S(T)]$ are the average total precipitation and the inter-annual variability, it is important to understand the potential error associated with forecasting these parameters. Therefore, the two-sample T-test is also used to compare the RMSE of each 2001-2009 downscaling simulation to the RMSE of their counterparts in Chapter 4 for both $E[S(T)]$ and $Var[S(T)]$. The results of these tests for each parameter between time period are summarized in Table 5.13. The testing shows that there is no significant difference in the RMSE of $E[S(T)]$ between the two time periods for each simulation, and that the 2001-2009 downscaling simulations have a significantly larger RMSE for $Var[S(T)]$ than their 1979-2000 counterparts. Therefore, while the results of this analysis indicate the potential errors associated with using these downscaling variations in a forecast, there are some clear results for the relationship of each simulation between time periods. For values of μ and $E[S(T)]$ the error associated with each simulation would be no different in a forecasting situation compared to the calibration period. However, for values of σ^2 and $Var[S(T)]$, the error of each simulation can be larger in during a forecast period compared to a calibration period. Finally, while there are many seasonal patterns for each simulation, the AW-DFIT and AW-MOM simulations show smaller error for values of $Var[S(T)]$ than the remaining simulations.

5.1.4. Probability Distribution Function Analysis

In section 5.1.3.1, the error for each 2001-2009 simulation for values of μ is shown to have a similar pattern to the 1979-2000 simulations across the domain. This is also reflected in the PDFs of nonzero precipitation generated for each station by each downscaling simulation. As an example, consider the month of January for several stations in the domain. From Section 5.1.3.1 the difference in RMSE of μ between simulations is less than 2.2 mm, and is less than 30 mm^2 for σ^2 . This small difference is also indicated by the PDFs for most locations in the domain. Figure 5.13 shows a comparison between the observed and generated PDFs for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for January for station 090969 in Blairsville, GA. Each simulation accurately reproduces much of the observed PDF for this station in the 2001-2009 time period. In addition, although each downscaling simulation accurately reproduces the PDF for this station, LWR-MOM (Figure 5.15a), AW-DFIT (Figure 5.15b), and AW-MOM (Figure 5.15d) overestimate the frequency of small precipitation amounts, but do not overestimate the frequency of precipitation at the capping value. The overestimation of the frequency of small amounts compared to other precipitation amounts causes each simulation to underestimate the value of μ for this station by 6 mm. In contrast, LWR-DFIT (Figure 5.15c) does overestimate the frequency of precipitation at the capping value for this station. As such, where this simulation would normally underestimate μ , the overestimate of the frequency of precipitation amounts at the capping value counters the overestimate of the frequency of small precipitation amounts. This causes the LWR-DFIT for this station to underestimate μ by less than 4 mm.

Figure 5.16 shows a comparison between the observed and generated PDFs for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for January for station 319457 in Wilmington, NC. Similarly to the results shown in Figure 5.15, each simulation reproduces much of the observed PDF. In contrast to the PDFs for station 090969, there are no instances where any of the simulations overestimate the frequency of precipitation at the

capping value for station 319457. However, each simulation has a tendency to overestimate the frequency of amounts between 10 and 20 mm.

Finally, Figure 5.17 shows a comparison between the observed and generated PDFs for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for January for station 286026 in Newark, NJ. This station is in the northeast corner of the domain, and the error of μ for the LWR-MOM, AW-DFIT, and AW-MOM simulations is less than 4 mm. However, the LWR-DFIT simulation overestimates the value of μ at this station by more than 7 mm. For this station it is apparent that the LWR-MOM (Figure 5.17a), the AW-DFIT (Figure 5.17b), and the AW-MOM (Figure 5.17d) capture almost all of the observed PDF. However, the LWR-DFIT (Figure 5.17c) does not capture the frequency of small precipitation amounts and overestimates the frequency of precipitation at the capping value. This is also shown to occur for stations in the same area in 1979-2000 domain for LWR-DFIT in Chapter 4, and causes the overestimate of μ at this station and others in the northeast corner of the domain. The error for values of μ shows similar spatial patterns in the domain between time periods, but has a smaller magnitude than in summer months, which is reflected by the PDFs in the domain. However, the values of σ^2 are consistent between simulations for most stations, which are reflected by the common tendency to capture much of the PDF between simulations for most stations for January and all winter months.

However, as described earlier the peak RMSE of μ for most simulations is in August and September and the error of μ is spread with no apparent pattern throughout the domain. In addition, the peak error for all simulations for σ^2 is also in August and September. For the three example stations shown in January, consider the PDFs produced for September at each station. Figure 5.18 shows a comparison between the observed and generated PDF for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) in September for station 090969 in Blairsville, GA. Each simulation produces PDFs that are similar at small precipitation amounts to the observed PDF. For this station, the LWR-MOM, AW-DFIT, and AW-MOM simulations underestimate the value of μ by less than 2 mm, and the LWR-DFIT underestimates by between 2 and 4 mm. Each station also underestimates the value of σ^2 by between 100 and 300 mm². While each of these 2001-2009 simulations reproduce the

PDF for precipitation amounts less than 25 mm, each of these simulations has a tendency to overestimate the frequency of precipitation amounts between 25 and 40 mm as well at the capping value for this station. Both the LWR-MOM (Figure 5.18a) and LWR-DFIT (Figure 5.18c) hit the capping value with equal frequency, while the AW-MOM (Figure 5.18d) hits the capping value more frequently than the other simulations. Given that each simulation captures the frequency of small amounts at this station and hits the capping value associated with this station, it is likely the underestimation of μ and σ^2 is the result of underestimating the frequency of heavy precipitation. The LWR-MOM, AW-DFIT, and AW-MOM simulations have a larger overestimation of the frequency of amounts between 25 and 40 mm than the LWR-DFIT simulation, which may contribute to better error for values of μ for the LWR-MOM, AW-DFIT, and AW-MOM simulations, by offsetting the lack of precipitation larger than the capping value.

Figure 5.19 shows a comparison between the observed and generated PDFs for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for September for station 319457 in Wilmington, NC. For this station, each simulation also appears to reproduce the observed PDF of nonzero precipitation in September. This is consistent with the results for the error of μ which shows that each simulation has values within 2 mm of the observed value of this station. In contrast to the other simulations, the PDF produced by AW-MOM (Figure 5.19d) for this station includes an overestimation of the frequency of precipitation at the capping value. The overestimation of the frequency of precipitation at the capping value by AW-MOM contributes the difference in the error of σ^2 between AW-MOM and the other simulations. The AW-MOM simulation underestimates the value of σ^2 by between 100 and 300 mm^2 at this station, while the remaining simulations underestimate the value of σ^2 by 300 to 500 mm^2 . Though the PDF produced by AW-MOM for this station shows some slight differences for these amounts, the PDF for the remaining precipitation amounts is reproduced accurately, which offsets the error from the overestimation of the capping value.

Finally, Figure 5.20 shows a comparison between the observed PDF and those generated by LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for station 286026 in Newark, NJ for September. At this station, the LWR-MOM and LWR-DFIT

simulations overestimate the value of μ by more than 7 mm, and the AW-DFIT and AW-MOM simulations estimate the value of μ within 2 mm. The cause of the error of μ for the LWR-MOM (Figure 5.20a) and LWR-DFIT (Figure 5.20c) is apparent in the PDF produced by each simulation. Each of the LWR simulations underestimates the frequency of the smallest precipitation amounts, and overestimates the frequency of precipitation at the capping value. This combination of underestimation and overestimation causes the LWR-MOM and LWR-DFIT simulations to overestimate the value of μ at this station in September. This tendency is common for all stations in the northeast corner of the domain for the summer months where both LWR-MOM and LWR-DFIT overestimate the value of μ by more than 7 mm. In addition, each simulation underestimates the value of σ^2 by up to 300 mm², suggesting that although the PDFs produced by each simulation are indicative of the overestimation of μ , the errors in the PDF of each simulation allows each simulation to have similar error for σ^2 . While there are slight differences in the PDFs produced between time periods, the general patterns are consistent between time periods.

Recall that in the earlier analysis it was shown that the RMSE of σ^2 is significantly larger in the 2001-2009 time period, not significantly different between simulations, but follows a similar spatial pattern to the error of σ^2 in the 1979-2000 time period. For most months, the error of σ^2 is less than 300 mm² in magnitude across the domain for each simulation. However, as shown previously there are several locations (particularly in September) where the error of σ^2 is underestimated by more than 700 mm². As an example, Figure 5.21 shows a comparison of the observed PDF with those generated by LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for station 092283 in Cornelia, GA for September. Each 2001-2009 simulation underestimates the value of σ^2 for September by 700 mm². The PDF generated by each downscaling simulation shows how each simulation hits the capping value inherent to GiSTR and overestimates the frequency of the smallest precipitation amounts. The result is that each simulation does not capture the PDF for the largest precipitation amounts. These larger precipitation amounts are primary contributors to values of σ^2 . Although they occur rarely, these events increase the observed values of σ^2 and, when they are not captured by the simulation, cause an underestimation of the value of σ^2 at

this station. This pattern was common for all stations where σ^2 was underestimated by more than 700 mm² in the domain. While these extreme values of events may be captured if the capping value was increased or removed, as also shown in Chapter 3, there is another possible source for this error. While this study has assumed the COOP stations as the truth, there may be some data points for some stations which passed quality control but are not good quality data, which would have an impact on the results particularly with regards to evaluating the generated values of σ^2 . Therefore, the data quality during the 1971-2000 period may have an impact on the variability of precipitation amounts during the forecast period.

5.1.5. Dry and Wet Spell Frequency Analysis

Following the analysis shown in Chapter 4, the frequencies of dry and wet spells of various lengths are also evaluated for each 2001-2009 downscaling simulation in the domain. Figure 5.22 shows the observed frequency of wet spells of various lengths in January (a) compared to those generated by LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d), and AW-MOM (e) for all stations in the domain. From this comparison it is clear that the wet spell frequencies of each simulation for various lengths are similar to each other. However, each 2001-2009 simulation underestimates the frequency of one day and two day wet spells in the domain. In addition, each simulation produces wet spell frequencies three days and longer accurately. Figure 5.23 shows the observed frequency of wet spells of various lengths in July (a) compared to those generated by LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d), and AW-MOM (e) for all stations in the domain. In contrast to the January wet spells, each 2001-2009 downscaling simulation reproduces the frequency of wet spells of various lengths in July with very few discrepancies. However, the wet spell frequencies produced by AW-DFIT (Figure 5.23c) and LWR-DFIT (Figure 5.23d) are slightly more accurate than the frequencies produced by the other simulations.

Figure 5.24 shows the observed frequency of dry spells of various lengths for January (a) compared to those generated by LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d), and

AW-MOM (e) for all stations in the domain. Similarly to the January wet spells, each 2001-2009 simulation produces frequencies similar to the other downscaling simulations. In addition, each 2001-2009 downscaling simulation has a tendency to overestimate the frequency of short dry spells (less than 3 days), which is consistent with the underestimation of short wet spells for the same month. In addition each simulation also replicates the frequency of dry spells three days and longer accurately. Finally, Figure 5.25 shows the observed frequency of dry spells of various lengths in July (a) compared to those generated by LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d), and AW-MOM (e) for all stations in the domain. Similarly to the July wet spells, there is also very little discrepancy between the observed and generated dry spell frequencies for each simulation in July. However, AW-DFIT (Figure 5.25c) and LWR-DFIT (Figure 5.25d) are slightly more accurate for reproducing the frequency of July dry spells. While AW-DFIT and LWR-DFIT appear to be slightly more accurate for both dry and wet spells in July, the difference between each downscaling simulation is minimal.

These results are consistent with the earlier T-tests on P_{0I} and P_{1I} , which indicated no significant difference in the RMSE of each parameter between downscaling simulations. In addition, the fact that each downscaling simulation improves the representation of the dry and wet spell frequencies in summer months agrees with the trend in the RMSE of P_{0I} and P_{1I} (Tables 5.1 and 5.2), which indicated that the RMSE of each parameter was smallest in summer months. The majority of the results described in this section are consistent with the results presented in Chapter 4 for the comparison between simulations, and to the observed frequencies. However, the analysis in this section also shows some distinct increases in error for a potential forecasting situation. While this section does show some potential errors in forecasting with each downscaling variation, CFSR uses data assimilation and therefore does not reflect exact conditions in most forecasting models. Therefore, it is recommended as future work that this analysis is repeated with the CFS Reforecast, to assess the full potential associated with downscaling for seasonal forecasting by these methods. Given the comparable patterns of error in this time period and that it has lower or comparable error for most parameters in this time period, the LWR-MOM variation replicates most parameters

with the lowest error in a “forecasting” situation. However, while there is comparable error between simulations, an additional conclusion is that for most parameters the AW-MOM simulation did not have the lowest error, indicating that any of the other variations are more accurate than the traditional approach to downscaling with weather generators, which is consistent with the conclusions reached in Chapter 4. Following the analysis shown in Chapter 4, the next section focuses on the differences with regards to downscaling with GiSTR and WGEN under a fixed set of assumptions.

5.2. Weather Generator Comparison – Downscaling Context

Following the analysis presented in Chapter 4, this section will focus on the evaluation of WGEN and GiSTR in a downscaling context. As mentioned previously WGEN has been chosen for this analysis because it is a standard for building Markov chain generators, and GiSTR has been chosen since it replicates the spatial structure and better replicates the variability associated with extreme precipitation, which is a historical problem for most weather generators. The 2001-2009 time period and the associated stations for cross validation described at the beginning of this chapter are also used in this analysis. Following the analysis in the previous chapter, the LWR-MOM is used as the downscaling assumptions for this analysis. The GiSTR downscaling simulation used is the same as LWR-MOM simulation in the previous section. While the previous chapter indicates that the LWR-MOM simulation has the lowest error for most parameters, Section 5.1 of this chapter indicates that the LWR-MOM simulation has the lowest error for several parameters and comparable error for the remaining parameters. Given these collective results, the LWR-MOM is again used to keep consistency between this section and Section 4.2.

5.2.1. Temporal Parameters

In Section 4.2.1 the error of the temporal parameters is shown to be larger for WGEN than for GiSTR. This section presents a similar analysis and also indicates that in the error is larger for the WGEN simulation than the GiSTR simulation, for the 1st order Markov Transition probabilities (P_{01} and P_{11}), unconditional probability of rain (π), and persistence (γ). For values of P_{01} , the difference in the RMSE between the two simulations is less than 0.02 on average, with WGEN having less error than GiSTR. For values of P_{11} , the RMSE of the WGEN simulation is larger than GiSTR by 0.06 to 0.20. A comparison of the RMSE between simulations for both P_{01} and P_{11} across all months is shown in Table 5.14. For both parameters, there is no seasonal trend in the RMSE for each simulation. However, there is little difference in RMSE for P_{01} for all months, while the difference in RMSE for P_{11} is more than 0.06 for all months. The pattern and magnitude of the RMSE for both generators is consistent with those shown in Chapter 4, which indicates that WGEN has a larger RMSE than GiSTR for these parameters in a downscaling context. In addition, the error of each simulation for values of π and γ is similar to the results of Chapter 4. For each parameter, the RMSE of WGEN is larger on average than the GiSTR simulation by between 0.08 and 0.1. Table 5.15 shows a comparison of the RMSE of each simulation for all months for both parameters. For values of π , the WGEN downscaling simulation has a larger RMSE for all months by 0.03 to 0.15. Similarly, for values of γ , the RMSE of the WGEN downscaling simulation is larger than the RMSE for the GiSTR downscaling simulation by up to 0.17. In addition, neither simulation shows a seasonal trend in the RMSE for either parameter. The two-sample T-test results, summarized in Table 5.16, also show similar results to those presented in Chapter 4. That is, the test results in both time periods indicate that the RMSE of P_{11} , π , and γ for the WGEN downscaling simulation is significantly larger than the RMSE for the GiSTR downscaling simulation. This suggests that the error of GiSTR for these temporal parameters is less than the WGEN simulation. This in turn suggests that GiSTR would have much lower error than WGEN in a “forecast” situation for the occurrence of precipitation in this domain.

5.2.2. Spatial Structure Parameters

As mentioned in Chapter 4, the spatial structure is not directly affected by the CFSR data. However, the spatial structure when GiSTR is used for downscaling is indirectly affected by the impact of the CFSR data on the temporal parameters (P_{01} , P_{11} , π , and γ). For the correlation matrices representing the spatial structure, the WGEN simulation has a consistently larger RMSE than GiSTR by up to 0.21 on average. Figure 5.26 shows a RMSE comparison between simulations for the correlation matrix of precipitation (a, ρ), the correlation matrix of precipitation events (b, ρ_{ev}), the correlation matrix of precipitation amounts (c, ρ_{am}), and the correlation matrix of precipitation extreme events (d, ρ_{ex}) for all months. This comparison indicates that the GiSTR simulation has the smallest RMSE for all months for all four parameters. The results are similar for the average decay function correlograms produced by each simulation. Figure 5.27 shows the observed average decay function correlogram compared to those generated by the WGEN and GiSTR simulations for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d) for January. For all four parameters the GiSTR simulation better replicates the observed correlogram, which is consistent to the RMSE of each correlation matrix for each simulation. Similarly to the results in Section 4.2.2 the GiSTR simulation has the best match to the observed correlogram for precipitation events (Figure 5.27b). The correlograms produced by each simulation in July also have similar results to the RMSE comparison. Figure 5.28 shows the observed average decay function correlogram compared to those generated by the WGEN and GiSTR simulations for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d) for July. As with the results for January, the GiSTR simulation consistently provides a better estimate of the observed correlogram than the WGEN simulation for all four parameters. Similarly to the results from Section 5.1.2 the GiSTR simulation has the best estimate of the observed correlogram for precipitation amounts (Figure 5.28c) in July. The results are consistent with the RMSE comparison in Figure 5.26 and with the results discussed in Section 4.2.2, which also suggests that the two-sample T-tests will also provide similar results.

The two-sample T-test results comparing the RMSE of each simulation for all four parameters are summarized in Table 5.17. The results of the tests between simulations show that the RMSE for the WGEN simulation is significantly larger for each parameter than GiSTR. In all cases, these results reflect the results described in Section 4.2.2 and the comparison between generators in Section 3.2.2, which both indicated that RMSE of the spatial structure parameters for WGEN is significantly larger than GiSTR. However, since the spatial structure parameters are fixed to the observed correlation matrices for 1971-2000, the impact on the spatial structure resulting from the indirect effect of the downscaling approach can be determined by comparing the RMSE of these 2001-2009 simulations with those simulations not used for downscaling. For this brief analysis, the downscaling simulations in this section will be compared to their counterparts from Chapter 3 (the WGEN Control and the GiSTR Control respectively). The RMSE of each correlation matrix for the WGEN and GiSTR simulations are tested using the two-sample T-test against the RMSE of each matrix for the WGEN and GiSTR Control simulations, with the result summarized in Table 5.18. This test shows that the GiSTR downscaling simulation has significantly larger RMSE for ρ_{ev} and ρ_{am} than the GiSTR Control simulation. The test also shows that the WGEN downscaling simulation has significantly larger RMSE than the WGEN Control for ρ_{am} only. These results are not consistent with the analysis shown in Chapter 4. In Chapter 4, only GiSTR has a significantly larger RMSE for ρ_{am} only. While the difference in the RMSE are statistically significant, the difference between the 2001-2009 downscaling simulations and their respectively controls is less than 0.05 in these situations. The test indicates that the impact on spatial structure is strongest when GiSTR is used for downscaling. As described in Chapter 2, both GiST and GiSTR generate precipitation events and amounts using both the spatial and temporal information, and for downscaling the spatial structure was held fixed since the resolution of CFSR cannot resolve changes in the individual correlations between stations. However, since GiSTR uses both the temporal and spatial structure in the generation process the spatial structure output by GiSTR is indirectly affected by the impact of CFSR on the temporal parameters. The result of this is significantly increased RMSE for the spatial structure for precipitation events and amounts.

Although there is a clear impact on the ρ_{ev} and ρ_{am} , the RMSE of ρ is not significantly different between the GiSTR downscaling simulation and the GiSTR Control. This suggests that while there is a statistically significant impact on some aspects of spatial structure, the impact on the overall spatial structure produced when downscaling with GiSTR is negligible. In addition, since WGEN does not take into account the spatial structure of precipitation events or amounts and there is only one instance of significantly larger RMSE for the WGEN downscaling simulation compared to the WGEN Control (ρ_{am}). The impact of the CFSR data on the spatial structure resulting from the WGEN downscaling simulation is also negligible. These results still lead to the conclusion that the GiSTR produces a more accurate spatial structure for precipitation than the WGEN when used in a downscaling context, which is consistent with the results presented in Chapters 3 and 4.

5.2.3. Precipitation Amount Parameters

While this analysis has assumed that the differences between weather generators used in a downscaling context are similar to when they are not used for downscaling, there are several instances where this assumption is not valid.

5.2.3.1. Mean Daily Nonzero Precipitation (μ)

In Section 4.2.3.1, the WGEN simulation was shown to have larger error than the GiSTR simulation for values of the mean daily nonzero precipitation (μ). The comparison of the WGEN and GiSTR simulation in this “forecast” simulation also shows that WGEN has a larger error than GiSTR for values of μ . The average RMSE for the WGEN and GiSTR simulations is 7.55 and 6.10 mm respectively. The RMSE comparison across all months, shown in Figure 5.29, indicates that the RMSE of the WGEN simulation has similar error to the GiSTR simulation for most months in the domain. However, the comparison also shows that the RMSE of WGEN in April is 28 mm, or 235% larger than the GiSTR simulation and

279% larger than the average RMSE for the WGEN simulation, which has also been shown by the same comparison in section 4.2.3.1. For values of μ in April, there are several stations where the values are overestimated in both simulations. Figure 5.30 shows the error of μ for the WGEN simulation (a) and the GiSTR simulation (b) in April. For the WGEN simulation (Figure 5.30a) there are several stations on the coast of Georgia where the value of μ is overestimated by 180 to 300 mm. For the same locations, the GiSTR simulation (Figure 5.30b) overestimates μ by between 20 and 40 mm. These results are consistent with those shown in Chapter 4 for both magnitude and location, and indicate that WGEN has a higher error than GiSTR in a downscaling context. The errors at these stations have the largest impact on the RMSE of μ in April for the WGEN simulation. This information also suggests that removing the outlier stations will improve the error associated with the WGEN simulation in a forecast situation, throughout the domain. However, as the analysis of the PDFs for each simulation suggests, the LWR-MOM variation may have produced some erroneous results for the input parameters required for both GiSTR and WGEN. Therefore, a different simulation may provide better estimates of this parameter at these outlier stations.

5.2.3.2. Variance of Daily Nonzero Precipitation (σ^2)

In Section 4.2.3.2, the WGEN simulation was shown to have larger error than the GiSTR simulation for the variance of daily nonzero precipitation (σ^2). In this “forecast” situation, the error of the WGEN simulation is also larger than the GiSTR simulation. The average RMSE of σ^2 for the WGEN and GiSTR downscaling simulations is 342.91 and 182.89 mm² respectively. Comparing the RMSE of each simulation for all months, shown in Figure 5.31, indicates that the RMSE is similar between simulations for all months, except for the month of September. In September, the RMSE for the WGEN simulation is more than 1500 mm². This RMSE for the WGEN simulation in April is 272% larger than the RMSE of the GiSTR simulation in April, and 341% larger than the average RMSE for the WGEN simulation. These results are consistent to the results presented in Chapter 4, though inconsistent with the results from Chapter 3, which indicates that GiSTR is more accurate for

this parameter in a downscaling context than WGEN. Similarly to the analysis for μ in April, there are stations where the error of the WGEN simulation in September is larger than surrounding station for values of σ^2 . Figure 5.32 show the difference between generated and observed values of σ^2 for the WGEN (a) and GiSTR (b) simulations. For values of σ^2 in September, there are several stations in the southern Appalachian Mountains where the WGEN simulation (Figure 5.32a) overestimates by more than 3500 mm². In contrast, for the same stations the GiSTR simulation (Figure 5.32b) estimates values of σ^2 within 500 mm² of the observed value. Therefore, similarly to the previous section and Section 4.2.3.2, the error associated with individual outlier stations has an impact on the RMSE of the WGEN simulation and removing these outlying stations will likely improve the error of the WGEN simulation with respect to values of σ^2 . However, altering the variation used may also improve the results produced by both WGEN and GiSTR for these stations.

5.2.3.3. Average Total Precipitation ($E[S(T)]$)

The error for the average total precipitation ($E[S(T)]$) for each simulation in this time period follows a similar pattern to the results from Section 4.2.3.3 and to the results from Section 5.2.3.1. The error of $E[S(T)]$ is largest for the WGEN simulation across the domain. The average RMSE of $E[S(T)]$ for the WGEN and GiSTR downscaling simulation is 86.09 and 79.61 mm respectively. A comparison of the RMSE between simulations for all months, shown in Figure 5.33, indicates that the RMSE for GiSTR is less than or similar to the WGEN simulation. However, for the month of April, the RMSE of the WGEN simulation is 254.54 mm, which is 142% larger than the RMSE of the GiSTR simulation for the same month. In addition, the RMSE of the WGEN simulation for April is also 196% larger than the average RMSE of the WGEN simulation for all months. These results are consistent with those shown in Chapter 4, but are larger in magnitude particularly for April (222.72 mm in 1979-2000, 254.51 mm in 2001-2009). Given the direct relationship between $E[S(T)]$ and μ , described by Equation 2.7, it is likely that the increased RMSE for the WGEN simulation in April for $E[S(T)]$ is also related to the high error of μ for individual stations shown in Section

5.2.3.1. Therefore, it is suggested that the error of the WGEN simulation for $E[S(T)]$ would improve if the outlier stations for values of μ are removed.

5.2.3.4. Inter-annual Variability ($Var[S(T)]$)

The inter-annual variability ($Var[S(T)]$), which is impacted by both the mean daily nonzero precipitation (μ) and the variance of the daily nonzero precipitation (σ^2), is shown to be better replicated by the GiSTR simulation in Section 4.2.3.4. In this “forecast” period, GiSTR again has the smallest error for $Var[S(T)]$. The average RMSE for the WGEN and GiSTR simulations are 16743.59 and 3231.09 mm² respectively. Figure 5.34 shows a comparison of the RMSE between simulations for all months. This comparison shows that the RMSE of each simulation is similar for most months except April and September. In April, the RMSE of the WGEN simulation is 126442.7 mm², which is over 3200% larger than the April RMSE for the GiSTR simulation and 655% larger than the average RMSE for the WGEN simulation. In September, the RMSE of the WGEN simulation is 26965.59 mm². The RMSE for September for this WGEN simulation is 430% larger than the September RMSE for the GiSTR simulation and 61% larger than the average RMSE of the WGEN simulation. The consistent pattern between each of these parameters reflects the direct relationship between μ , σ^2 , and $Var[S(T)]$ described in Chapter 2 (Equation 2.8). The increased April RMSE for the WGEN simulation reflects the increased error associated with the individual stations in April for values of μ . The increased September RMSE for the WGEN simulation reflects the increased error associated with individual stations for values of σ^2 . All of these results are consistent with those shown in Chapter 4. However, the magnitude of the error of WGEN increases by almost 6000 mm² on average from the 1979-2000 time period to the 2001-2009 time period.

While these errors are much larger than the error for the remaining stations, this does not indicate that the RMSE of μ , σ^2 , $E[S(T)]$, and $Var[S(T)]$ for the WGEN simulation is significantly larger than the RMSE for the GiSTR simulation. The two-sample T-test results, summarized in Table 5.19, indicate that there is no significant difference in the RMSE of any

of parameters between the two downscaling simulations. In addition, the error for values of μ and σ^2 is consistent across the domain for the remaining months and less than 10 mm and 500 mm² respectively, such as in the example shown in the error of μ and σ^2 for January in Figure 5.35. Therefore, removing the individual stations which are outliers in error may cause the RMSE of these parameters to reflect the results shown in Chapter 3, by reducing the error associated with April and September in the WGEN simulation.

5.2.4. Probability Distribution Function Analysis

The previous section shows that the increased error associated with each of the amount parameters for the WGEN simulation is related to increased error with individual stations in the domain. This raises a question of how well the WGEN simulation captures the PDF for these stations in April and September for this “forecast” period. While the WGEN simulation captures the PDF for most stations, when values of μ are overestimated by over 180 mm the WGEN downscaling fails to capture the PDF. As an example, consider a comparison between the observed PDF and the PDF produced by WGEN for station 091345 in Brunswick, GA in April shown in Figure 5.36. For this station the value of μ is overestimated by more than 200 mm in April and the PDF produced by this WGEN downscaling simulation fails to capture the observed PDF for this station. When values of σ^2 are overestimated by over 3500 mm² in September, the differences between the observed and generated PDFs are equally apparent. For example consider a comparison between the observed PDF and the PDF produced by WGEN for station 085973 near Lake Wales, FL in September shown in Figure 5.37. For this month and station, the WGEN simulation overestimates the value of σ^2 by 16300 mm². The WGEN simulation has a tendency to overestimate the frequency of the smallest precipitation amounts and overestimate the frequency of precipitation amounts larger than 200 mm. This combination of overestimates causes the WGEN simulation to overestimate the value of σ^2 for this station.

Given that the errors for these same stations for WGEN is smaller when the generator is not used for downscaling compared to when the generator is used for downscaling with the

LWR-MOM variation it is possible that LWR-MOM variation produced erroneous input parameters. The two parameter gamma distribution is used by GiSTR and WGEN to represent precipitation amounts. The LWR-MOM uses the locally weighted regression and the moment estimators of the gamma distribution to determine the two parameters of the gamma distribution used by each generator. Therefore, it is possible that either the locally weighted regression or the moment estimation produced values of the shape and scale of the gamma distribution. The GiSTR simulation also has smaller error than the WGEN simulation for the same stations where WGEN overestimates μ by more than 180 mm or σ^2 by more than 3500 mm². Given that GiSTR has a capping value to prevent unrealistic precipitation amounts from occurring, and that the same input parameters are used in both the GiSTR and WGEN simulations, it is likely that the capping value present in GiSTR counters the error associated with erroneous input parameters produced by the LWR-MOM at these stations. As a result, the GiSTR simulation has lower error than the WGEN simulation for these stations though the input parameters are erroneous.

The trends in the PDFs where values of μ are overestimated by over 180 mm and σ^2 are overestimated by over 3500 mm² by WGEN are similar between stations where this occurs, and is also consistent with the PDF comparison in Section 4.2.5. The overestimation of the occurrence of small amounts combined with the underestimation of the occurrence of larger precipitation amounts could be the cause of the overestimation of the σ^2 for this station and month. Therefore, removing stations where the WGEN simulation overestimates μ by over 180 mm and overestimates σ^2 by over 3500 mm² potentially improving error associated with forecasting with this WGEN simulation. In addition, one important note is that those stations which overestimate μ by over 180 mm and σ^2 by over 3500 mm² in this time period are the same as those outlier stations in the 1979-2000 time period.

5.2.5. Changes with Outliers Removed

In order to determine the impact associated with removing the stations with outlier error in the WGEN simulation, similar analyses to Section 5.2.3 are performed with these

stations removed. Every station was removed where the error of the mean daily nonzero precipitation (μ) is more than 180 mm in the WGEN simulation, and where the error of the variance of the daily nonzero precipitation (σ^2) is more than 3500 mm² from both simulations. These seven outlier stations are removed in both simulations since while these stations were outliers in the WGEN simulation, they were also outliers in the GiSTR simulation but with smaller magnitude than the WGEN simulation. For the temporal and spatial structure parameters, there was no difference in the results with the outlier stations removed. However, the error for values of μ improves for both simulations with these outliers removed. The average RMSE of the WGEN and GiSTR simulations is 6.25 and 5.80 mm respectively. In addition, the RMSE comparison by month, shown in Figure 5.38, indicates that the RMSE of each simulation is again similar for most months. However, the WGEN simulation has an RMSE of 14.91 mm for April, which is 84% larger than the GiSTR simulation for April and 139% larger than the average RMSE for the WGEN simulation. In addition, the RMSE of WGEN simulation improves by approximately 50%.

The results for values of $E[S(T)]$ are similar to the results for values of μ . The WGEN simulation has larger RMSE on average than the GiSTR simulation for values of $E[S(T)]$, and the RMSE of each simulation is smaller with the outlier stations removed. The average RMSE for WGEN and GiSTR is 76.96 and 75.51 mm respectively. For the RMSE comparison across all months, shown in Figure 5.39, it is apparent that each simulation has similar RMSE for most months. However, the RMSE of the WGEN simulation in April is 183.75 mm which is 76% larger than GiSTR in April and 139% larger than the average RMSE for WGEN. In addition, the error of the WGEN simulation improves by 28%. This combination suggests that while removing outliers improves the error of μ and $E[S(T)]$ for both simulations, the WGEN simulation still has the largest error for both parameters in a “forecast” situation. This result is in contrast to what was shown in Section 4.2.5, where removing the outlier stations caused the WGEN simulation to have less error for values of μ .

The variance of daily nonzero precipitation (σ^2) for each simulation shows similar error with the outliers removed to the results shown in Section 5.2.3.2. While the RMSE for each simulation improves, the WGEN simulation has the largest RMSE on average. The

average RMSE for WGEN and GiSTR is 249.43 and 178.67 mm² respectively. The RMSE comparison for all months, shown in Figure 5.40, indicates several results which are different from the earlier results shown in Section 5.2.3.2. For all months, the GiSTR simulation has smaller RMSE than WGEN, which is consistent with the earlier results. However, while each simulation has a peak RMSE in September, the difference between simulations for this month is smaller compared to results with outlier stations included. The RMSE of the WGEN simulation in September is 459.10 mm², which is 14% larger than GiSTR in September and 84% larger than the average RMSE for WGEN. The RMSE of the WGEN simulation improves by 70%. These results indicate an improvement in the error of values of σ^2 for both simulations, though the GiSTR simulation has the lowest error which is consistent with previous results in Chapter 3 and Chapter 4.

Finally, for the inter-annual variability ($Var[S(T)]$), there is an improvement in the error of this parameter for each simulation with the outlier stations removed. However, although the error of both simulations is improved, the RMSE of the GiSTR simulation is smaller than the RMSE for WGEN. On average, the RMSE of the WGEN and GiSTR simulation with outliers removed is 5908.23 and 3017.58 mm² respectively. The RMSE comparison between simulations by month, shown in Figure 5.41, shows that the GiSTR simulation has smaller RMSE than WGEN for most months. In addition, the comparison also shows that the RMSE of WGEN is 18730.42 mm² in April. This value is 412% larger than the corresponding value for GiSTR and 217% larger than the average RMSE for WGEN. This is an improvement for the WGEN simulation by 85%. While removing the outlier stations improved the RMSE of WGEN for April compared to GiSTR, it is apparent that the error of this month is still associated with the error for values of μ . In contrast, the RMSE of WGEN in September with outliers removed is 5971.96 mm², which is only 23% larger than the corresponding value for GiSTR. This is an improvement for the WGEN simulation by 79%. It has been mentioned that the error of σ^2 for the WGEN simulation also improved for September with the outlier stations removed. The improved error of σ^2 translated to an improved error for $Var[S(T)]$, causing the RMSE of WGEN for $Var[S(T)]$ to be similar to GiSTR for the month of September. Although the error of both simulations

improves with the outlier stations removed, the WGEN simulation still has larger error than the GiSTR simulation, which is consistent with the results of Sections 4.2.3.4 and 4.2.5 and Section 5.2.3.4.

While the results of the error comparison for each parameter are similar the test for significance show an interesting result compared to the earlier tests for each of the amount parameters. Table 5.20 show the results of the two-sample T-test for the RMSE of μ , σ^2 , $E[S(T)]$, and $Var[S(T)]$ between simulations with the outlier stations removed. Similarly to the results with outlier stations included, there is no significant difference in the RMSE of μ and $E[S(T)]$ between simulations. In contrast, with the outlier stations removed, the RMSE of σ^2 and $Var[S(T)]$ for the WGEN simulation is significantly larger than GiSTR. Recall that the outlier stations are removed based upon either the RMSE of μ or the RMSE of σ^2 produced by the entire WGEN simulation. Therefore, it is possible that while certain stations which are outlier stations for μ are not outlier stations for σ^2 which may cause the error of the WGEN simulation to be significantly larger for σ^2 and by aggregation $Var[S(T)]$. Regardless of the exact reason, this analysis indicates that the GiSTR simulation has lower error for most parameters in this “forecast” situation, even with outliers removed, which is consistent with the results shown in Chapter 4.

5.2.6. Dry and Wet Spell Frequency Analysis

The frequencies of the dry and wet spells of various lengths are also evaluated for each simulation. Given that the error of the temporal parameters is not different when the outlier stations are removed, as described in Section 5.2.5, the dry and wet spells analysis is not considered with the outliers removed. The dry and wet spells frequency in this analysis shows similar results to Section 4.2.6, which is also consistent with the results for the temporal parameters in Section 5.2.1 and Section 4.2.1. Figure 5.42 shows the observed frequency of wet spells of various lengths in January (a) compared to those produced by the WGEN simulation (b), and the GiSTR simulation (c) for all stations in the domain. For wet spells in January, the WGEN simulation underestimates the frequency of one and two day

wet spells compared to the GiSTR simulation. Figure 5.43 shows the observed frequency of wet spells of various lengths in July (a) compared to those produced by the WGEN simulation (b), and the GiSTR simulation (c) for all stations in the domain. Similarly for wet spells in January, WGEN also underestimates the frequency of one and two day wet spells compared the GiSTR simulation. For each month, the WGEN simulation also overestimates the frequency of wet spells which are 5 days and longer.

Figure 5.44 shows the observed frequency of dry spells of various lengths in January (a) compared to those produced by the WGEN simulation (b), and the GiSTR simulation (c) for all stations in the domain. Each downscaling simulation matches the pattern of the observed dry spell frequencies in the domain. Figure 5.45 shows the observed frequency of dry spells of various lengths in July (a) compared to those produced by the WGEN simulation (b), and the GiSTR simulation (c) for all stations in the domain. Similarly to dry spells in January, there is also little difference between the observed dry spell frequencies and the frequencies resulting from each downscaling simulation in July. However, in both months the one-day and two-day dry spells are slightly overestimated by the GiSTR downscaling simulation and slightly underestimated by the WGEN downscaling simulation. These for the dry and wet spells frequencies are consistent for all months, which is consistent with the results presented in Chapter 4. Given the tendency to of the WGEN simulation to overestimate the frequency of wet spells longer than five days and to underestimate the frequency of wet spells of one or two days and the increased RMSE of P_{11} it is likely that the WGEN simulation overestimates values of P_{11} across much of the domain. An overestimate of P_{11} would result make it more likely that the downscaling simulation would produce more long wet spells than short wet spells. It would also cause the frequency of long dry spells to be reduced, which is not apparent in this analysis given that the frequency of long dry spells for most months is smaller than 0.05 for dry spells longer than ten days. This is also shown by the results in the previous chapter.

Based upon the analysis in this section, the GiSTR simulation provides the lowest error for all parameters, with and without outlier stations, for this “forecast” period. This result is consistent with the results presented in Section 4.2. However, it is important to note

that this analysis and that in Chapter 4 assumes the patterns between generators would be the same when used in a downscaling context despite the relationship variation used. Both chapters have shown that this assumption may not be valid, and that the error associated with downscaling with WGEN may be more sensitive to the variation used. Future work in this area includes downscaling with all possible combinations of weather generator and downscaling variations to determine the influence of each variation on each weather generator in the context of downscaling.

5.3. Summary Discussion

5.3.1. Downscaling Simulation Comparison

Section 4.1 and 5.1 focuses on evaluating four variations of the two downscaling assumptions described in Chapter 2. These variations include the LWR-MOM, AW-DFIT, LWR-DFIT, and AW-MOM. The AW-MOM variation is the most commonly used variation for downscaling with weather generators. In Chapter 4, the downscaling focuses on the 1979-2000 time period while this chapter focuses on the 2001-2009 time period to assess the potential errors associated with seasonal forecasting. Several of the results in this chapter are consistent with the results presented in Chapter 4.

While the summary in Table 5.21 may show some discrepancies between the results of this chapter and Chapter 4, the T-tests done in each chapter show that there is no significant difference in the RMSE of the temporal parameters (P_{01} , P_{11} , π , and γ) and the spatial structure parameters (ρ , ρ_{ev} , ρ_{am} , and ρ_{ex}) between the downscaling variations. The difference between downscaling simulations is only significant for those parameters influencing precipitation amounts (μ , σ^2 , $E[S(T)]$, and $Var[S(T)]$). In Section 5.1, the AW-DFIT is shown to have the smallest RMSE across the domain for values of μ and $E[S(T)]$, which is consistent with the results from Section 4.1. In addition, the LWR-MOM and LWR-DFIT simulations were shown to have significantly larger RMSE for values of μ and

$E[S(T)]$ than the AW-DFIT and AW-MOM simulations. In contrast to the results of Chapter 4, the AW-MOM simulation is shown to have the smallest RMSE for values of σ^2 and the AW-DFIT simulation is shown to have the smallest RMSE for values of $Var[S(T)]$. However, the T-tests between each downscaling simulation shows that there is no significant difference in the RMSE between each simulation for σ^2 and $Var[S(T)]$, which is consistent with the results from the previous chapter.

As an additional analysis, the RMSE of each downscaling simulation in the 2001-2009 time period for each parameter are tested against the RMSE of the corresponding simulation for the 1979-2000 time period (from Chapter 4). For all the temporal parameters, the RMSE of each 2001-2009 downscaling simulation is significantly larger than the RMSE of the corresponding 1979-2000 downscaling simulation. For all the spatial structure parameters, the RMSE of each 2001-2009 downscaling simulation is not significantly different from the RMSE of the corresponding 1979-2000 downscaling simulation. Finally for the amount parameter, the RMSE of μ and $E[S(T)]$ for each downscaling simulation was not significantly different between time periods, while the RMSE of σ^2 and $Var[S(T)]$ for the 2001-2009 time period was significantly larger than the 1979-2000 time period. The RMSE of μ and $E[S(T)]$ is not significantly different between time periods, sources of error for these parameters are similar between time periods.

For the 1979 to 2000 time period the primary sources of error for μ resulted from localized areas where μ is overestimated or underestimated. For all the 2001-2009 simulations in this domain, the spatial patterns of the error were similar to the 1979-2000 time period, with a few exceptions. For instance, each LWR simulation shows a tendency to overestimate values of μ in the northeast corner of the domain, which is consistent with the results shown in Chapter 4. In contrast, each simulation shows a tendency to underestimate values of μ in the Carolinas and have less error in Florida in the summer months than the 1979-2000 simulation.

Each downscaling simulation does show some small localized sources of error for values of σ^2 in the domain, which is consistent with the spatial patterns shown in Chapter 4. These small sources are in the same locations, but the peak error is in September, compared

April and October for the 1979-2000 time period. The primary sources of error for values of σ^2 for the 2001-2009 time period are on the northern Gulf Coast, southern Alabama, and the east coast of North Carolina. While the stations with primary sources of error for σ^2 does not change between time periods, the magnitude of error between the two time periods becomes larger (by up to 200 mm² for some stations) which is the cause for the significant increase in the RMSE of both σ^2 and $Var[S(T)]$ for the 2001-2009 time period.

There are several possible reasons for why the RMSE of σ^2 and $Var[S(T)]$ is significantly larger in the 2001-2009 time period for each downscaling simulation. Consider the difference between the observed values of σ^2 in each time period. For the 1979-2000 time period, shown in Figure 5.46, the observed maximum value of σ^2 (in September) is 1600 mm² along the northern Gulf Coast and southern Alabama. For the 2001-2009 time period, shown in Figure 5.47, the observed maximum value of σ^2 (also in September) is 2300 mm² in eastern Maryland. In addition, between the two time periods there is a broad increase in σ^2 across the domain of 200 to 300 mm², which is most evident in summer. The increase in σ^2 in this time period may be related to decrease length of the time period. However, this increase may also be related to the several extreme events in the 2001-2009 time period. First, four hurricanes made landfall in Florida during the 2004 Atlantic hurricane season which produced extreme precipitation events, causing the observed values of σ^2 to rise and reducing the overestimation of σ^2 by each downscaling simulation in Florida. Second, the 2005 Atlantic hurricane season led to several land falling hurricanes in the Southeast U.S. which increases the number of extreme precipitation events and the observed values of σ^2 . Finally, these two wet events across the southeast were followed by the record drought of 2007. This dry situation led to a reduction of precipitation events, which would lead to more days of zero precipitation. The result is an extremely dry situation, which would also cause the observed values of σ^2 to increase. The combination of extreme wet events and extreme dry events during four years of the 2001-2009 time period caused the observed values of σ^2 to increase, which improves the values of σ^2 in Florida, but increases the underestimation of values of σ^2 in the rest of the domain, resulting in the significant increase in the RMSE of σ^2 between time periods. Given the direct relationship between σ^2 and $Var[S(T)]$, this increase

in the RMSE of $Var[S(T)]$ between time periods for each downscaling simulation is also caused by the increase in the observed values of σ^2 which is not captured by any of the downscaling simulations. In that way, one might conclude that 1979-2000 is not representative of the extremes observed in 2001-2009. However, it is also possible that using a shorter time period can inherently increase the observed variance. Finally, in the previous chapter, the edge and topographic effects were discussed as the possible causes for the LWR-MOM simulations and LWR-DFIT simulations having larger error for values of μ . In these simulations, there are very few over or underestimates more than 4 mm of μ on the inland edges of the domain. However, there are several overestimates of μ along the coastlines of the domain. This suggests a possible edge effect with regards to what station information is available for the LWR interpolation on the coast. Without information from buoys, there is a lack of offshore information about the required parameters for downscaling, which could be source of error in the interpolation. While this is a possibility, there are several coastal stations where the error of μ is less than 4 mm, which suggests that this coastal edge effect is not the only possible source of error. The high error of μ in the western Carolinas and northeast Georgia suggests that topography may affect the LWR interpolation, leading to overestimates of μ in summer. However, if topography alone caused the error in this region, then it would also be likely that higher error of μ would exist in the northeast corner of domain and along the Appalachian Mountains since there is also complex topography in this region. Therefore, topography may not be the only source of error for values of μ in the LWR simulations. Given that the LWR interpolation makes use the elevation at each station and at each CFSR grid point, it is possible that the modeled land surface elevation in CFSR is too coarse, which could lead to erroneous interpolation when used in the LWR variations. Regardless of the sources of error or the difference between time periods or the comparable error between variations, there is one common pattern. For most parameters, there are combinations of the two downscaling assumptions which have lower RMSE than the commonly used AW-MOM combination.

5.3.2. Weather Generator Comparison – Downscaling Context

The second analysis presented in this chapter is an analysis of the difference between weather generators used in downscaling when the downscaling assumptions are fixed to the LWR-MOM assumptions. This analysis is done to determine if the error patterns between weather generators when they are used for downscaling are consistent with the comparison of weather generators presented in Chapter 3 for a non-downscaling context. The results from this analysis, summarized in Table 5.22, mirror the results shown in Chapter 4. In addition, those stations which had outlying error values for the WGEN simulation were removed. These results of this additional analysis, summarized in Table 5.23, also reflect most of the results presented for the same analysis in Chapter 4.

While the results for this chapter for the spatial parameters, σ^2 , and $Var[S(T)]$ are the same as the results for the comparison of weather generators in Chapter 3, the remaining parameters show the same pattern described in Chapter 4. For three of the temporal parameters (P_{11} , π , and γ) the WGEN simulation has significantly larger RMSE than the GiSTR simulation, which is in contrast to the results in Chapter 3. The error in these parameters is reflected in the error in the frequency of dry and wet spells produced by the WGEN downscaling simulation. The overestimation of the frequency of wet spells longer than five days suggests that values of P_{11} are overestimated in this domain. Since this causes to the precipitation amount generation process to be used more often, this may cause the errors in the values of all the amount parameters, causing an increase in the error of all the amount parameters when WGEN is used for downscaling. These same issues do not affect the GiSTR downscaling simulation which again suggests that the results from WGEN when used for downscaling are sensitive to the downscaling assumptions used. There is one caveat in this result. While the precipitation data at the cross-validation stations were checked for quality, it is possible that some of the observations used are poor quality which passed the quality control procedure used.

The results of the evaluation in this chapter and in Chapters 3 and 4 show distinct trends in the error of all the downscaling and control simulations evaluated. However, the

results from this evaluation also suggest several recommendations for future work involving weather generator downscaling. In addition, it is important to note that the CFSR is not used for seasonal forecasting. Therefore, while the results of this section provide insight into the potential errors associated with using each generator or downscaling variation in downscaling seasonal forecasts, it is recommended to expand this analysis to the CFS Reforecast. The CFS Reforecast data comes from the same seasonal forecast model used operationally by the Climate Prediction Center. Therefore, expanding this analysis to the Reforecast will help to determine the ability of each generator and simulation to downscale seasonal forecasts of precipitation in the region. The next chapter discusses the conclusions drawn from this research, the caveats, and the future work possible in studying downscaling with weather generators, and the potential applications to seasonal forecasting and crop, hydrology, and climate change modeling.

Table 5.1. RMSE Comparison for each 2001-2009 downscaling simulation across the Southeast U.S. for value of P_{0I} .

month	LWR-MOM	AW-DFIT	LWR-DFIT	AW-MOM
1	0.137	0.141	0.157	0.135
2	0.0981	0.0991	0.0987	0.0995
3	0.106	0.105	0.112	0.113
4	0.105	0.115	0.114	0.113
5	0.0924	0.0922	0.0883	0.0955
6	0.0840	0.0847	0.0958	0.0879
7	0.0805	0.0740	0.0815	0.0782
8	0.104	0.111	0.105	0.106
9	0.132	0.124	0.128	0.140
10	0.0984	0.0855	0.113	0.102
11	0.127	0.131	0.114	0.133
12	0.123	0.122	0.124	0.120
average	0.107	0.107	0.111	0.110

Table 5.2. RMSE Comparison for each 2001-2009 downscaling simulation across the Southeast U.S. for value of P_{II} .

month	LWR-MOM	AW-DFIT	LWR-DFIT	AW-MOM
1	0.122	0.115	0.122	0.121
2	0.135	0.128	0.127	0.140
3	0.106	0.100	0.0941	0.0940
4	0.117	0.114	0.112	0.123
5	0.120	0.121	0.136	0.131
6	0.119	0.132	0.136	0.127
7	0.128	0.129	0.143	0.126
8	0.125	0.124	0.135	0.139
9	0.132	0.131	0.134	0.128
10	0.147	0.129	0.132	0.133
11	0.0953	0.107	0.104	0.114
12	0.122	0.113	0.119	0.116
average	0.122	0.120	0.124	0.124

Table 5.3. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the Markov Transition Probabilities (P_{0I} and P_{1I}). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second. The Difference in the Average RMSE indicates is the difference between the average RMSE of the first simulation and the average RMSE of the second simulation.

Simulations Compared		P_{0I}		P_{1I}	
		Difference in the Average RMSE	P-value	Difference in the Average RMSE	P-value
LWR-MOM	AW-DFIT	0.000153	0.492	0.00231	0.320
LWR-MOM	LWR-DFIT	-0.00387	0.314	-0.00194	0.370
LWR-MOM	AW-MOM	-0.00309	0.346	-0.00202	0.353
AW-DFIT	LWR-DFIT	-0.00402	0.316	-0.00425	0.212
AW-DFIT	AW-MOM	-0.00324	0.347	-0.00434	0.183
LWR-DFIT	AW-MOM	0.000784	0.462	-8.60E-05	0.494

Table 5.4. Summary of Two-sample T-test results on the RMSE of the Markov Transition Probabilities (P_{0I} and P_{1I}) for each 2001-2009 downscaling simulation with the corresponding 1979-2000 downscaling simulation. P-values in red indicate that the RMSE of the 2001-2009 simulation is significantly larger than the 1979-2000 simulation.

Simulation	P_{0I}		P_{1I}	
	Difference in Average RMSE	P-value	Difference in Average RMSE	P-value
LWR-MOM	0.0517	2.370E-07	0.0617	4.363E-11
AW-DFIT	0.0504	6.437E-07	0.0607	2.875E-13
LWR-DFIT	0.0561	1.019E-07	0.0621	4.501E-10
AW-MOM	0.0525	2.393E-07	0.0624	8.215E-12

Table 5.5. RMSE Comparison for each 2001-2009 downscaling simulation across the Southeast U.S. for values of the conditional probability of rain (π).

month	LWR-MOM	AW-DFIT	LWR-DFIT	AW-MOM
1	0.135	0.139	0.147	0.132
2	0.104	0.105	0.102	0.112
3	0.0820	0.0852	0.0876	0.0913
4	0.105	0.108	0.107	0.118
5	0.0950	0.0945	0.0934	0.0950
6	0.0851	0.0833	0.101	0.0821
7	0.0924	0.0881	0.0977	0.0887
8	0.0999	0.104	0.100	0.100
9	0.109	0.103	0.106	0.113
10	0.104	0.0921	0.114	0.0951
11	0.113	0.119	0.108	0.131
12	0.118	0.111	0.118	0.112
average	0.103	0.103	0.107	0.106

Table 5.6. RMSE Comparison for each 2001-2009 downscaling simulation across the Southeast U.S. for values of persistence (γ).

Month	LWR-MOM	AW-DFIT	LWR-DFIT	AW-MOM
1	0.138	0.128	0.148	0.142
2	0.136	0.119	0.133	0.123
3	0.163	0.151	0.154	0.151
4	0.128	0.139	0.140	0.128
5	0.137	0.142	0.153	0.155
6	0.141	0.160	0.145	0.161
7	0.130	0.129	0.139	0.129
8	0.153	0.156	0.164	0.173
9	0.208	0.201	0.204	0.210
10	0.170	0.149	0.154	0.174
11	0.146	0.153	0.141	0.146
12	0.132	0.134	0.131	0.134
average	0.148	0.147	0.150	0.152

Table 5.7. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the unconditional probability of rain (π) and persistence (γ). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second. The Difference in the Average RMSE is the difference between the average RMSE of the first simulation and the average RMSE of the second simulation.

Simulations Compared		π		γ	
		Difference in the Average RMSE	P-value	Difference in the Average RMSE	P-value
LWR-MOM	AW-DFIT	0.000735	0.453	0.00171	0.426
LWR-MOM	LWR-DFIT	-0.00341	0.291	-0.00196	0.412
LWR-MOM	AW-MOM	-0.00245	0.351	-0.00376	0.351
AW-DFIT	LWR-DFIT	-0.00414	0.260	-0.00366	0.332
AW-DFIT	AW-MOM	-0.00319	0.315	-0.00547	0.284
LWR-DFIT	AW-MOM	0.000952	0.442	-0.00181	0.422

Table 5.8. Summary of Two-sample T-test results on the RMSE of the unconditional probability of rain (π) and persistence (γ) for each 2001-2009 downscaling simulation with the corresponding 1979-2000 downscaling simulation. P-values in red indicate that the RMSE of the 2001-2009 simulation is significantly larger than the 1979-2000 simulation.

Simulation	π		γ	
	Difference in Average RMSE	P-value	Difference in Average RMSE	P-value
LWR-MOM	0.0738	8.7E-12	0.0418	1.69E-05
AW-DFIT	0.0720	2.44E-11	0.0401	1.13E-05
LWR-DFIT	0.0756	8.8E-12	0.0443	1.2E-06
AW-MOM	0.0737	5.84E-11	0.0429	3.1E-05

Table 5.9. Average RMSE for each 2001-2009 downscaling simulation for each correlation matrix.

Parameter	LWR-MOM	AW-DFIT	LWR-DFIT	AW-MOM
ρ	0.144	0.143	0.148	0.145
ρ_{ev}	0.0991	0.0980	0.0993	0.0878
ρ_{am}	0.180	0.178	0.188	0.173
ρ_{ex}	0.121	0.124	0.123	0.117

Table 5.10. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the mean daily nonzero precipitation (μ) and the variance of the daily nonzero precipitation (σ^2). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second. The Difference in the Average RMSE is the difference between the average RMSE of the first simulation and the average RMSE of the second simulation.

Simulations Compared		μ Difference in the Average RMSE (mm)	P-value	σ^2 Difference in the Average RMSE (mm ²)	P-value
LWR-MOM	AW-DFIT	2.113	0.000443	-2.813	0.467
LWR-MOM	LWR-DFIT	0.148	0.395	-4.634	0.442
LWR-MOM	AW-MOM	1.598	0.00448	5.700	0.430
AW-DFIT	LWR-DFIT	-1.965	5.49E-06	-1.820	0.478
AW-DFIT	AW-MOM	-0.515	0.0590	8.513	0.401
LWR-DFIT	AW-MOM	1.450	0.00052	10.333	0.373

Table 5.11. Summary of Two-sample T-test results on the RMSE the mean daily nonzero precipitation (μ) and the variance of the daily nonzero precipitation(σ^2). P-values in red indicate that the RMSE of the 2001-2009 simulation is significantly larger than the corresponding RMSE of the 1979-2000 simulation.

Simulation	μ		σ^2	
	Difference in Average RMSE (mm)	P-value	Difference in Average RMSE (mm ²)	P-value
LWR-MOM	0.784	0.124	70.194	0.00708
AW-DFIT	0.255	0.236	48.054	0.0549
LWR-DFIT	0.735	0.0904	70.650	0.00568
AW-MOM	0.323	0.251	63.988	0.0145

Table 5.12. Summary of Two-sample T-test results comparing downscaling simulations for the RMSE of the average total precipitation ($E[S(T)]$) and the inter-annual variability($Var[S(T)]$). P-values in red indicate that either the first simulation has a significantly larger or smaller RMSE than the second. The Difference in the Average RMSE indicates is the difference between the average RMSE of the first simulation and the average RMSE of the second simulation.

Simulations Compared		$E[S(T)]$ Difference in the Average RMSE (mm)	P-value	$Var[S(T)]$ Difference in the Average RMSE (mm ²)	P-value
LWR-MOM	AW-DFIT	26.159	0.00184	749.914	0.0590
LWR-MOM	LWR-DFIT	3.875	0.324	86.950	0.425
LWR-MOM	AW-MOM	16.953	0.0258	715.727	0.0625
AW-DFIT	LWR-DFIT	-22.284	0.00115	-662.96	0.0674
AW-DFIT	AW-MOM	-9.206	0.0769	-34.186	0.468
LWR-DFIT	AW-MOM	13.078	0.0346	628.777	0.0710

Table 5.13. Summary of Two-sample T-test results on the RMSE of the average total precipitation ($E[S(T)]$) and the inter-annual variability ($Var[S(T)]$) for each 2001-2009 simulation with the corresponding 1979-2000 simulation. P-values in red indicate that the RMSE of the 2001-2009 simulation is significantly larger than the 1979-2000 simulation.

Simulation	$E[S(T)]$ Difference in Average RMSE		$Var[S(T)]$ Difference in Average RMSE	
	(mm)	P-value	(mm ²)	P-value
LWR-MOM	14.242	0.0776	996.405	0.0115
AW-DFIT	9.5101	0.0912	634.615	0.0339
LWR-DFIT	11.789	0.0768	917.913	0.00980
AW-MOM	10.565	0.0876	757.169	0.0146

Table 5.14. Comparison of the RMSE of the Markov Transition Probabilities (P_{01} and P_{11}) between the GiSTR downscaling simulation and the WGEN downscaling simulation. Time period of 2001-2009.

Month	P_{01}		P_{11}	
	GiSTR	WGEN	GiSTR	WGEN
1	0.137	0.149	0.122	0.284
2	0.0981	0.116	0.135	0.333
3	0.106	0.114	0.106	0.247
4	0.105	0.120	0.117	0.235
5	0.0924	0.0832	0.120	0.221
6	0.0840	0.104	0.119	0.204
7	0.0805	0.0941	0.128	0.215
8	0.104	0.0842	0.125	0.220
9	0.132	0.0880	0.132	0.201
10	0.0984	0.0880	0.147	0.233
11	0.126	0.106	0.0953	0.249
12	0.123	0.132	0.122	0.292
average	0.107	0.107	0.122	0.244

Table 5.15. Comparison of the RMSE of the unconditional probability of rain (π) and persistence (γ) between the GiSTR downscaling simulation and the WGEN downscaling simulation. Time period of 2001-2009.

Month	π		γ	
	GiSTR	WGEN	GiSTR	WGEN
1	0.135	0.269	0.138	0.218
2	0.104	0.245	0.136	0.299
3	0.0820	0.215	0.163	0.221
4	0.105	0.180	0.128	0.252
5	0.0950	0.152	0.137	0.232
6	0.0851	0.153	0.141	0.242
7	0.0925	0.130	0.130	0.260
8	0.0999	0.144	0.153	0.243
9	0.109	0.163	0.208	0.196
10	0.104	0.181	0.170	0.218
11	0.113	0.197	0.146	0.243
12	0.118	0.230	0.132	0.275
average	0.103	0.188	0.148	0.242

Table 5.16. Summary of the two-sample T-test results for the Markov Transition Probabilities (P_{01} and P_{11}), the unconditional probability of rain (π), and persistence (γ). P-values in red indicate that the RMSE of that parameter for WGEN is significantly larger than the RMSE of that parameter for GiSTR. Time period of 2001-2009.

Parameter	Difference in Average RMSE (WGEN - GiSTR)	P-value
P_{01}	-0.00065	0.468
P_{11}	0.122	8.53E-08
π	0.0850	1.17E-05
γ	0.0932	6.83E-09

Table 5.17. Summary of two-sample T-test results for the RMSE of the correlation matrix of precipitation (ρ), the correlation matrix of precipitation events (ρ_{ev}), the correlation matrix of precipitation amounts(ρ_{am}), and the correlation matrix of precipitation extreme events(ρ_{ex}). P-values in red indicate that the RMSE of that parameter for WGEN is significantly larger than the RMSE for GiSTR.

Parameter	Difference in Average RMSE (WGEN - GiSTR)	P-values
ρ	0.123	0.000191
ρ_{ev}	0.209	6.04E-08
ρ_{am}	0.102	2.99E-06
ρ_{ex}	0.0417	0.00962

Table 5.18. Summary of two-sample T-test results from the RMSE of the correlation matrix of precipitation (ρ), the correlation matrix of precipitation events (ρ_{ev}), the correlation matrix of precipitation amounts(ρ_{am}), and the correlation matrix of precipitation extreme events(ρ_{ex})between each 2001-2009 simulation and its Control from Chapter 3. The Difference in the Average RMSE is the Simulation – Control. P-values in red indicate that the RMSE of that simulation is larger than the Control simulations.

Parameter	GiSTR		WGEN	
	Difference in Average RMSE	P-value	Difference in Average RMSE	P-value
ρ	0.0178	0.142	0.00482	0.446
ρ_{ev}	0.0262	1.11E-05	0.00518	0.423
ρ_{am}	0.0453	6.16E-08	0.0383	0.0432
ρ_{ex}	0.0164	0.0773	-0.119	0.0873

Table 5.19. Summary of the two-sample T-test results for the mean daily nonzero precipitation (μ), the variance of the daily nonzero precipitation (σ^2), the average total precipitation ($E[S(T)]$), and the inter-annual variability ($Var[S(T)]$). P-values in red indicate that the RMSE of that parameter for the WGEN simulation is significantly larger than the RMSE of that parameter for GiSTR simulation.

Parameter	Difference in Average RMSE (WGEN - GiSTR)	P-value
μ	1.451 mm	0.242
σ^2	160.02 mm ²	0.0863
$E[S(T)]$	6.476 mm	0.362
$Var[S(T)]$	13512.5 mm ²	0.105

Table 5.20. Summary of the two-sample T-test results for the mean daily nonzero precipitation (μ), the variance of the daily nonzero precipitation (σ^2), the average total precipitation ($E[S(T)]$), and the inter-annual variability ($Var[S(T)]$). P-values in red indicate that the RMSE of that parameter for the WGEN simulation is significantly larger than the RMSE of that parameter for GiSTR simulation. Outlier stations removed.

Parameter	Difference in Average RMSE (WGEN - GiSTR)	P-value
μ	0.449	0.463
σ^2	70.756	0.0254
$E[S(T)]$	1.454	0.456
$Var[S(T)]$	2890.649	0.0209

Table 5.21. Summary of results from Chapters 4 and 5 for the downscaling simulation comparison by parameter.

Parameter	Best Downscaling Variation from Chapter 4	Best Downscaling Variation from Chapter 5
Probability of a wet day following a dry day (P_{01})	LWR-MOM and LWR-DFIT	LWR-MOM and AW-DFIT
Probability of a wet day following a wet day (P_{11})	AW-DFIT	AW-DFIT
Unconditional Probability of Rain (π)	LWR-MOM	LWR-MOM and AW-DFIT
Persistence (γ)	LWR-MOM	LWR-MOM and AW-DFIT
Precipitation correlation matrix (ρ)	LWR-MOM	LWR-MOM and AW-DFIT
Precipitation event correlation matrix (ρ_{ev})	LWR-MOM	LWR-MOM and AW-DFIT
Precipitation amount correlation matrix (ρ_{am})	LWR-MOM	LWR-MOM and AW-DFIT
Precipitation extreme event correlation matrix (ρ_{ex})	LWR-MOM	LWR-MOM and AW-DFIT
Mean daily nonzero precipitation amounts (μ)	AW-DFIT	AW-DFIT
Variance of daily nonzero precipitation amounts (σ^2)	LWR-MOM	LWR-MOM and AW-MOM
Average monthly total precipitation ($E[S(T)]$)	AW-DFIT	AW-DFIT
Variance of the monthly total precipitation / Inter-annual Variability ($Var[S(T)]$)	AW-MOM	AW-DFIT and AW- MOM

Table 5.22. Comparison Table between the Control (full details in Chapter 3), 1979-2000 downscaling simulations, and 2001-2009 downscaling simulations in the Southeast U.S. Comparison includes GiSTR and WGEN only.

Parameter	Best control simulation (Chapter 3)	Best Downscaling simulation (Chapter 4)	Best Downscaling Simulation (Chapter 5)
Probability of a wet day following a dry day (P_{01})	WGEN	both	both
Probability of a wet day following a wet day (P_{11})	WGEN	GiSTR	GiSTR
Unconditional Probability of Rain (π)	WGEN	GiSTR	GiSTR
Persistence (γ)	WGEN	GiSTR	GiSTR
Precipitation correlation matrix (ρ)	GiSTR	GiSTR	GiSTR
Precipitation event correlation matrix (ρ_{ev})	GiSTR	GiSTR	GiSTR
Precipitation amount correlation matrix (ρ_{am})	GiSTR	GiSTR	GiSTR
Precipitation extreme event correlation matrix (ρ_{ex})	GiSTR	GiSTR	GiSTR
Mean daily nonzero precipitation amounts (μ)	WGEN	both	Both
Variance of daily nonzero precipitation amounts (σ^2)	GiSTR	GiSTR	GiSTR
Average monthly total precipitation ($E[S(T)]$)	WGEN	both	Both
Variance of the monthly total precipitation / Inter-annual Variability ($Var[S(T)]$)	GiSTR	GiSTR	GiSTR

Table 5.23. Comparison Table between downscaling simulations in the Southeast U.S. Comparison includes GiSTR and WGEN only, with outlier stations removed.

Parameter	Best downscaling simulation (Chapter 4)	Best downscaling simulation (Chapter 5)
Probability of a wet day following a dry day (P_{01})	both	both
Probability of a wet day following a wet day (P_{11})	GiSTR	GiSTR
Unconditional Probability of Rain (π)	GiSTR	GiSTR
Persistence (γ)	GiSTR	GiSTR
Precipitation correlation matrix (ρ)	GiSTR	GiSTR
Precipitation event correlation matrix (ρ_{ev})	GiSTR	GiSTR
Precipitation amount correlation matrix (ρ_{am})	GiSTR	GiSTR
Precipitation extreme event correlation matrix (ρ_{ex})	GiSTR	GiSTR
Mean daily nonzero precipitation amounts (μ)	WGEN	GiSTR
Variance of daily nonzero precipitation amounts (σ^2)	GiSTR	GiSTR
Average monthly total precipitation ($E[S(T)]$)	GiSTR	GiSTR
Variance of the monthly total precipitation / Inter-annual Variability ($Var[S(T)]$)	WGEN	GiSTR

Cross Validation Station Map: 2001 - 2009

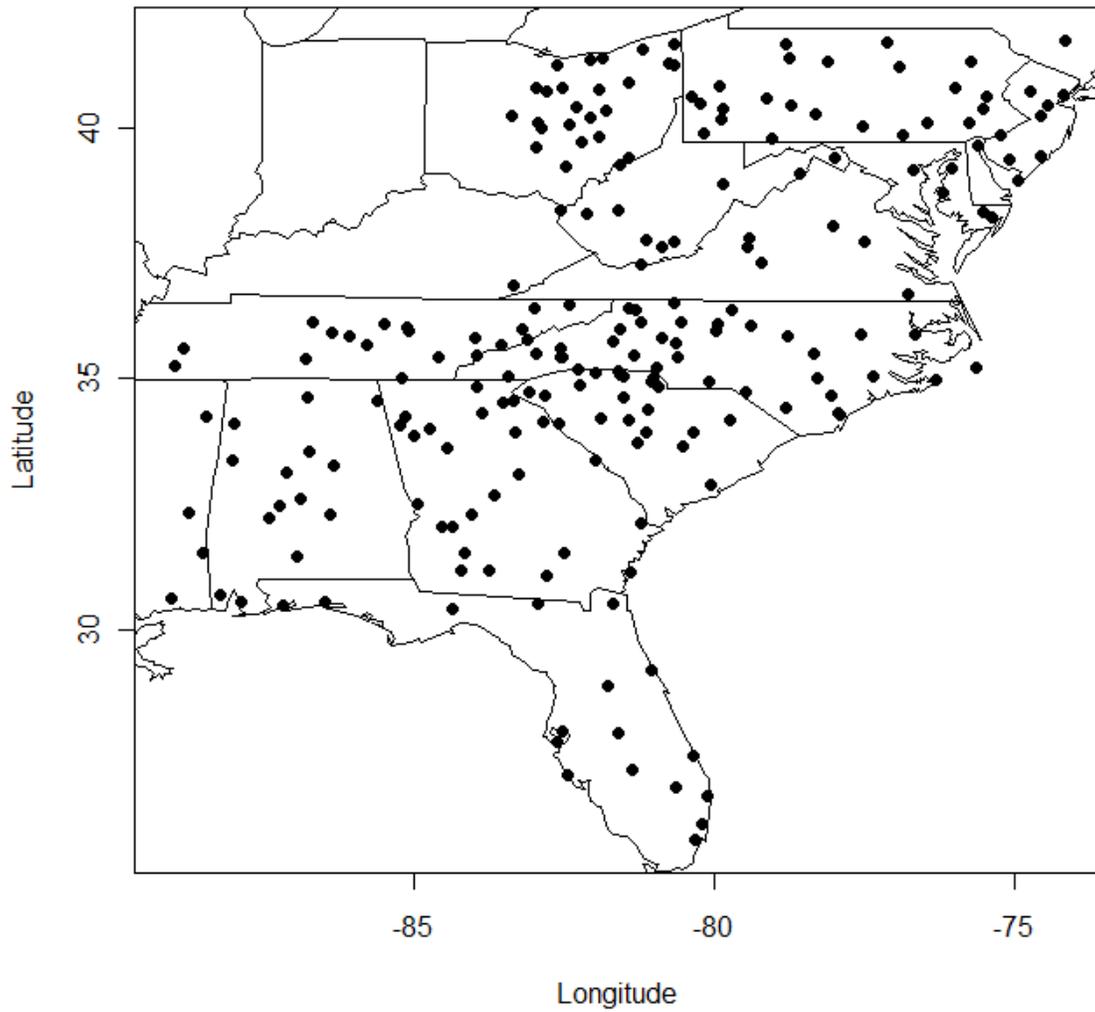


Figure 5.1. Cross Validation station map for 2001-2009.

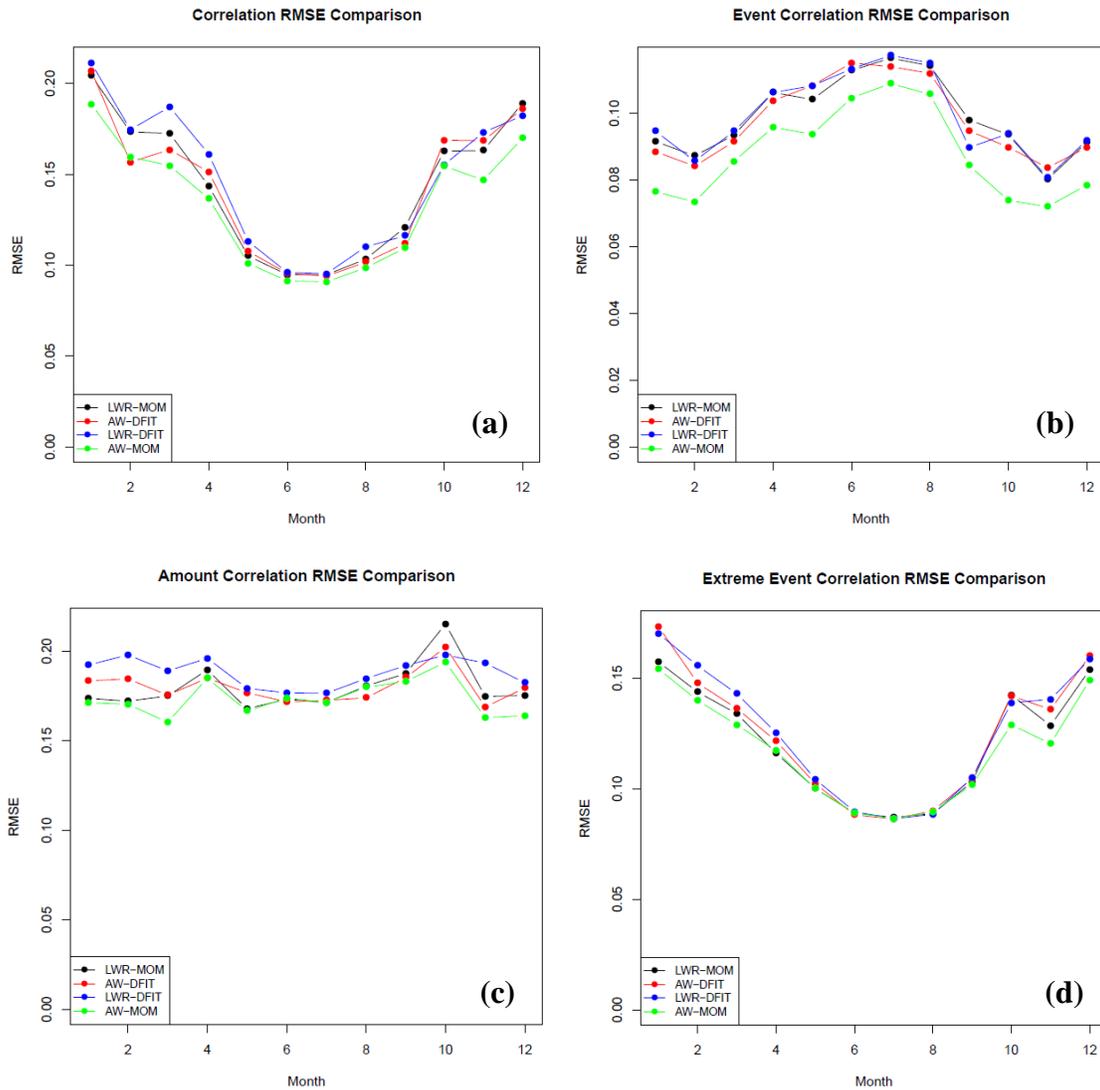


Figure 5.2. RMSE of each downscaling simulation by month for the correlation matrix of precipitation (ρ) (a), the correlation matrix of precipitation events (ρ_{ev}) (b), the correlation matrix of precipitation amounts (ρ_{am}) (c), and the correlation matrix of precipitation extreme events (ρ_{ex}) (d).

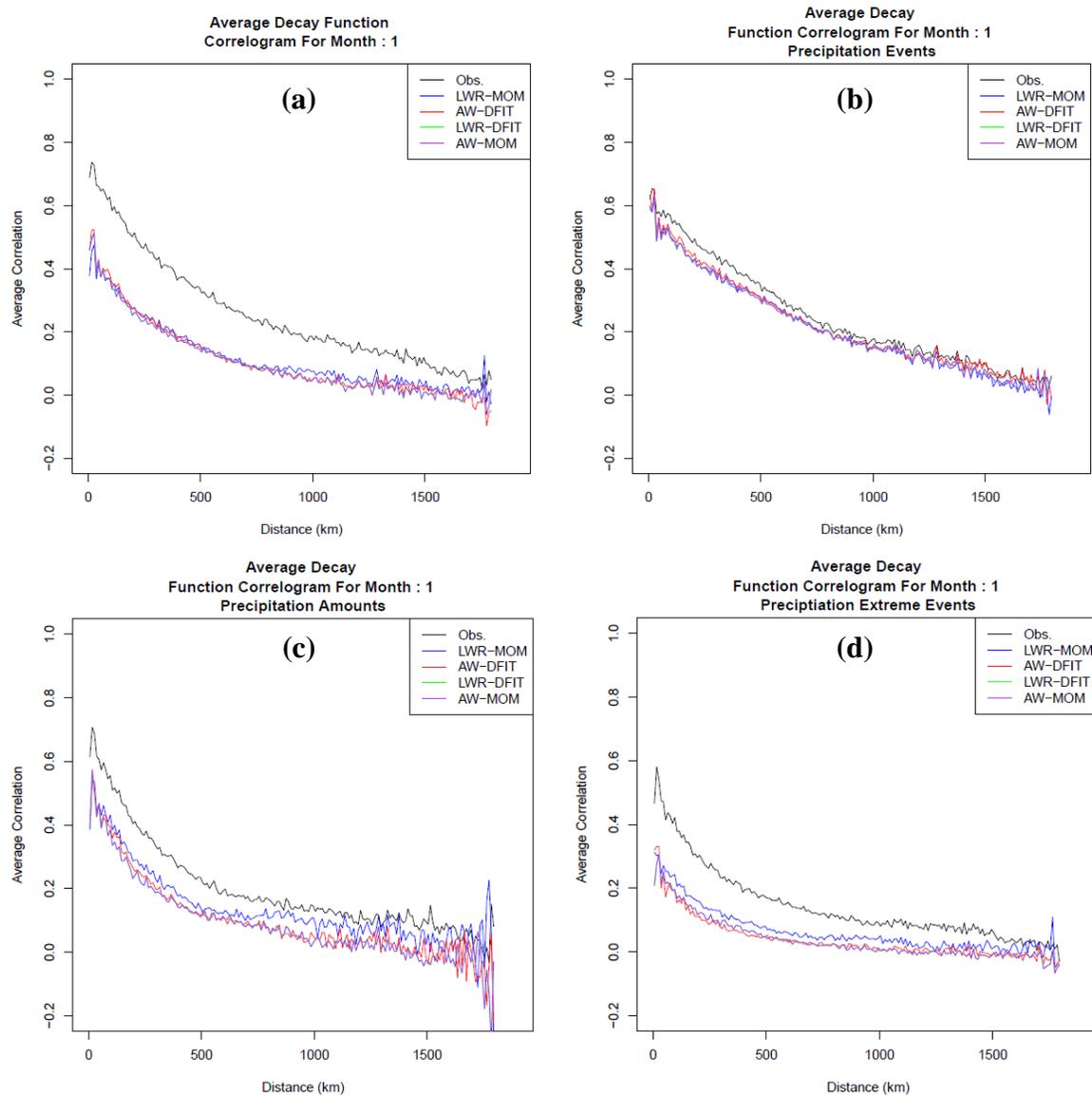


Figure 5.3. Average Decay Function Correlogram for each simulation compared to observations for January for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 2001-2009 time period.

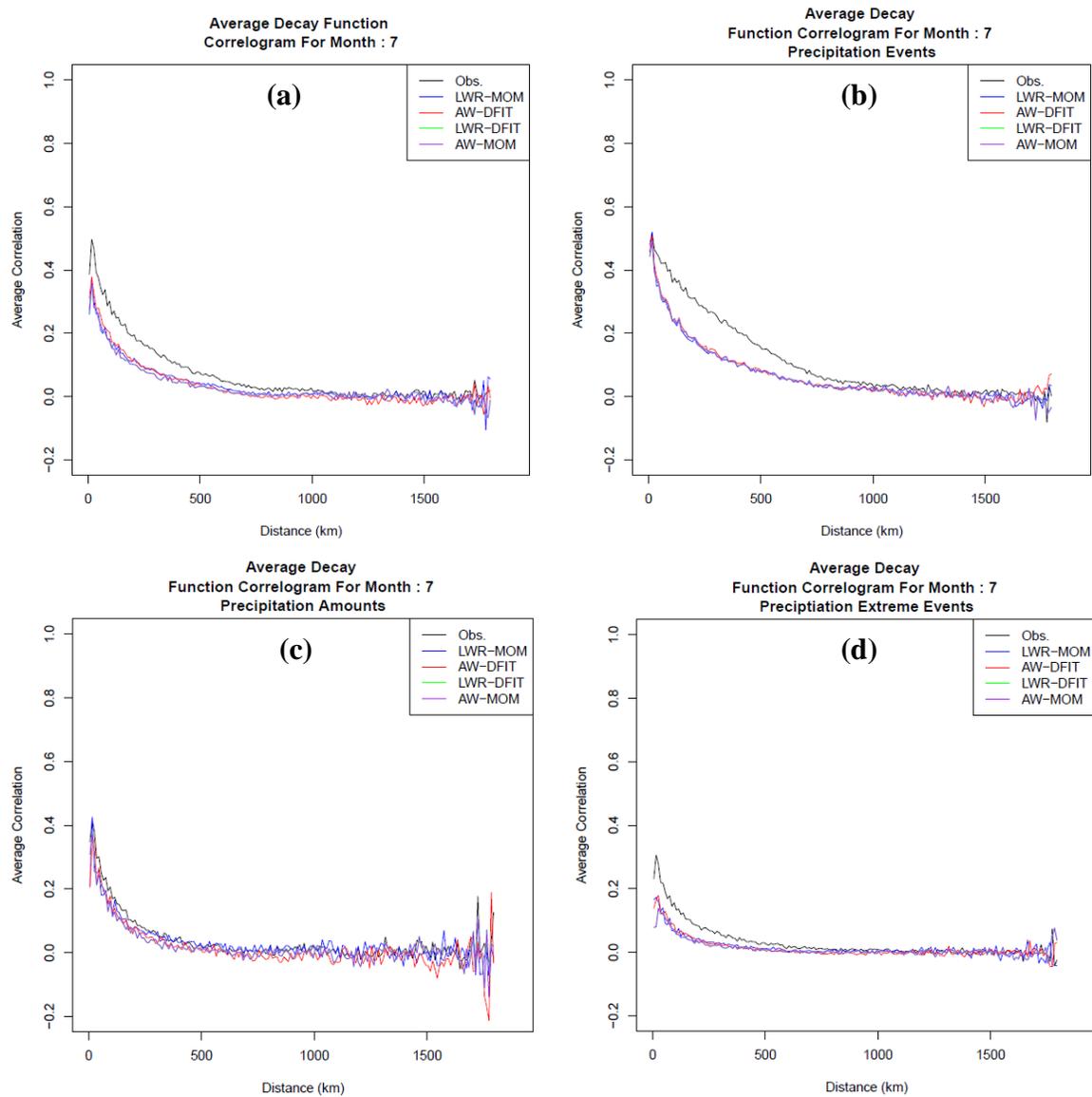


Figure 5.4. Average Decay Function Correlogram for each simulation compared to observations for July for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 2001-2009 time period.

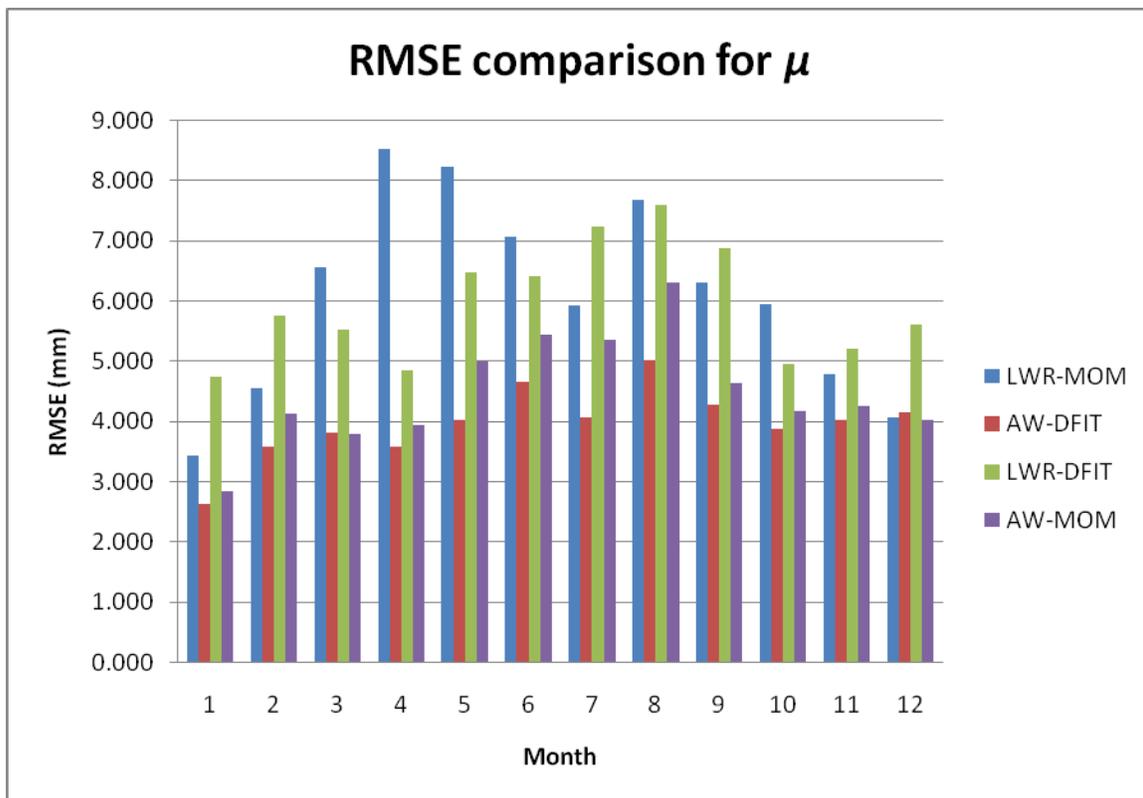
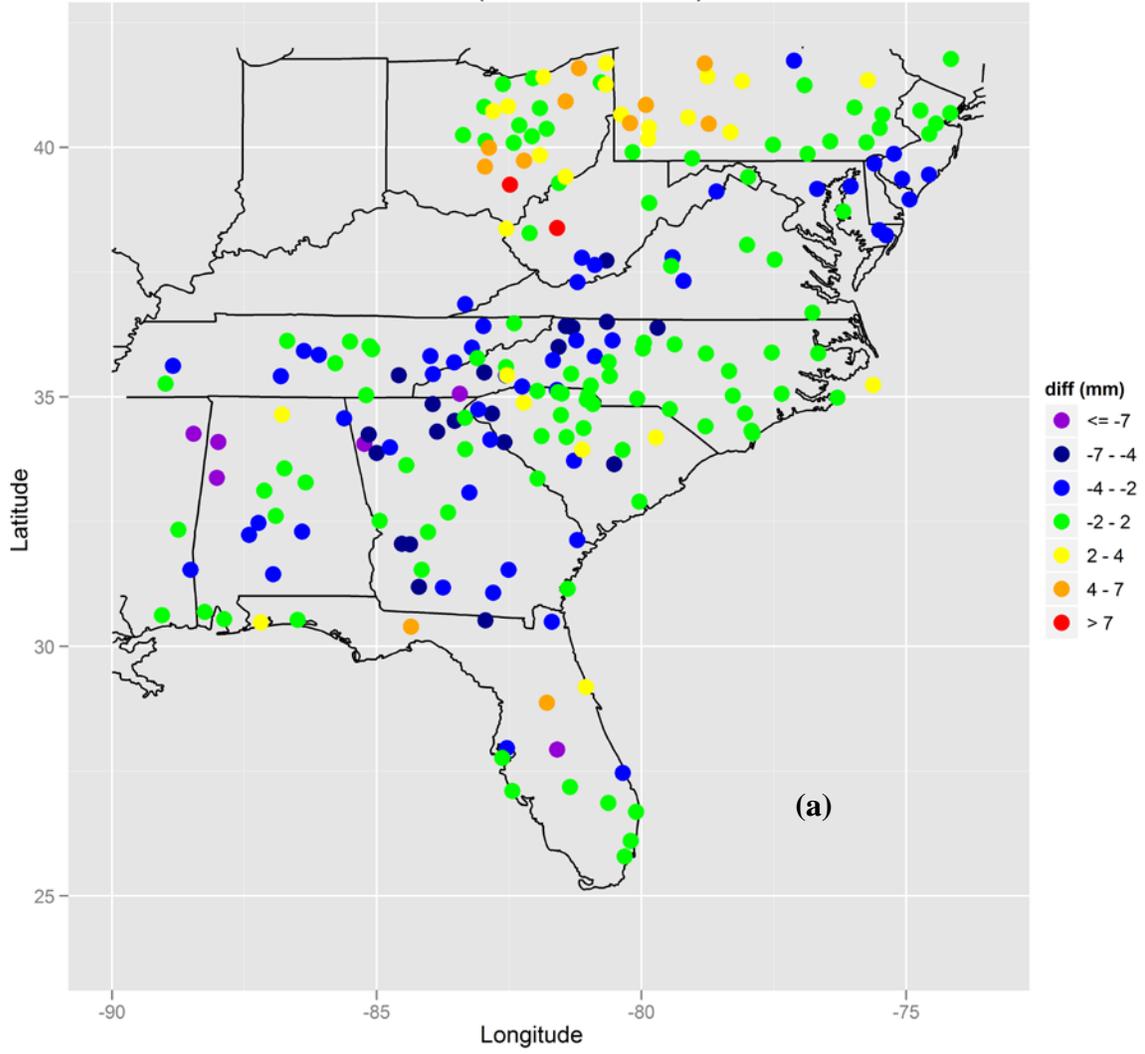
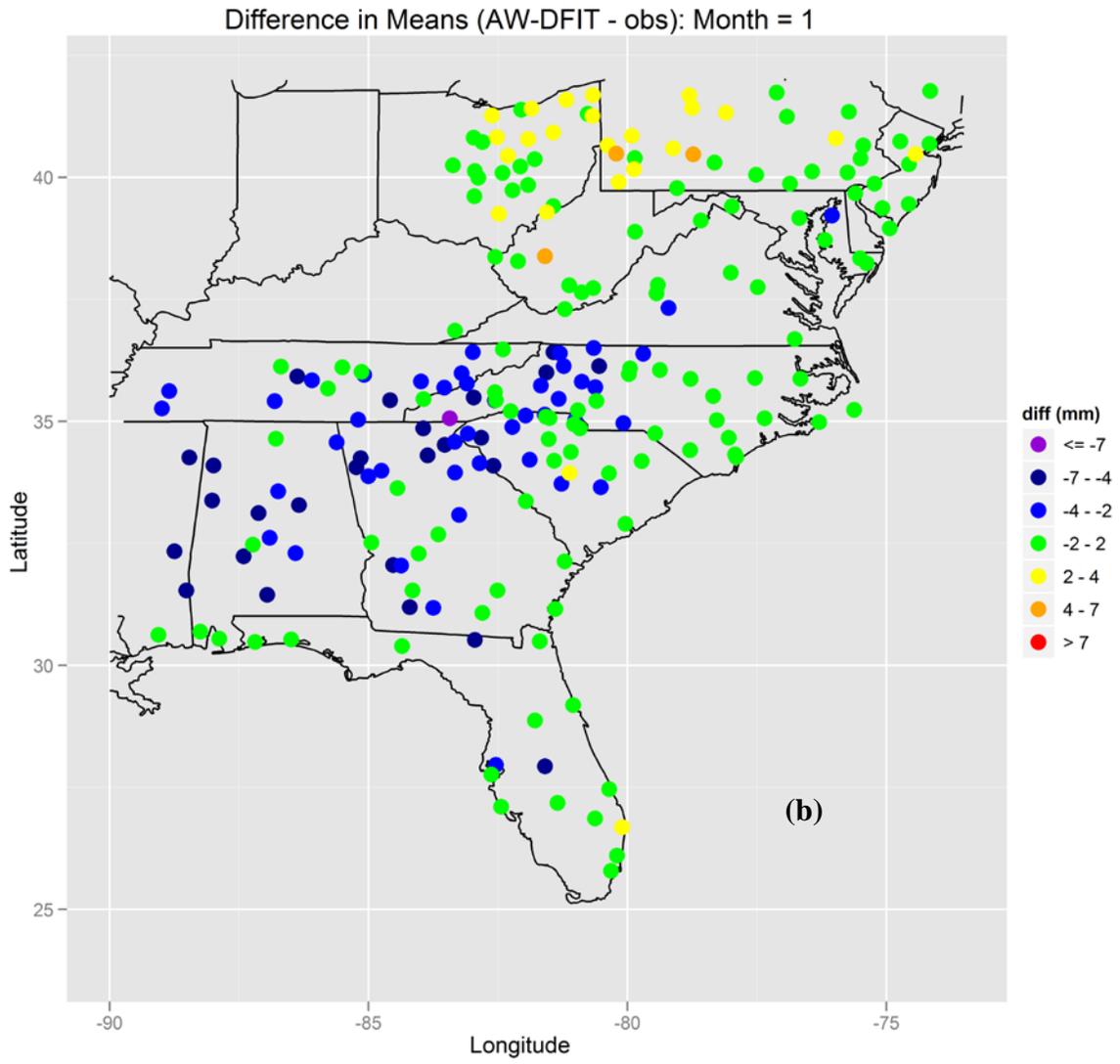


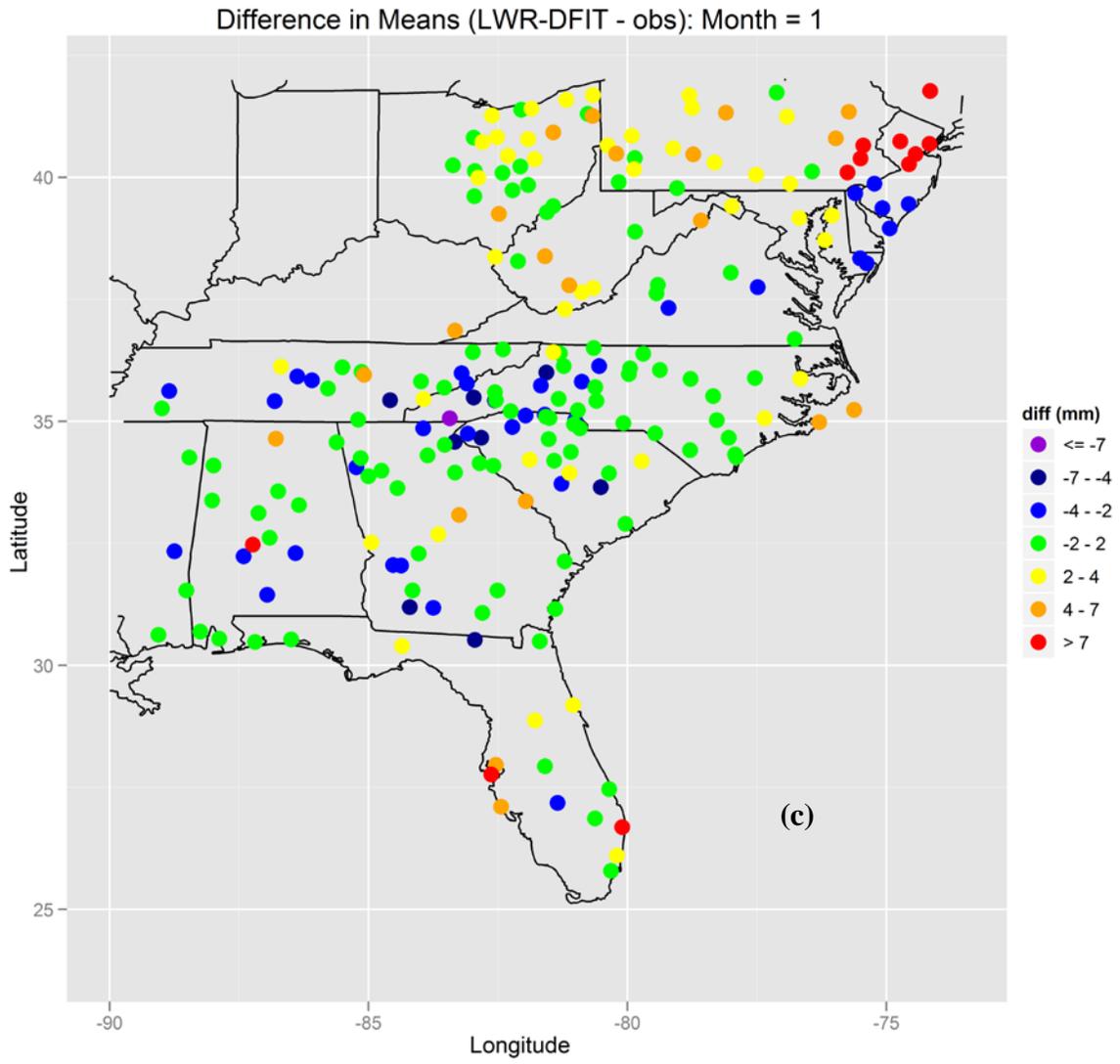
Figure 5.5. RMSE Comparison between 2001-2009 downscaling simulations for mean daily nonzero precipitation (μ).

Figure 5.6. Difference between generated and observed values of mean daily nonzero precipitation (μ) for each of the 2001-2009 downscaling simulations: LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for January.

Difference in Means (LWR-MOM - obs): Month = 1







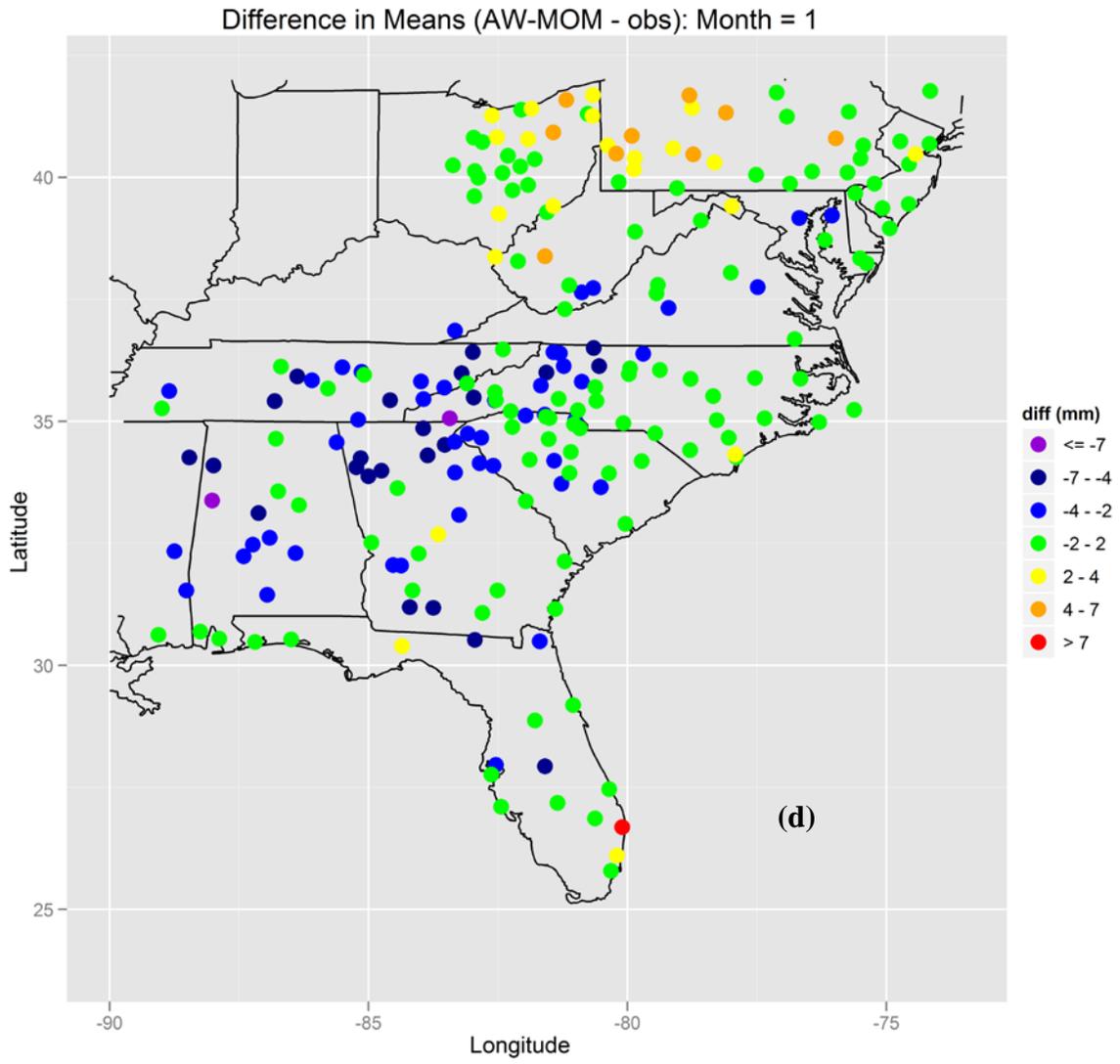
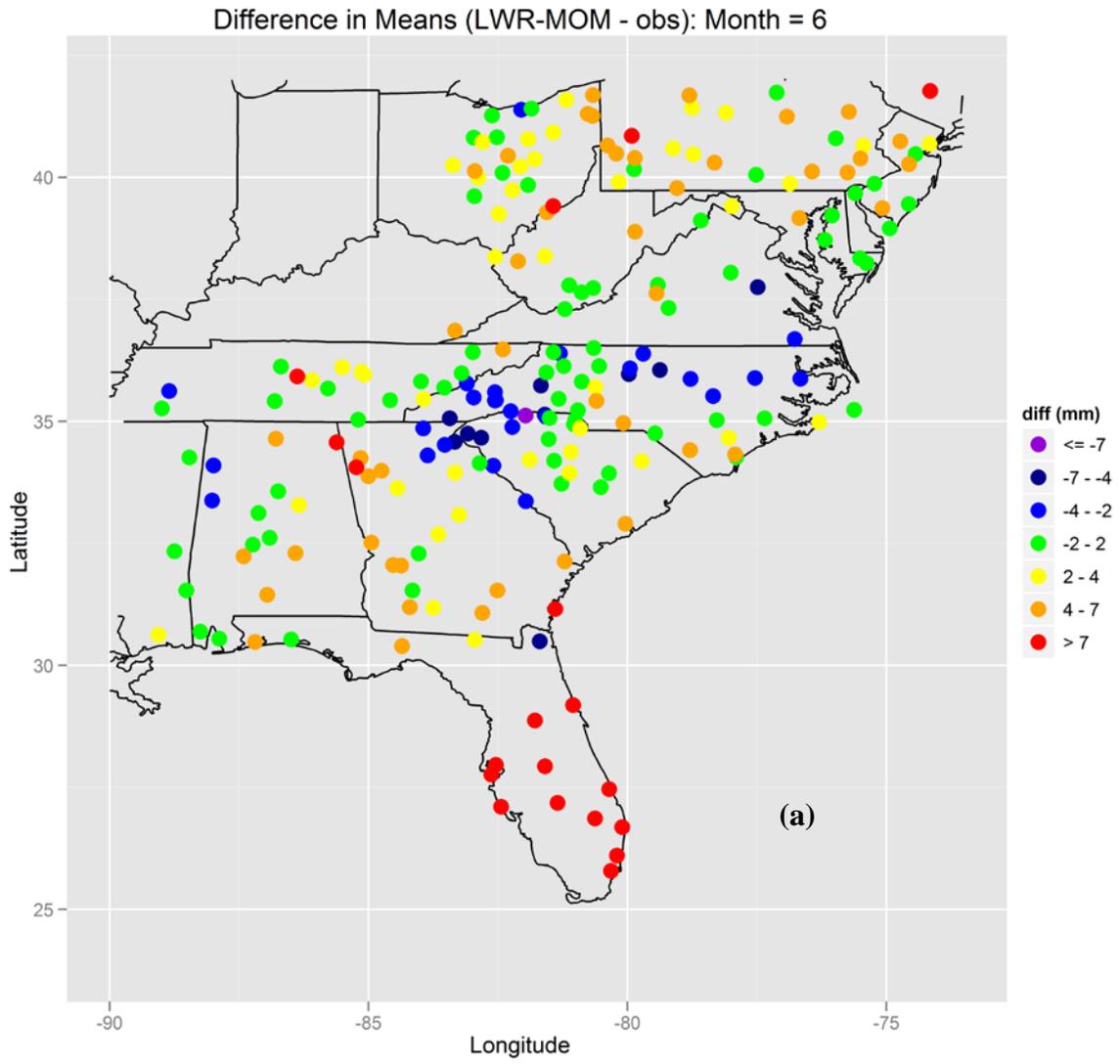
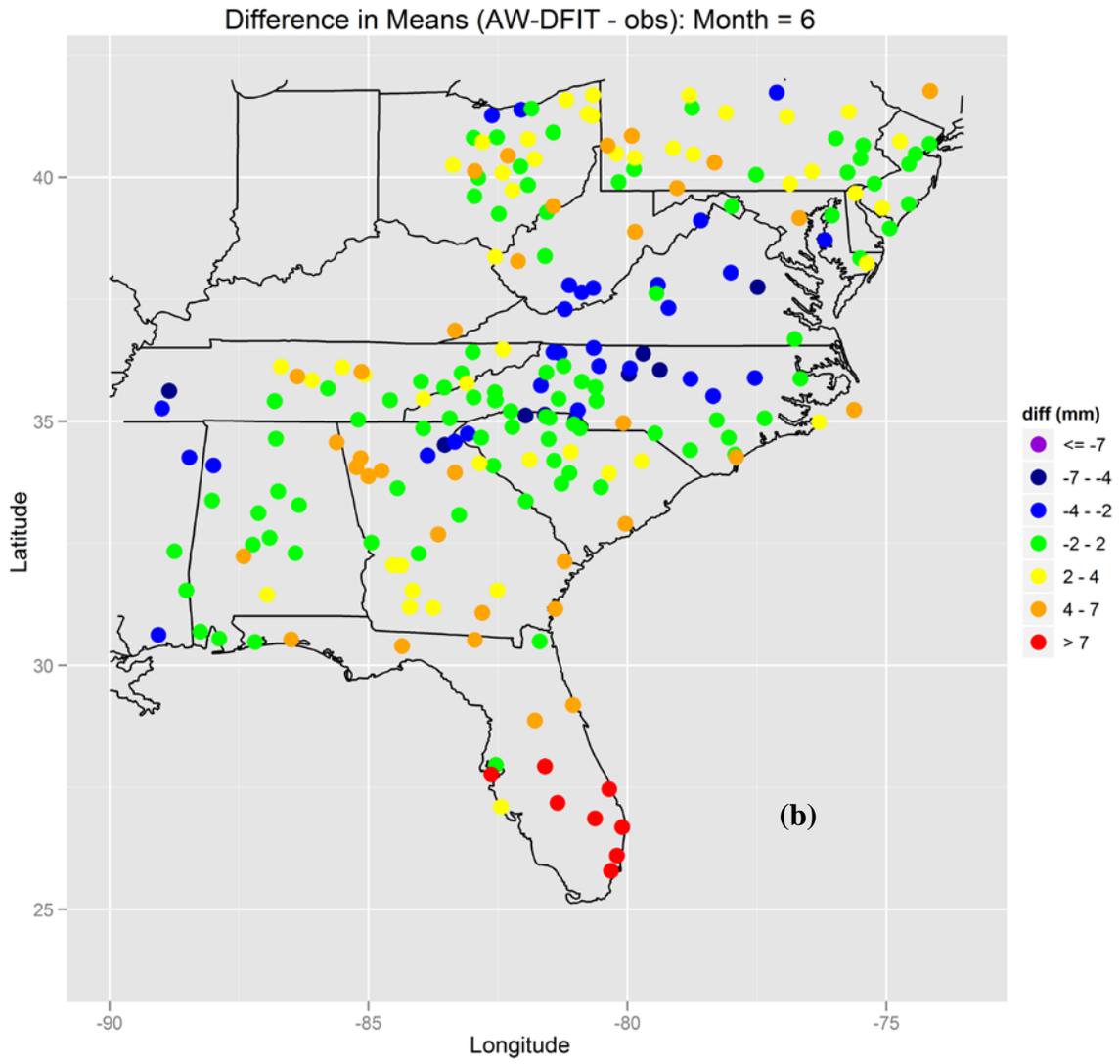
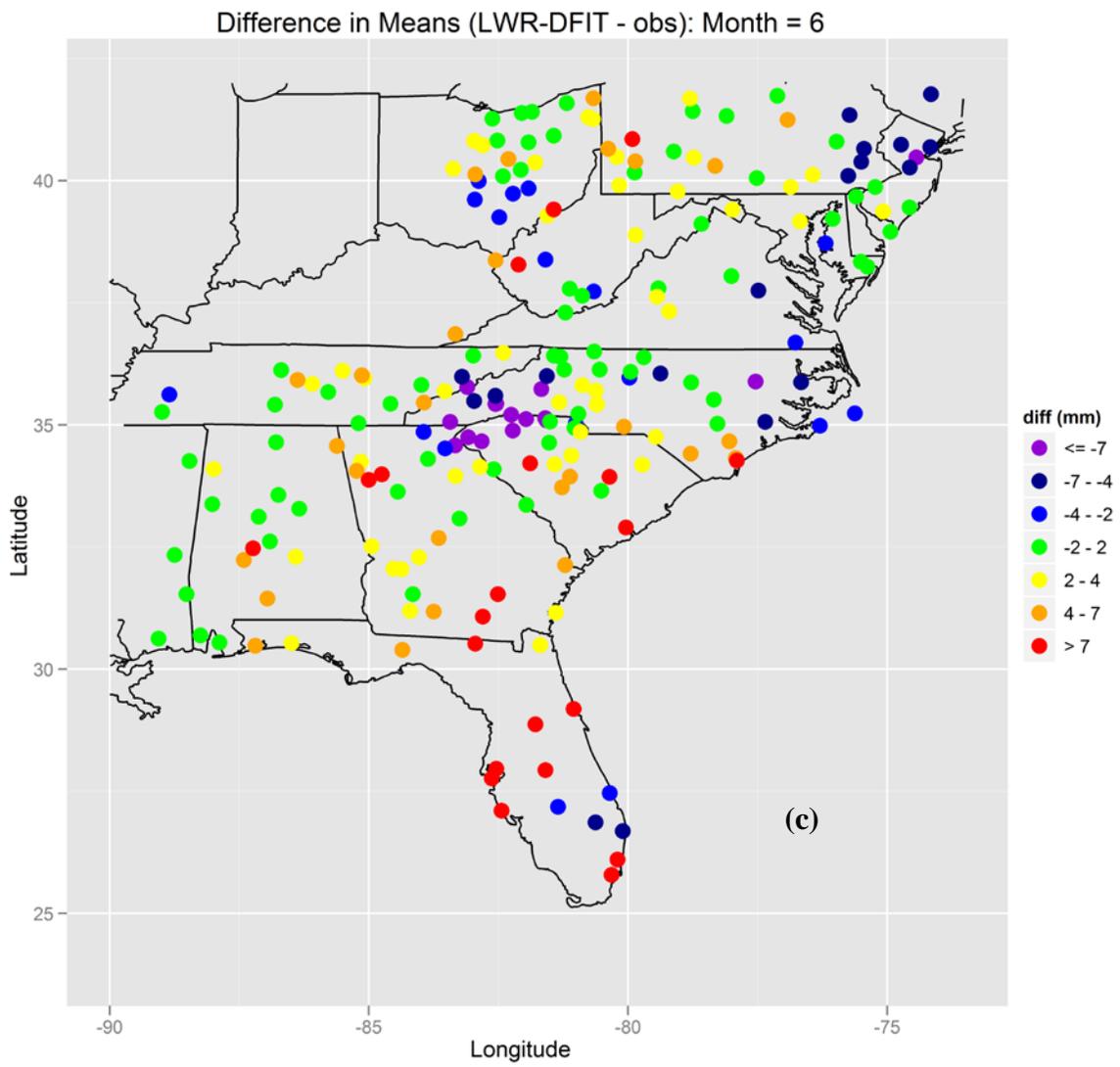


Figure 5.7. Difference between generated and observed values of mean daily nonzero precipitation (μ) for each of the 2001-2009 downscaling simulations: LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for June.







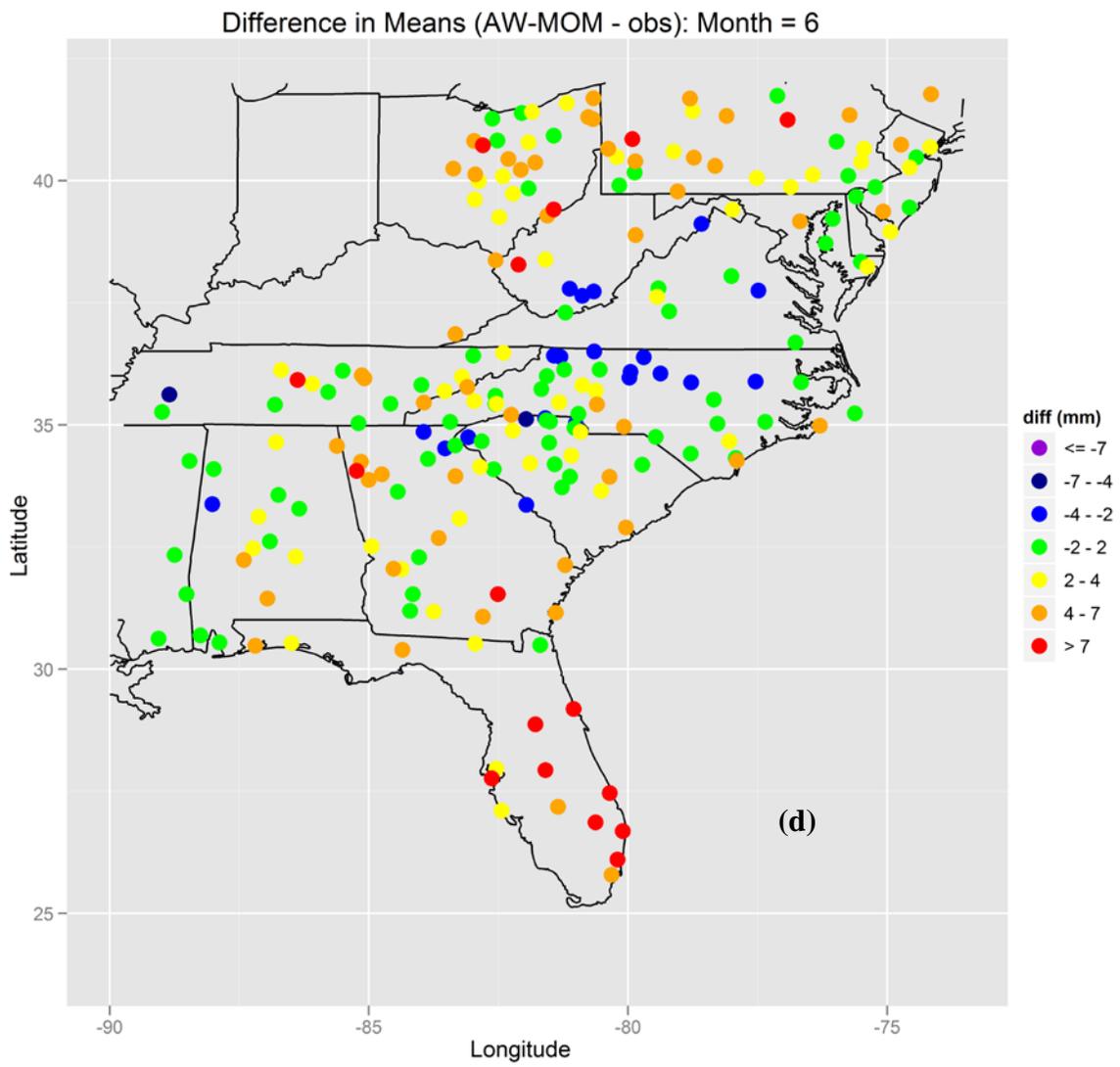
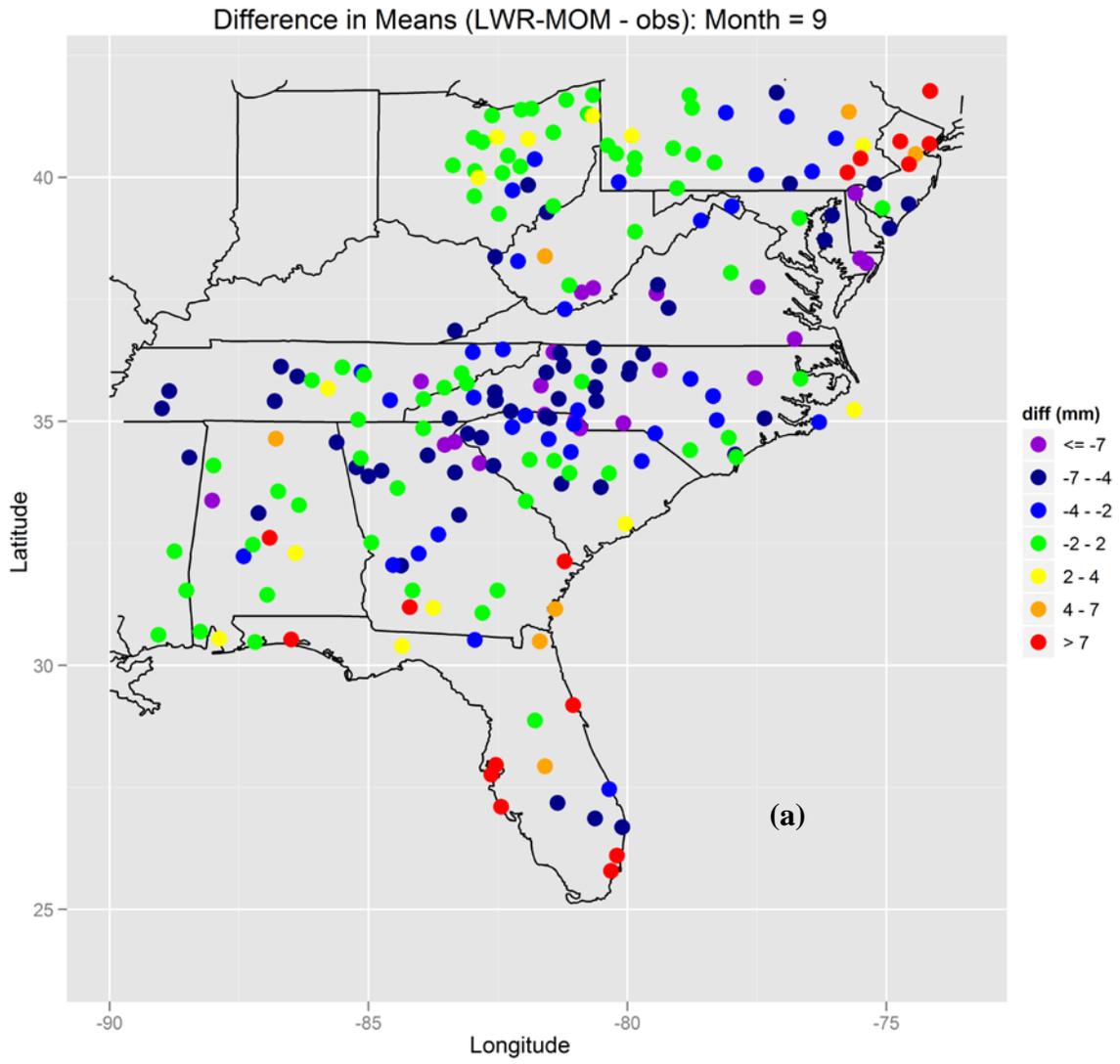
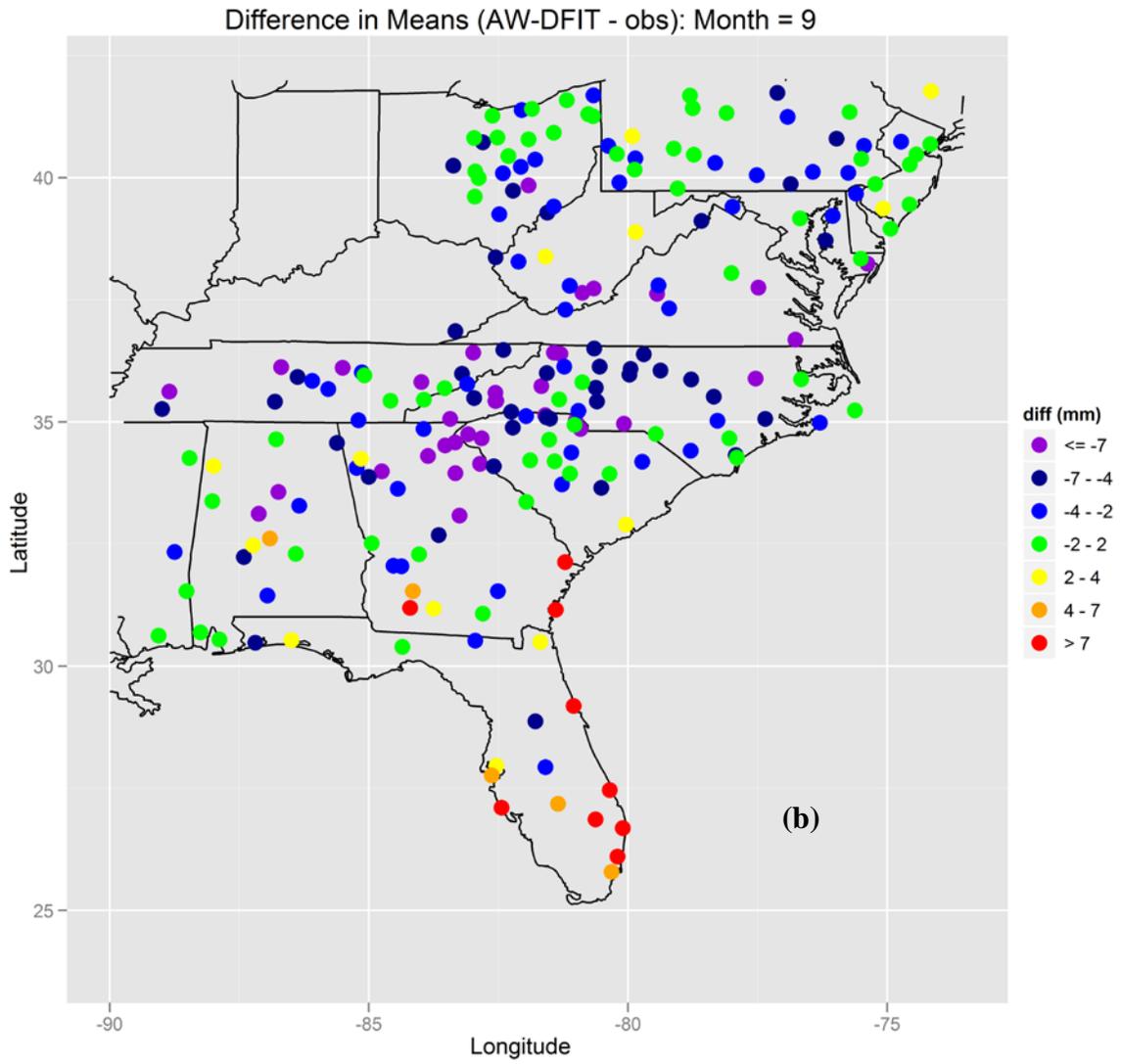
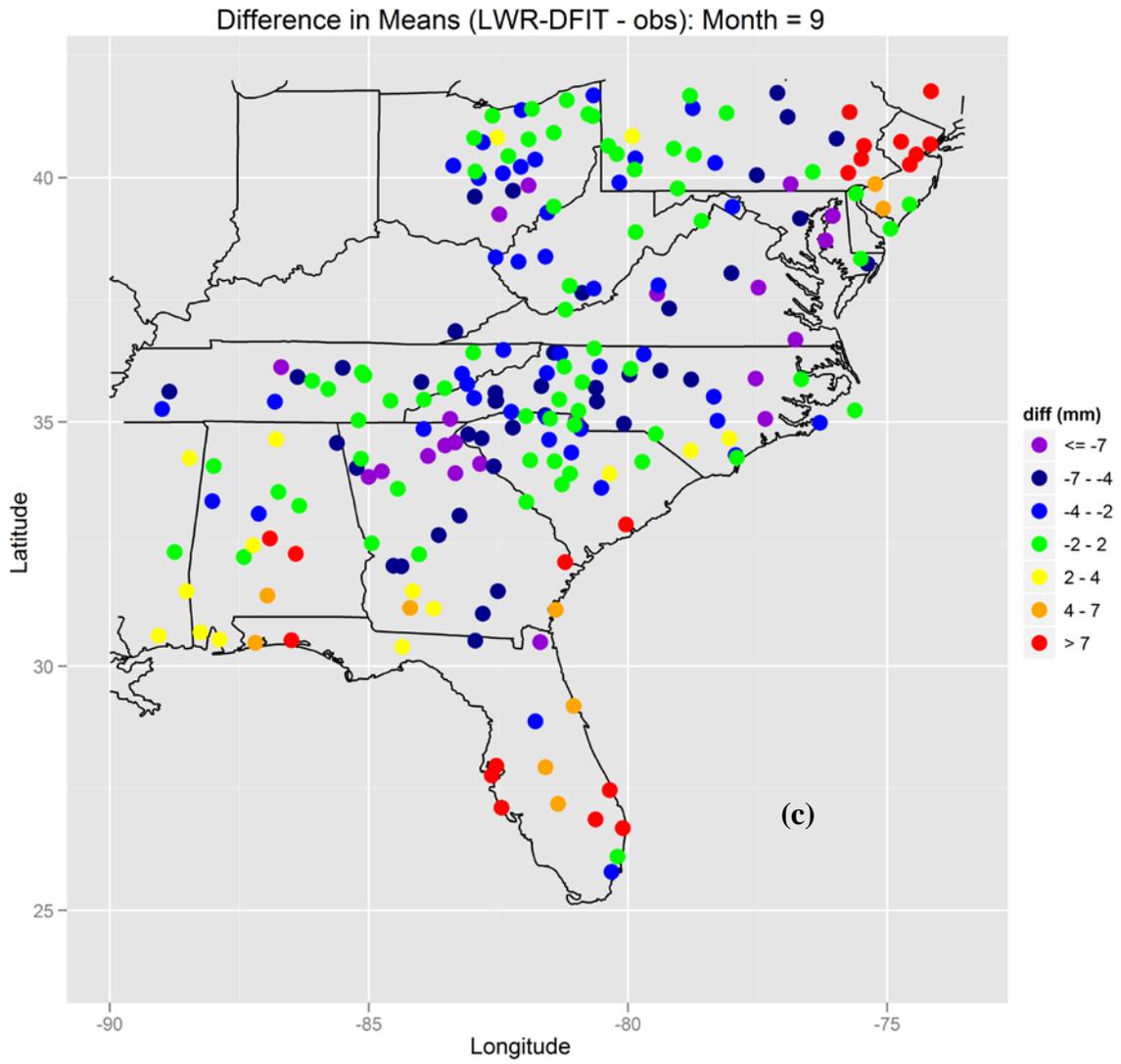
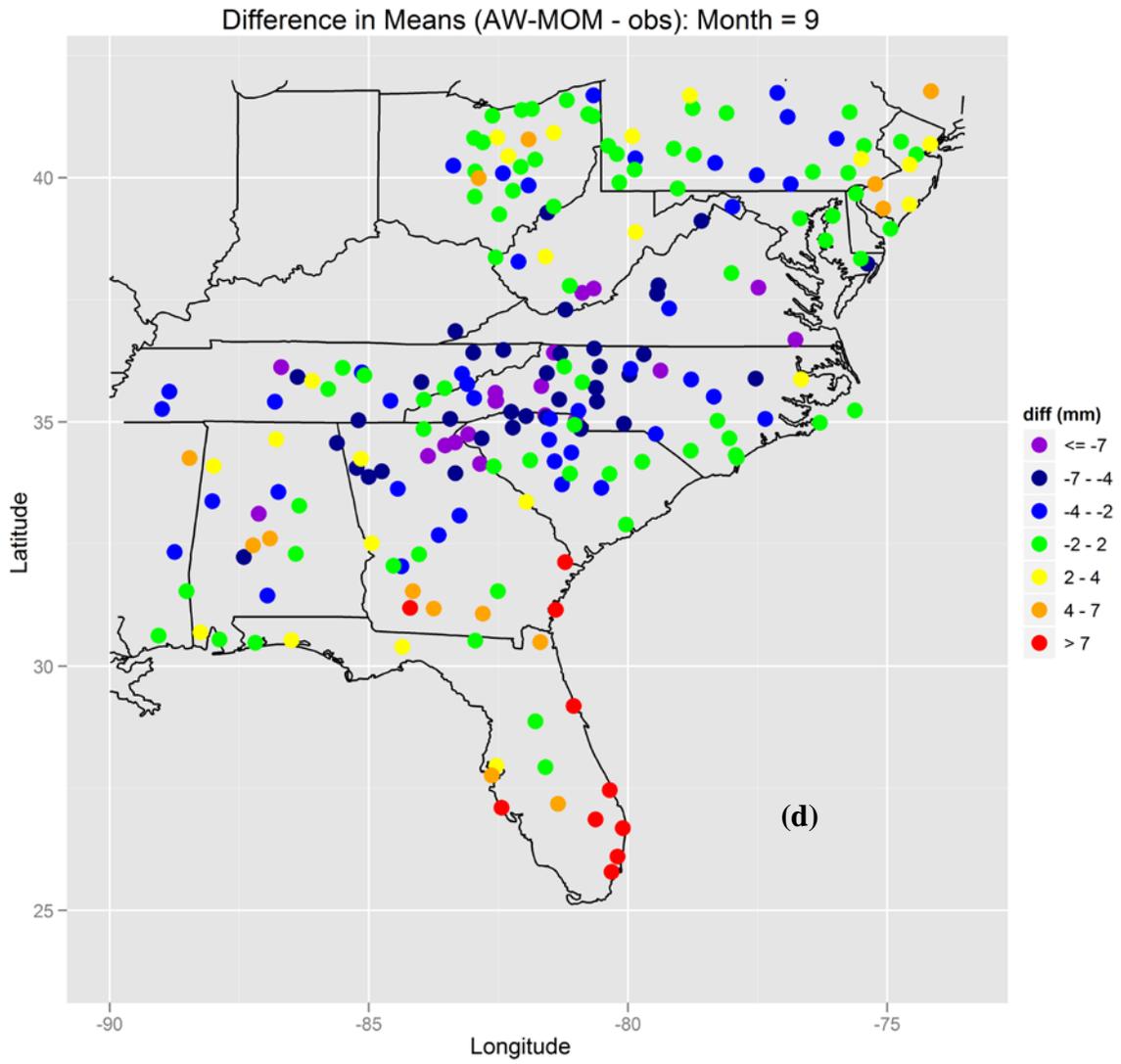


Figure 5.8. Difference between generated and observed values of mean daily nonzero precipitation (μ) for each of the 2001-2009 downscaling simulations: LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for September.









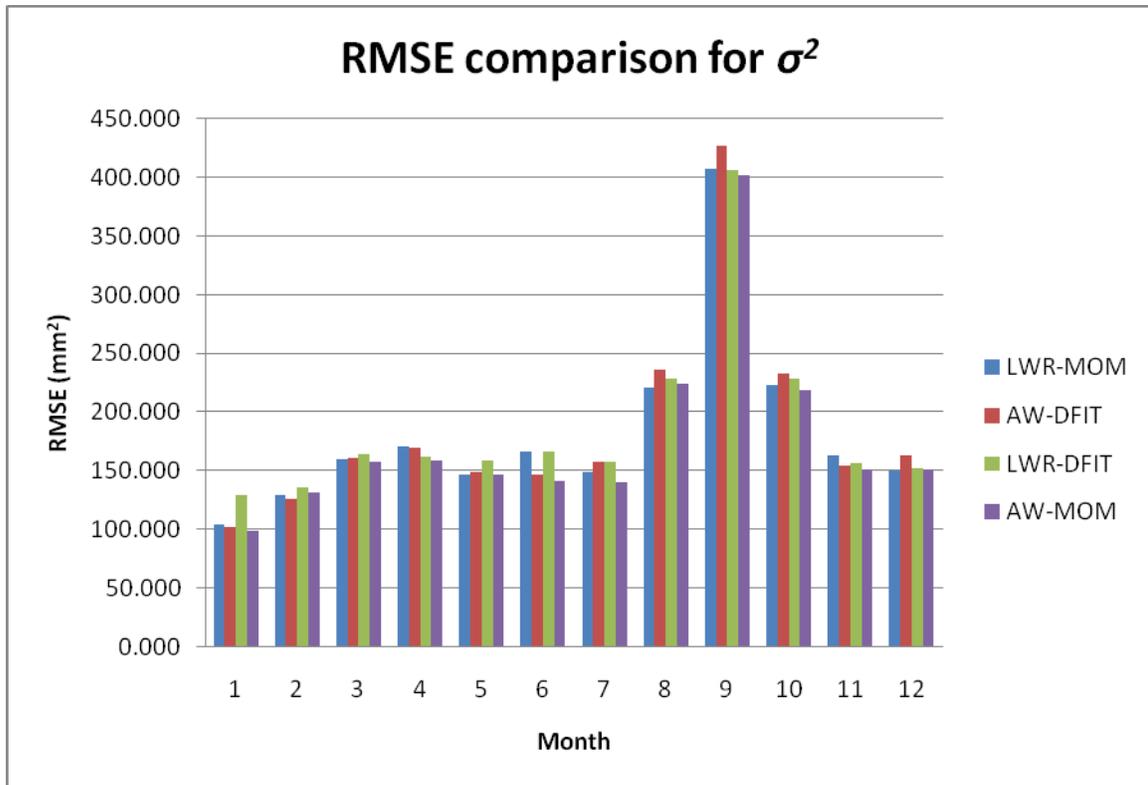
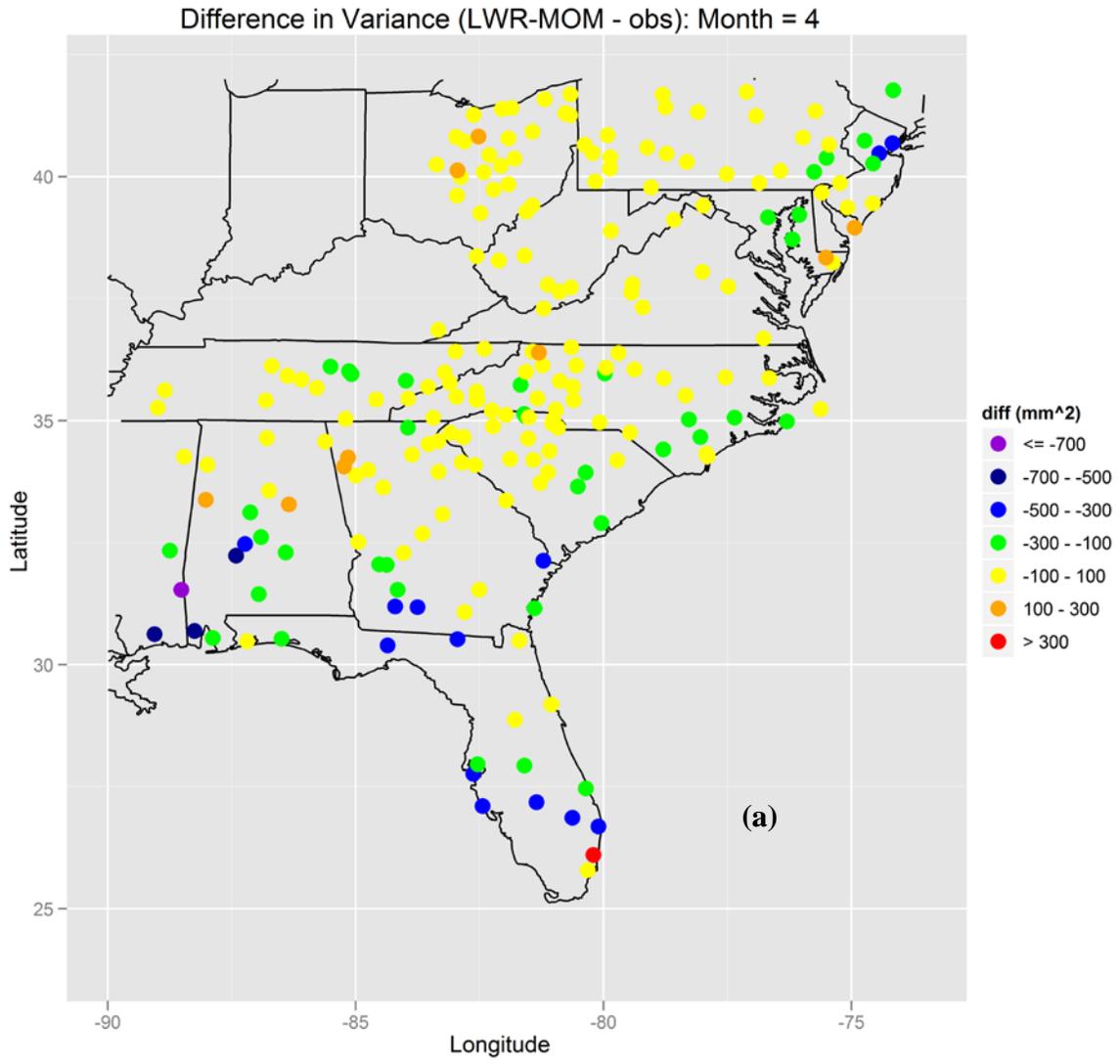
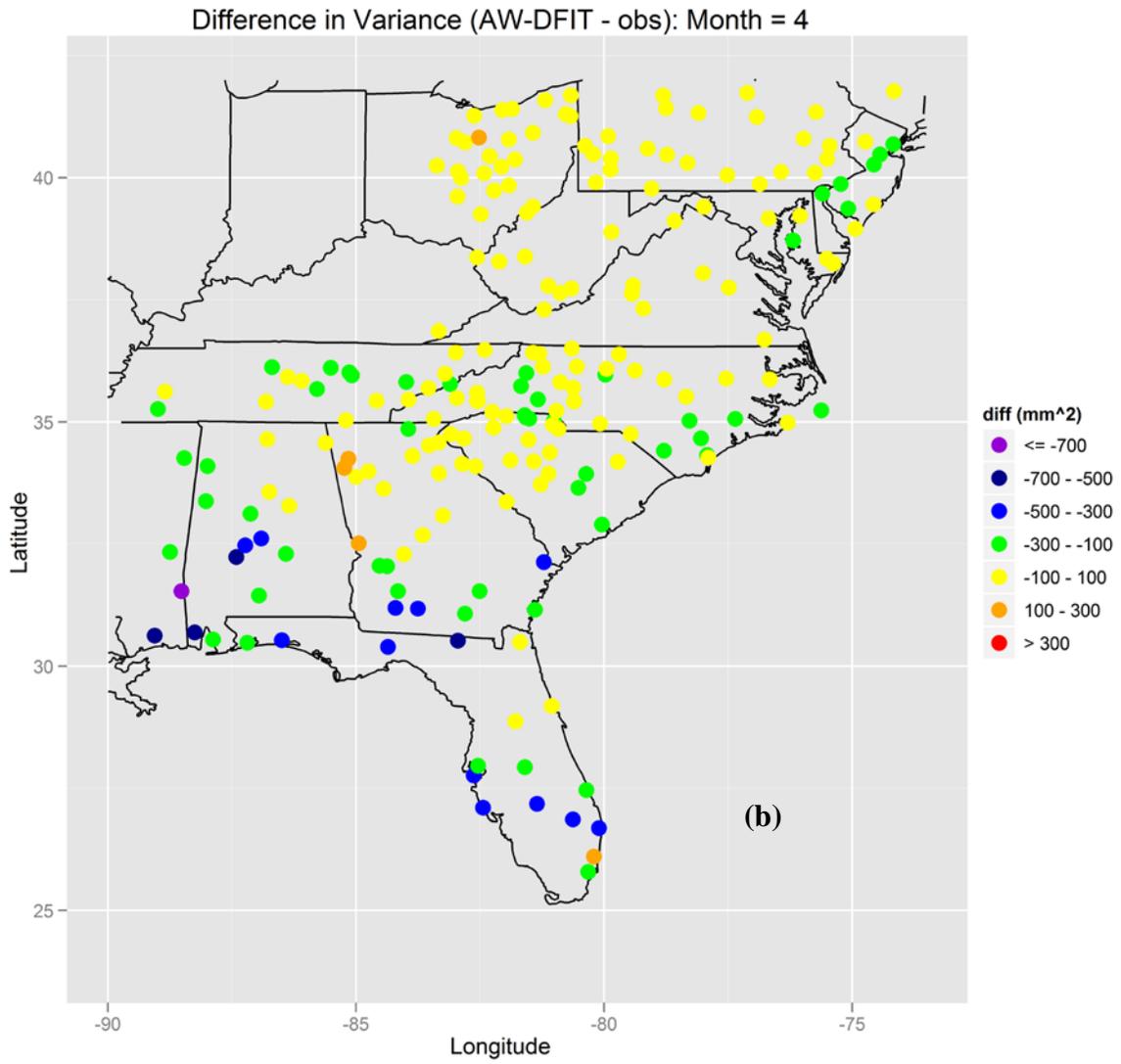
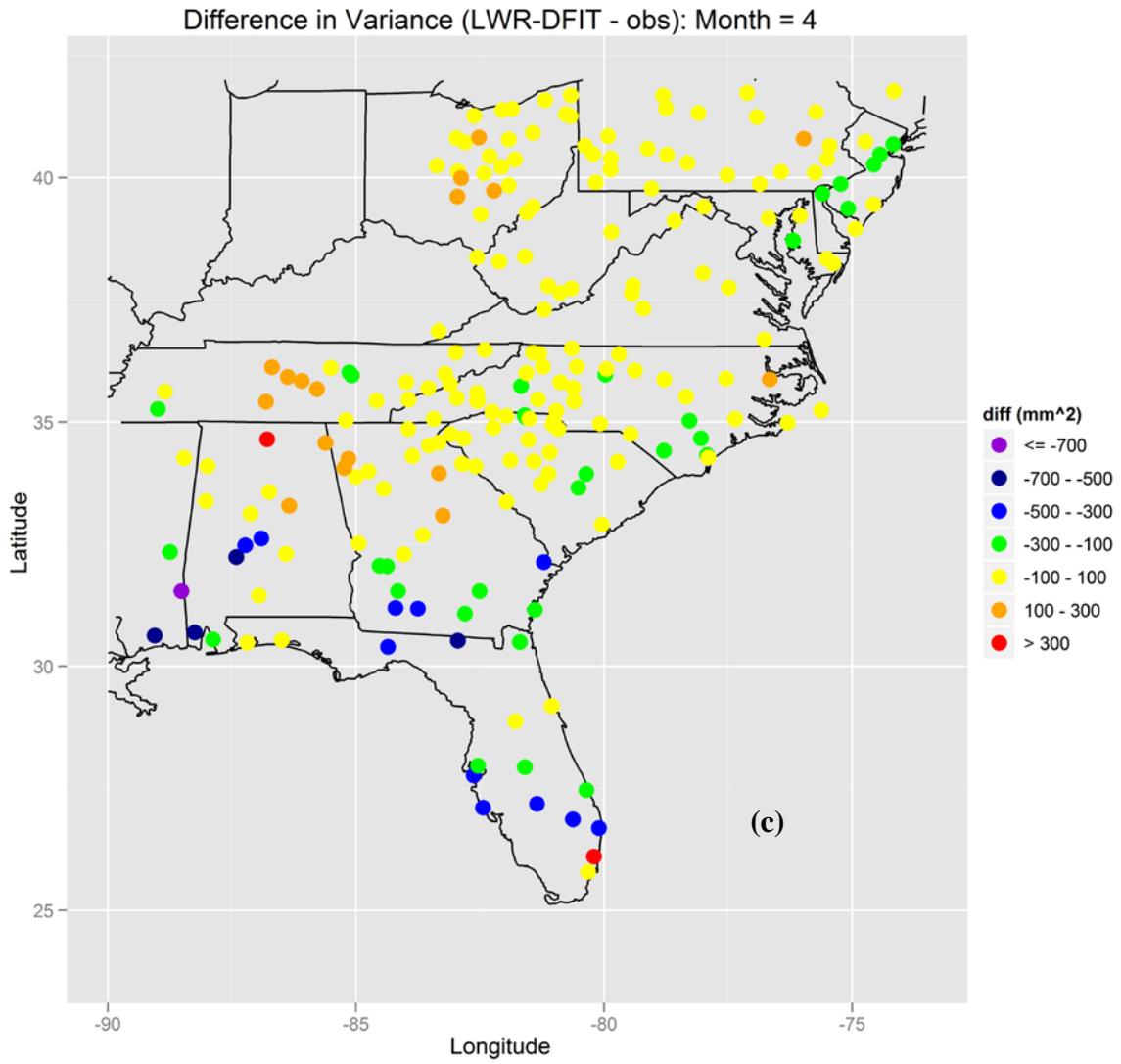


Figure 5.9. RMSE Comparison between 2001-2009 downscaling simulations for the variance of the daily nonzero precipitation (σ^2).

Figure 5.10. Difference between generated and observed values of the variance of the daily nonzero precipitation (σ^2) for each of the 2001-2009 downscaling simulations: LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for April.







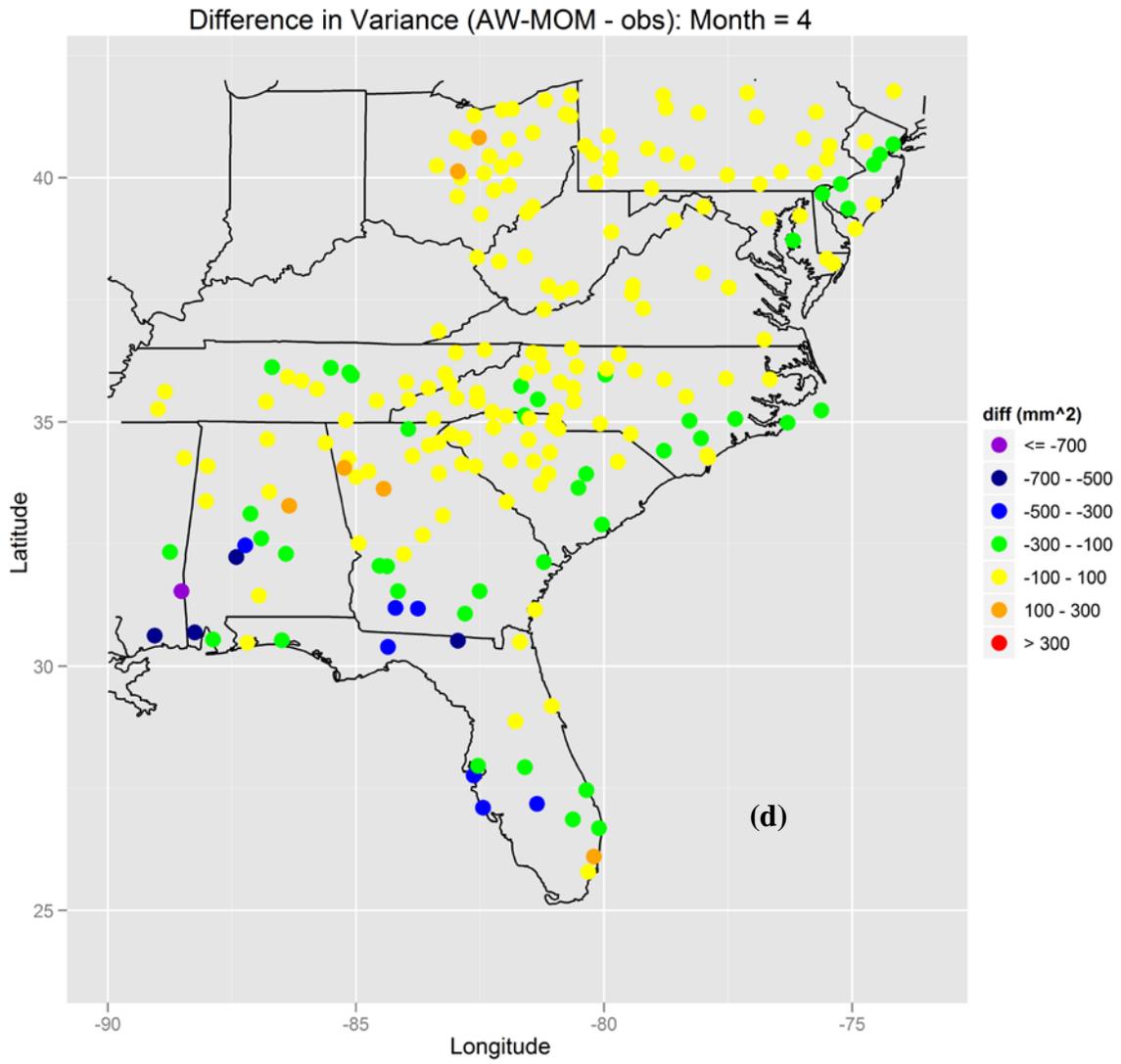
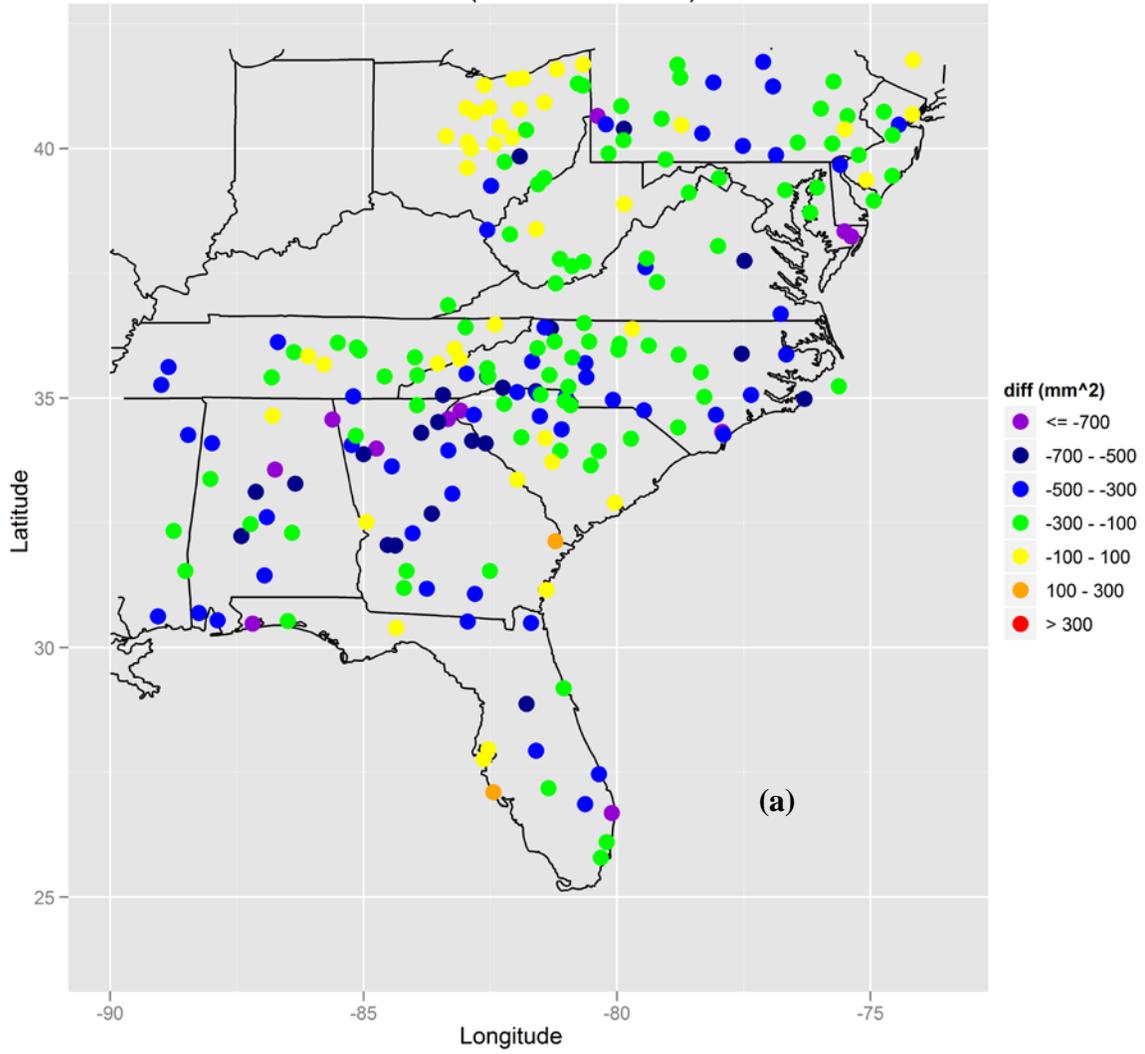
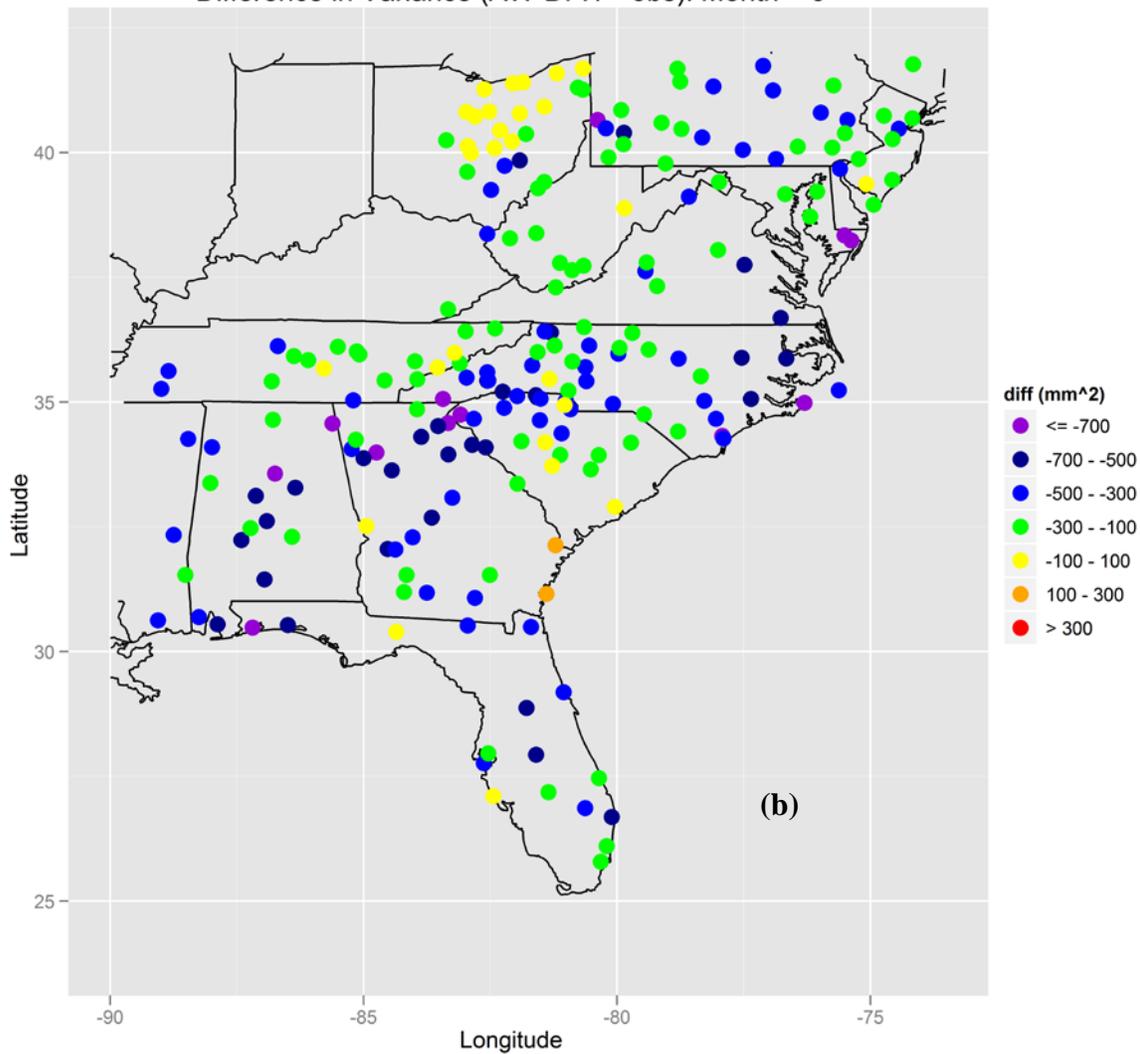


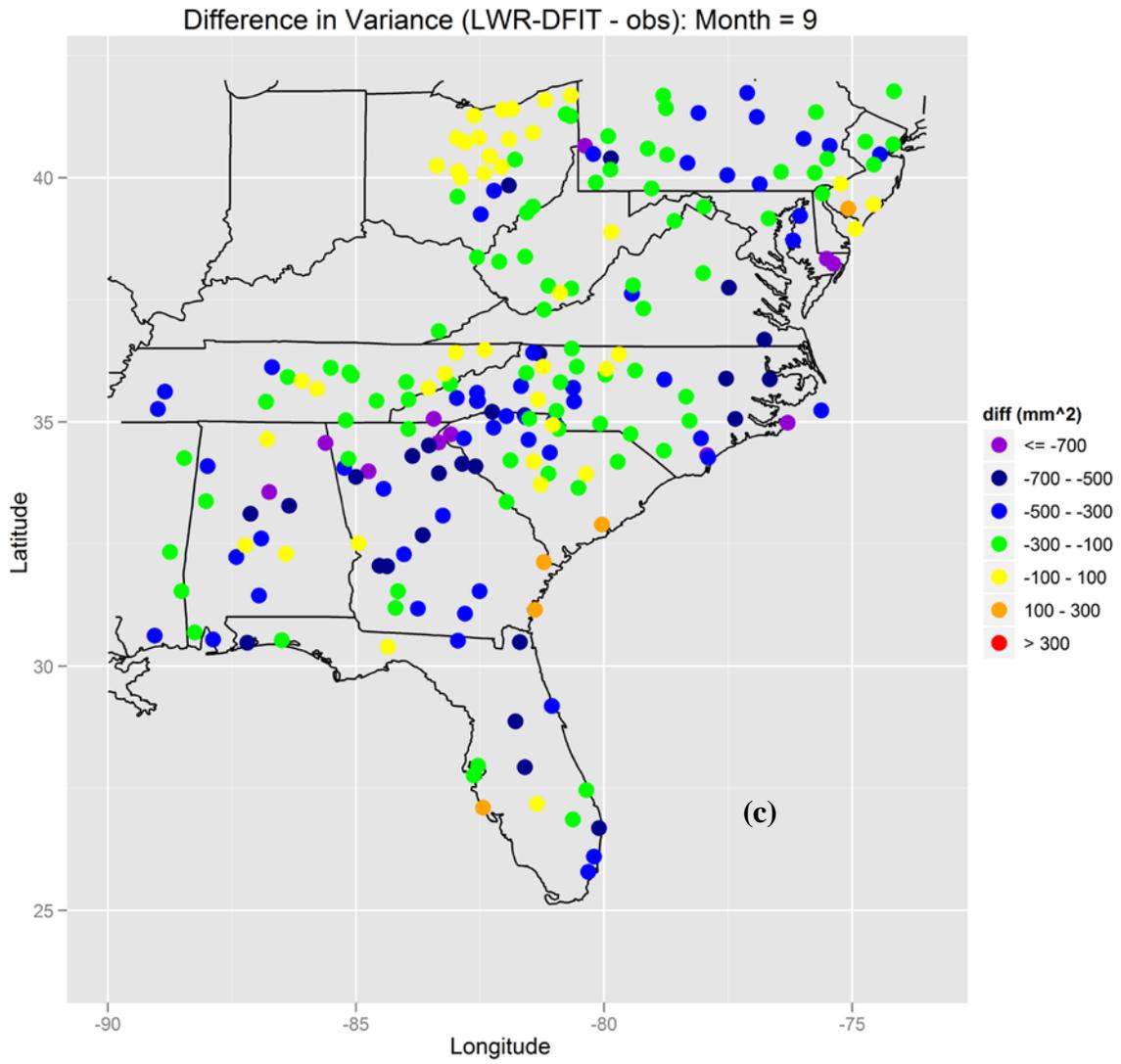
Figure 5.11. Difference between generated and observed values of the variance of the daily nonzero precipitation (σ^2) for each of the 2001-2009 downscaling simulations: LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for September.

Difference in Variance (LWR-MOM - obs): Month = 9



Difference in Variance (AW-DFIT - obs): Month = 9





Difference in Variance (AW-MOM - obs): Month = 9

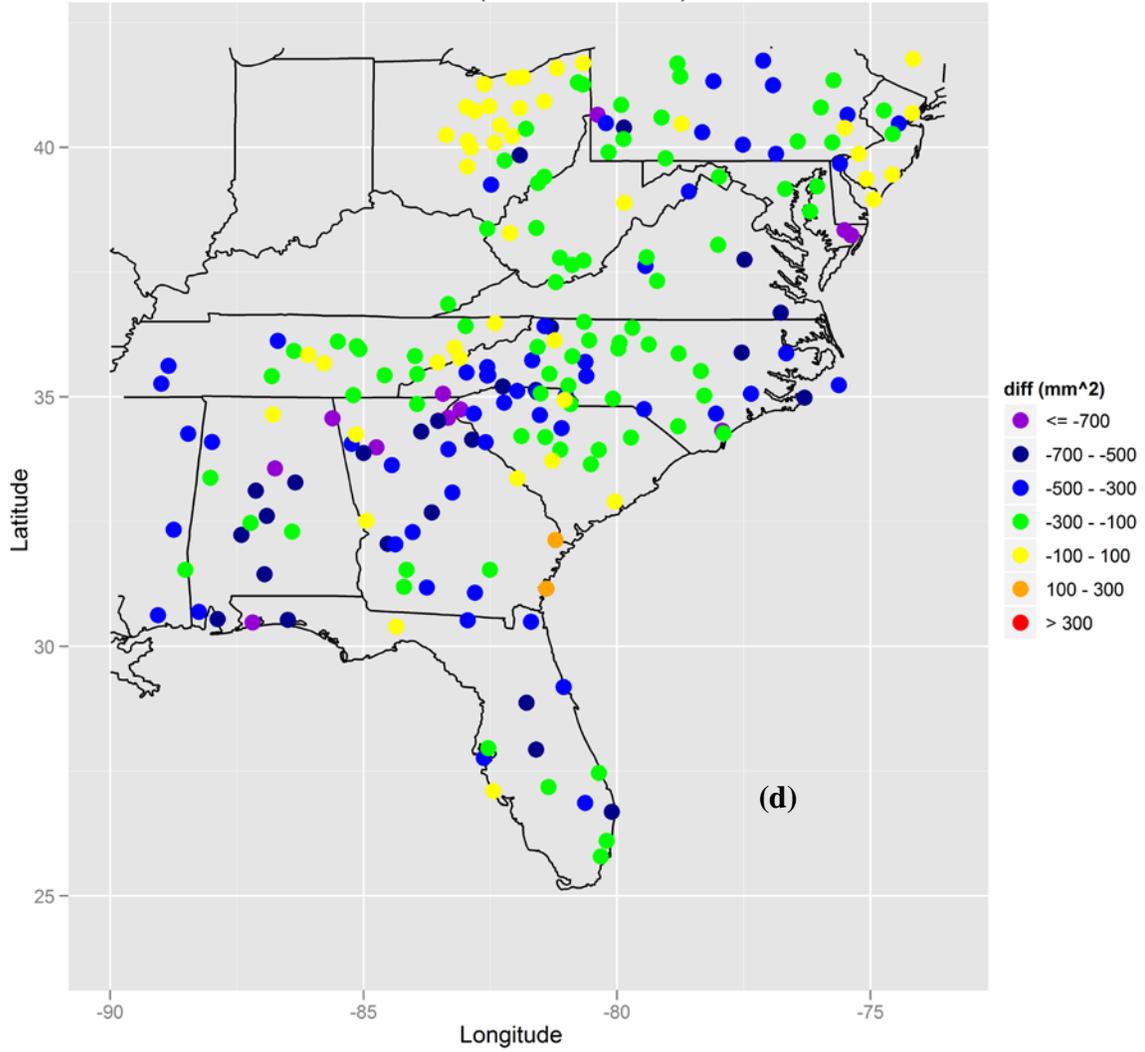
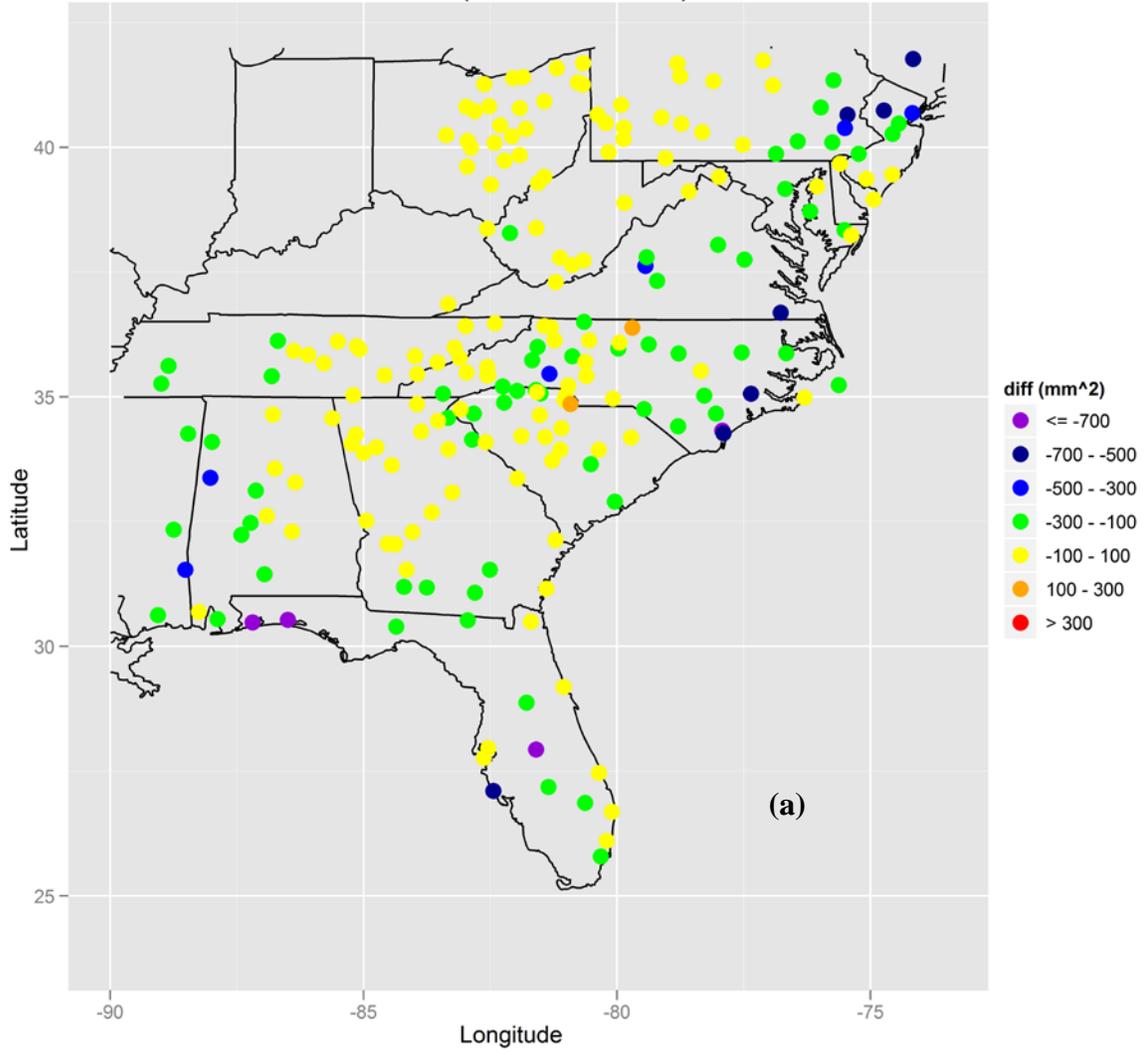
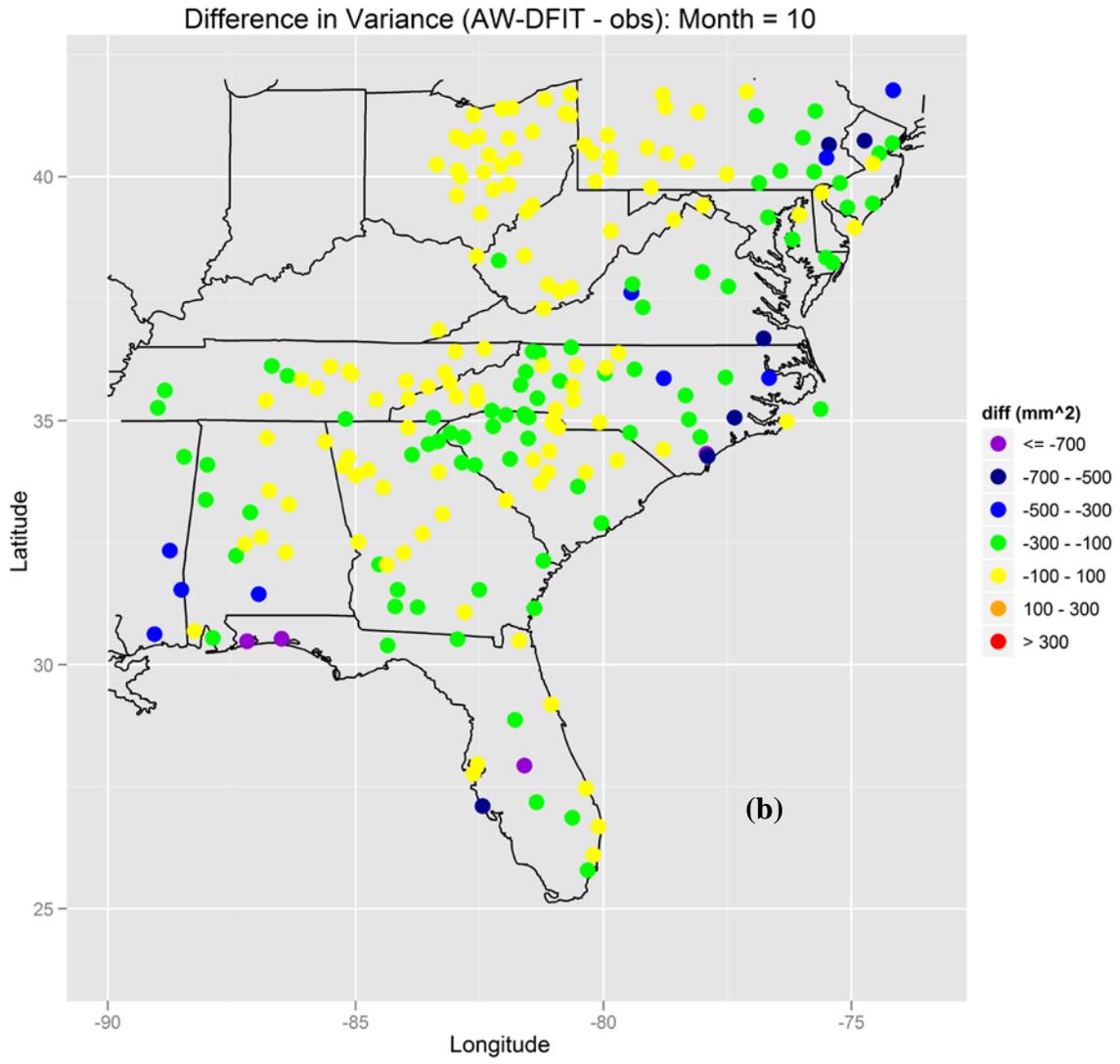
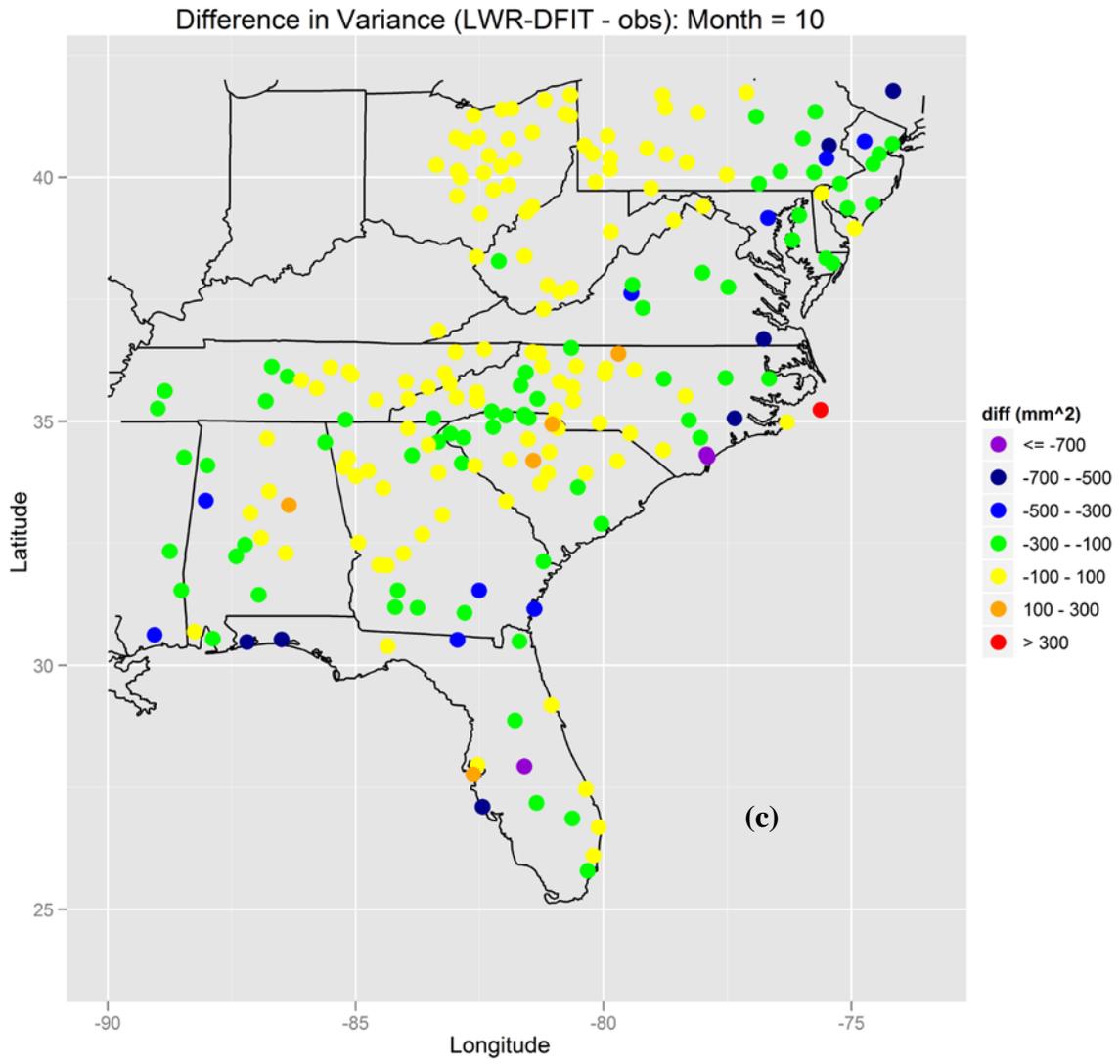


Figure 5.12. Difference between generated and observed values of the variance of the daily nonzero precipitation (σ^2) for each of the 2001-2009 downscaling simulations: LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d) for October.

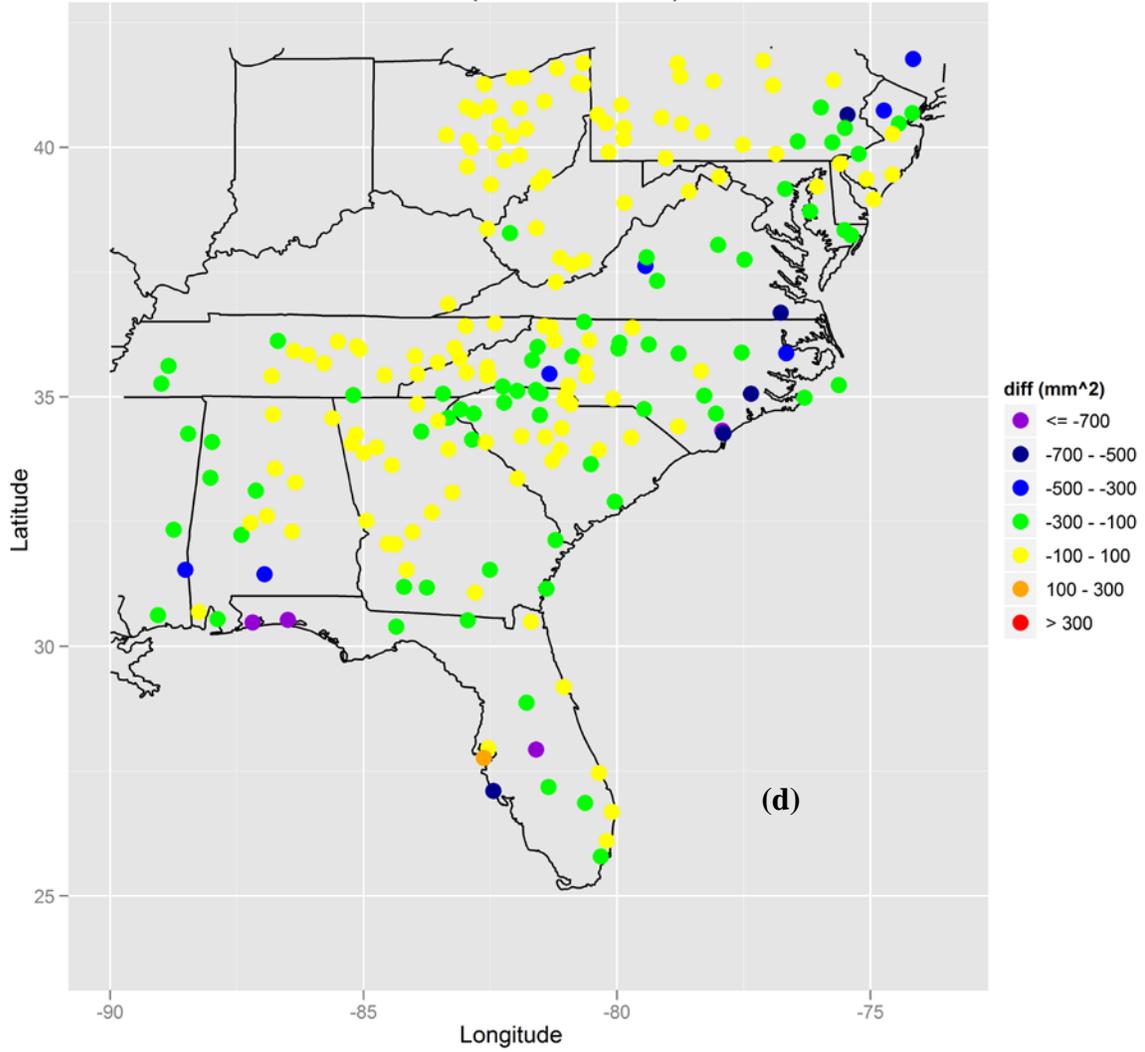
Difference in Variance (LWR-MOM - obs): Month = 10







Difference in Variance (AW-MOM - obs): Month = 10



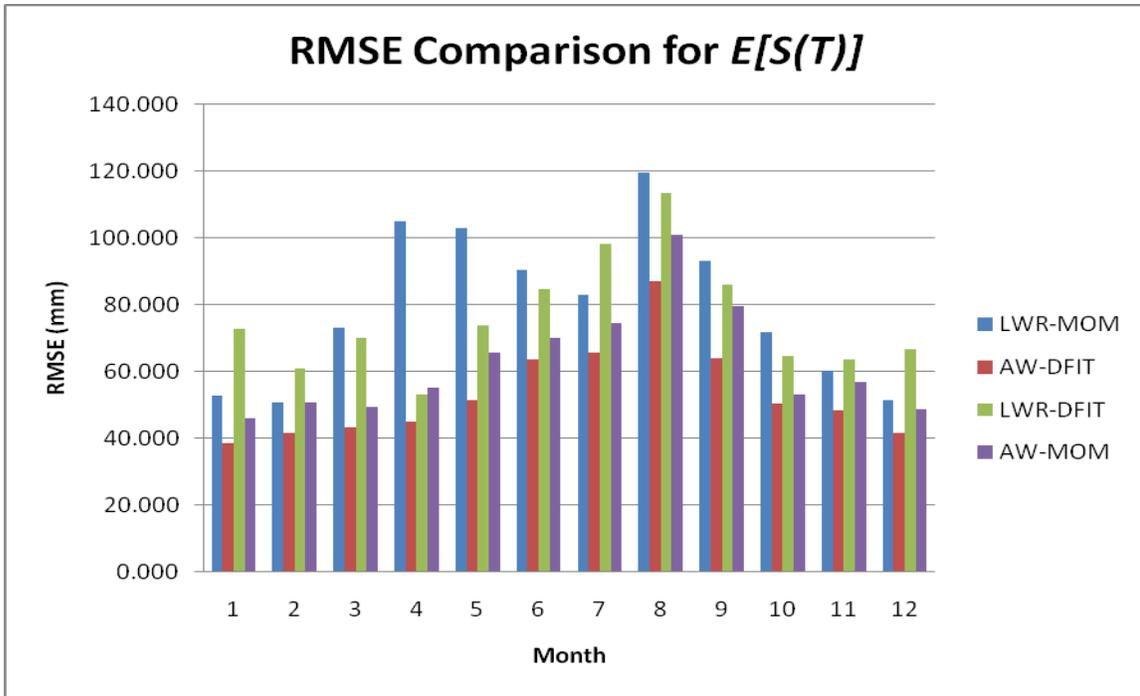


Figure 5.13. RMSE Comparison between 2001-2009 downscaling simulations for the average total precipitation ($E[S(T)]$).

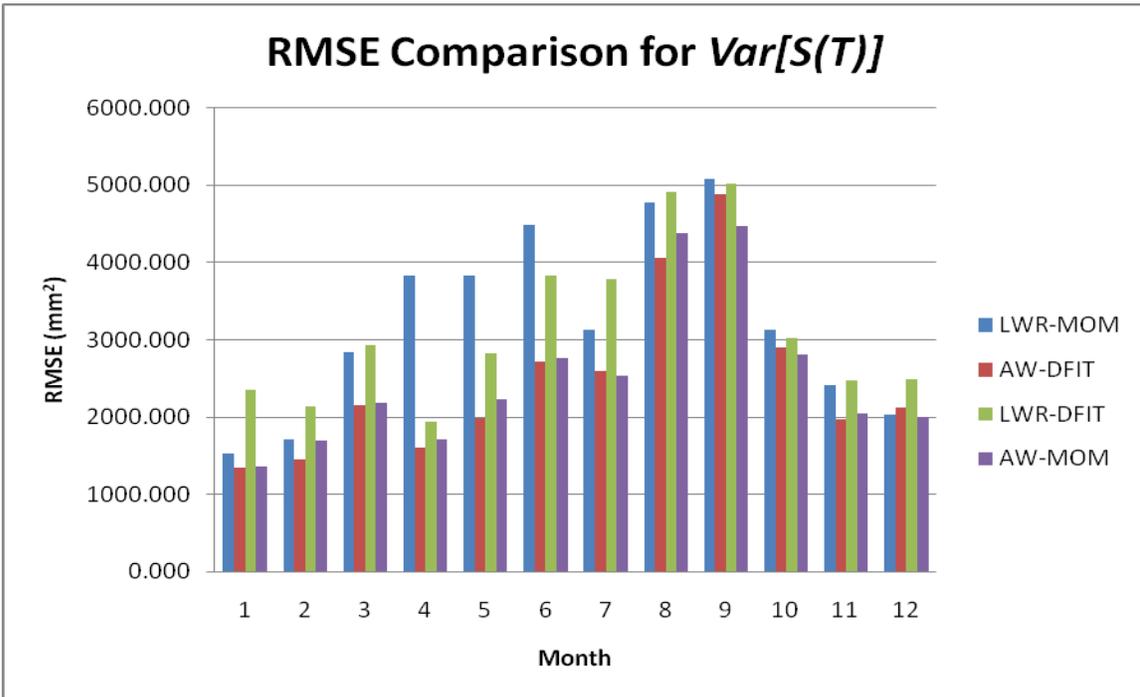
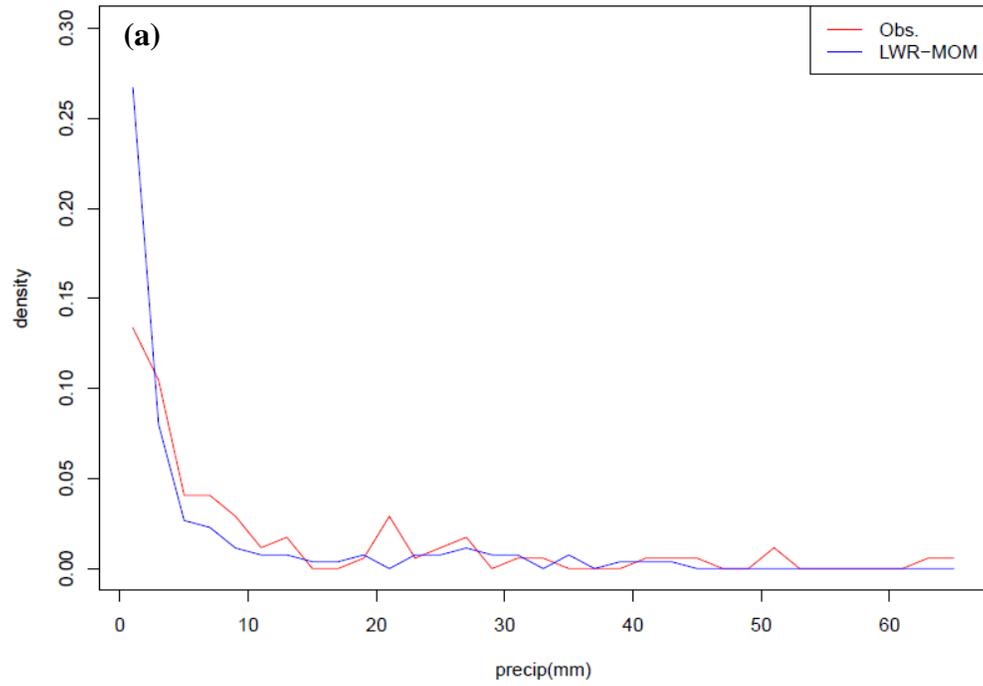


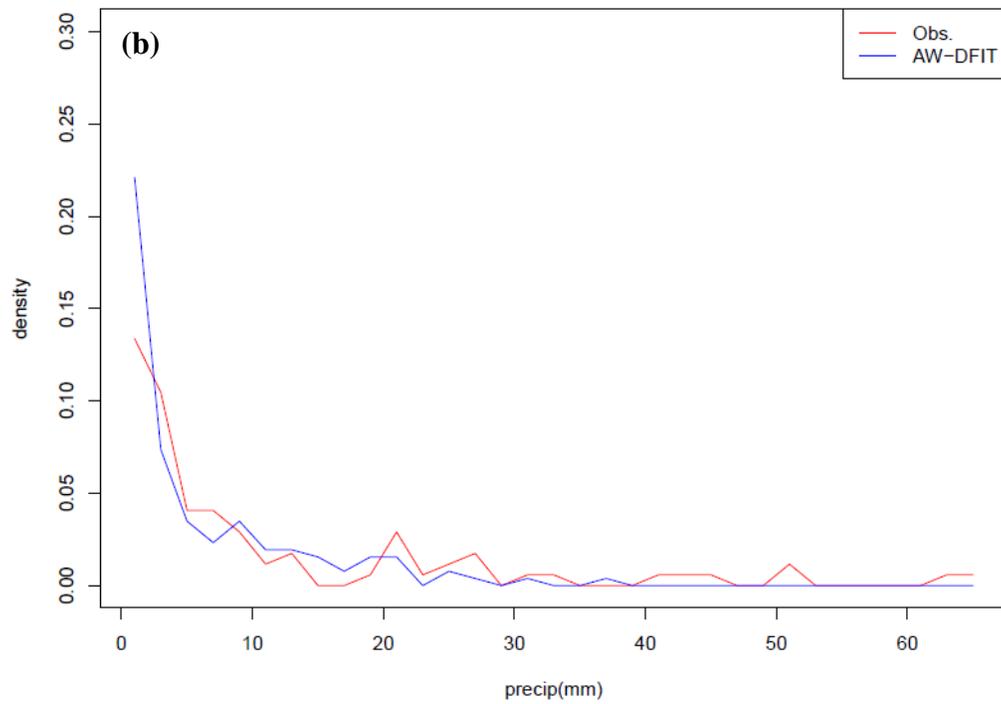
Figure 5.14. RMSE Comparison between 2001-2009 downscaling simulations for the inter-annual variability ($Var[S(T)]$).

Figure 5.15. Observed vs. Generated PDFs of nonzero precipitation for station 090969 in Blairsville, GA for January for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).

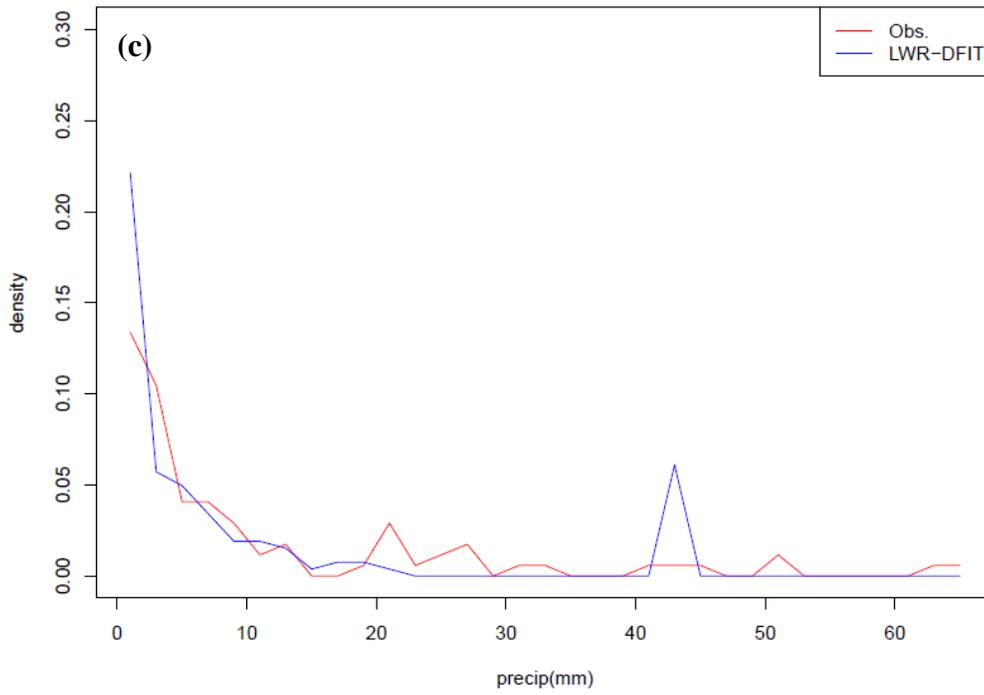
Comparison of Nonzero Rainfall PDF
Station: 090969 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 090969 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 090969 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 090969 Month: 1

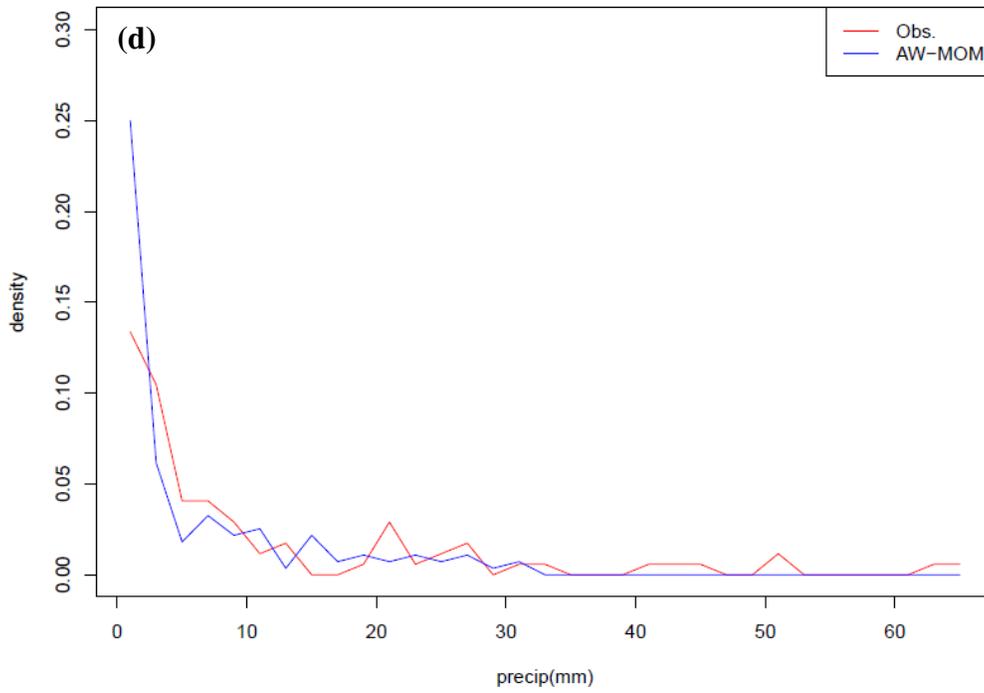
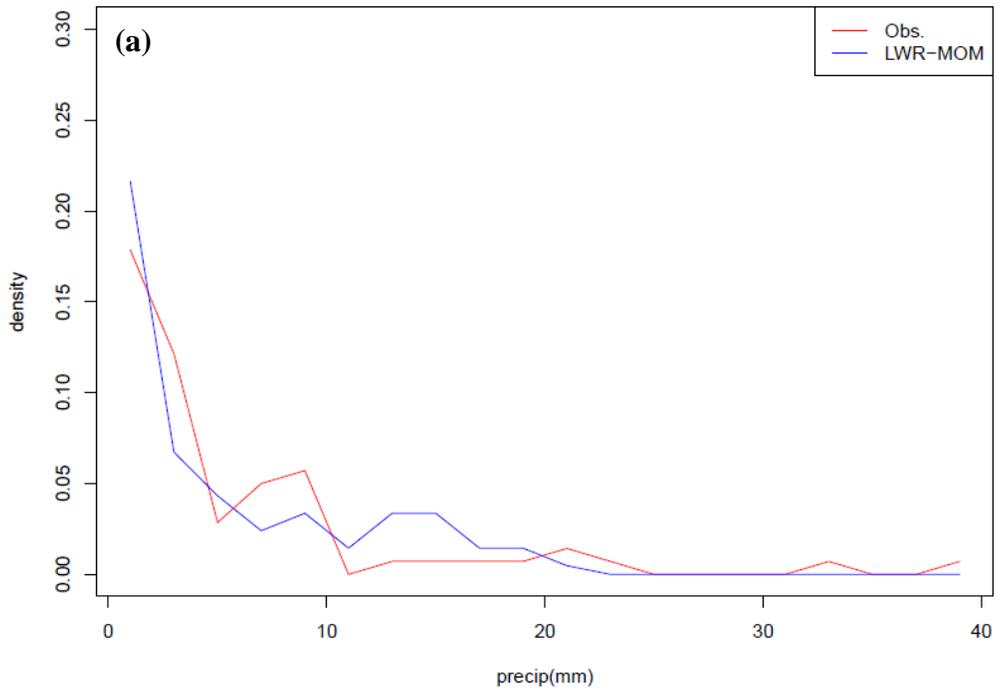
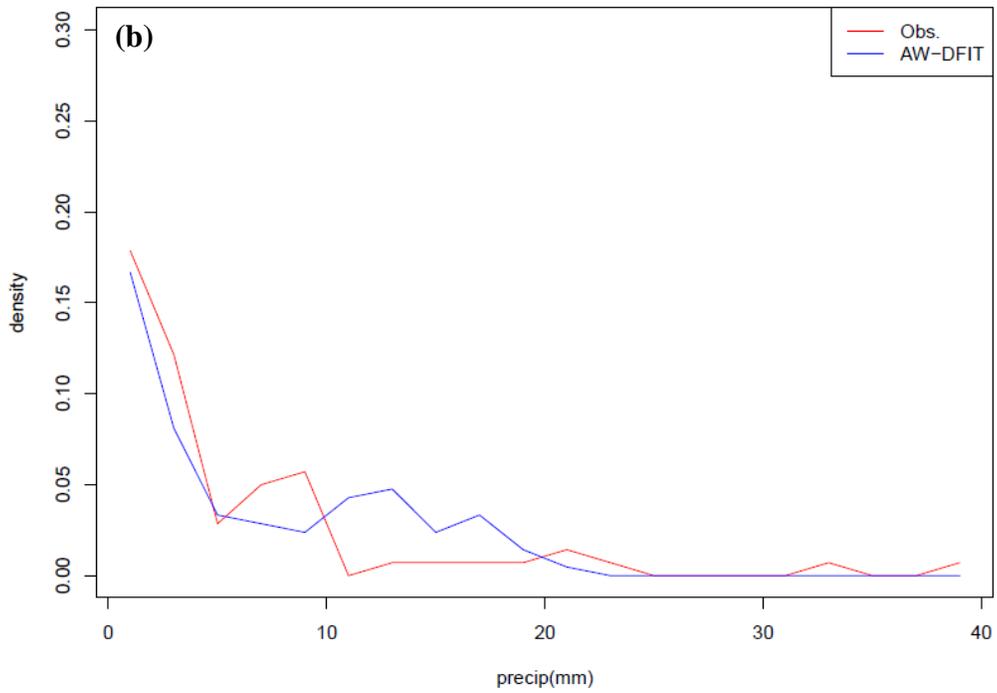


Figure 5.16. Observed vs. Generated PDFs of nonzero precipitation for station 319457 in Wilmington, NC for January for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).

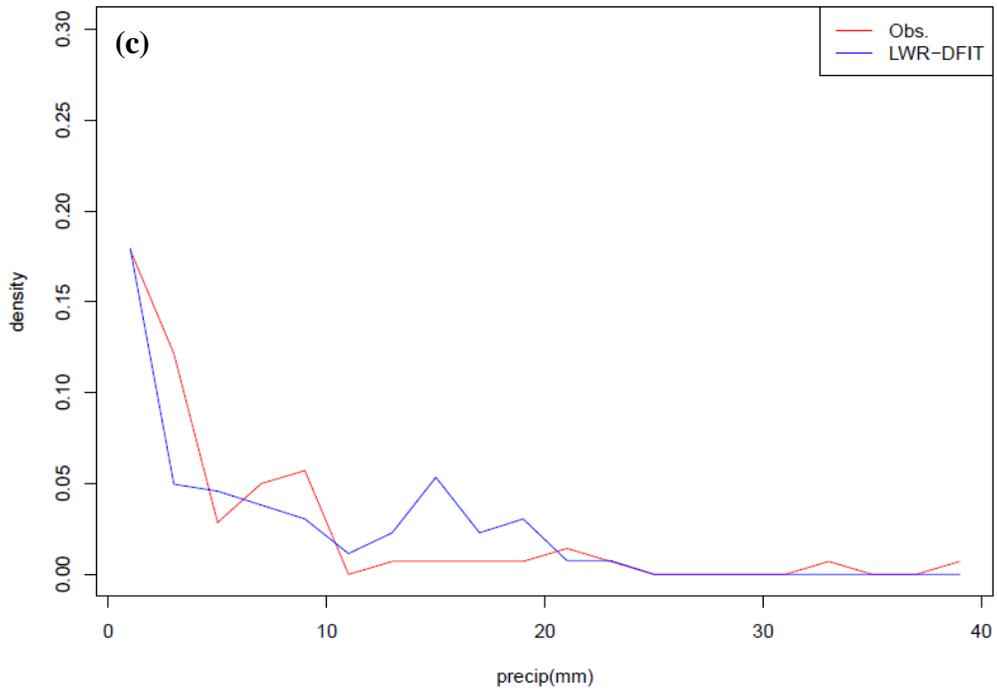
Comparison of Nonzero Rainfall PDF
Station: 319457 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 319457 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 319457 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 319457 Month: 1

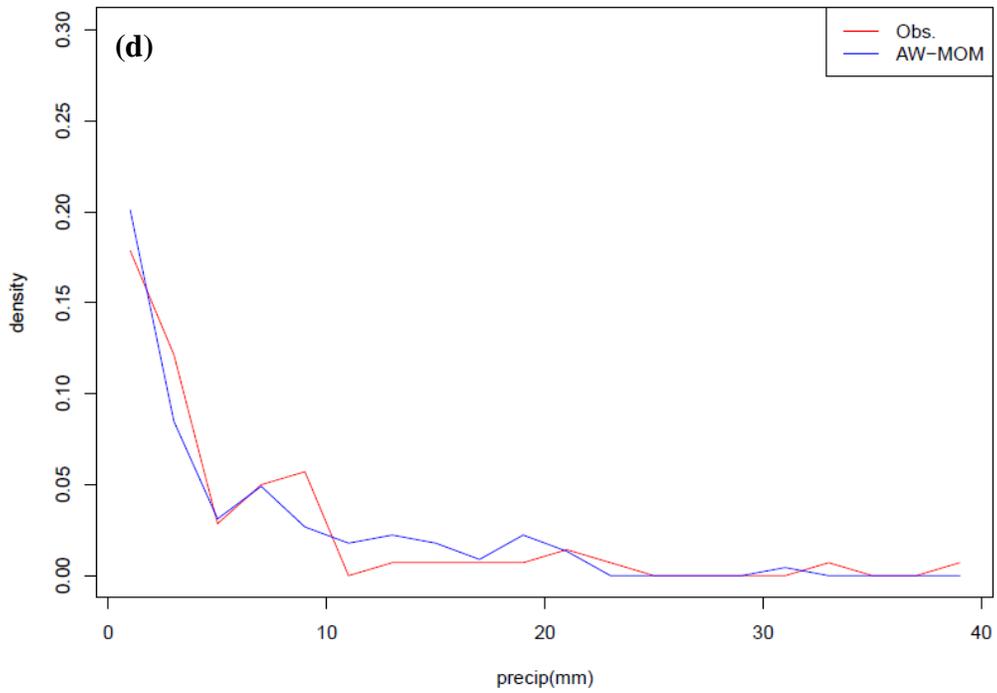
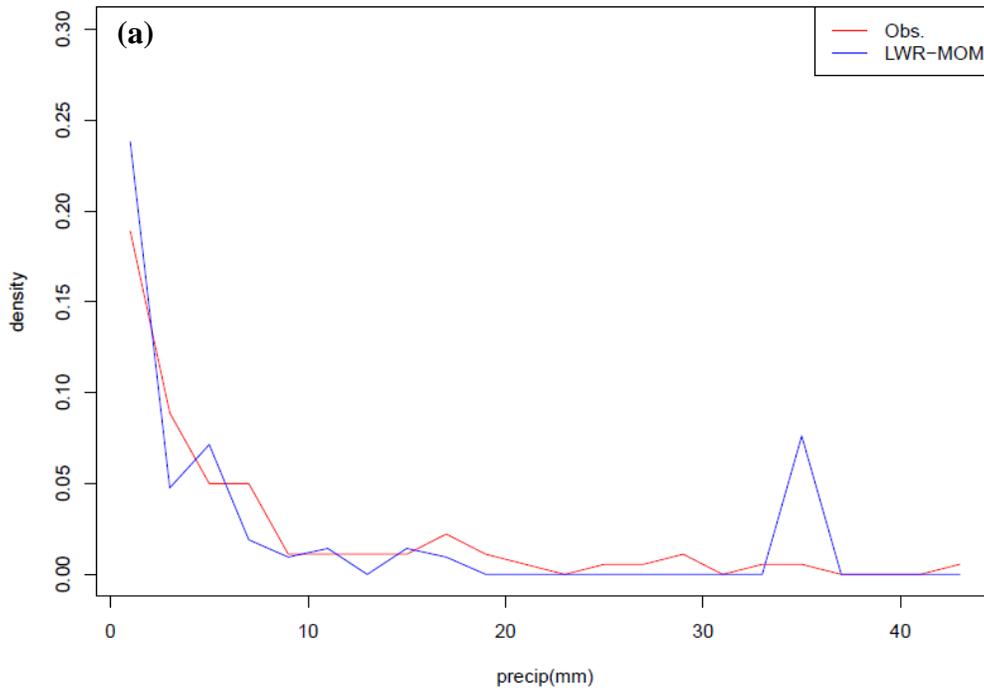
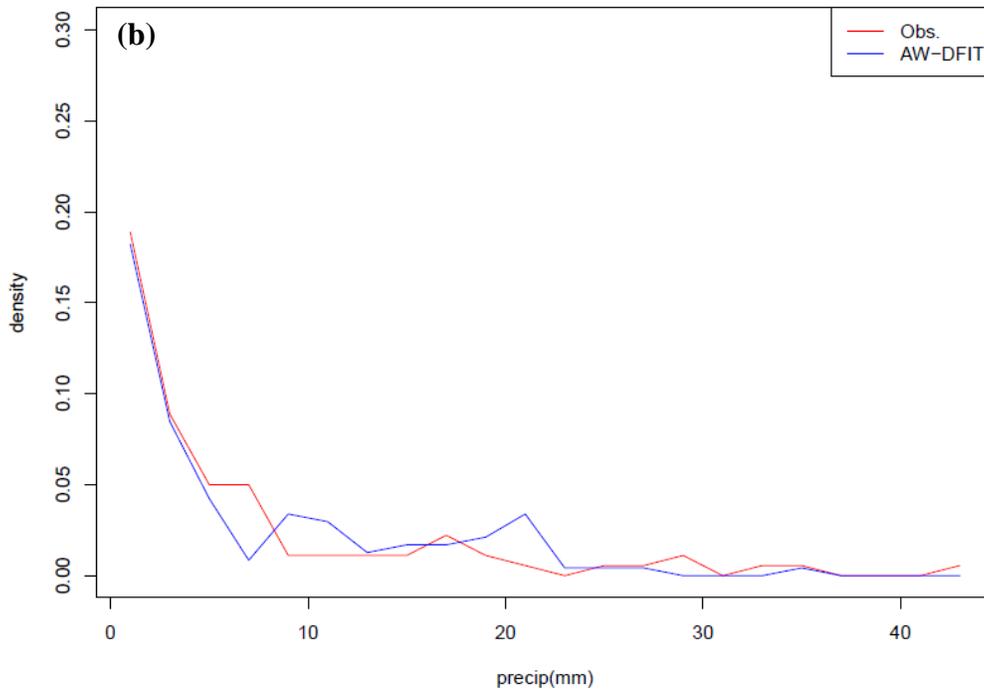


Figure 5.17. Observed vs. Generated PDFs of nonzero precipitation for station 286026 in Newark, NJ for January for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).

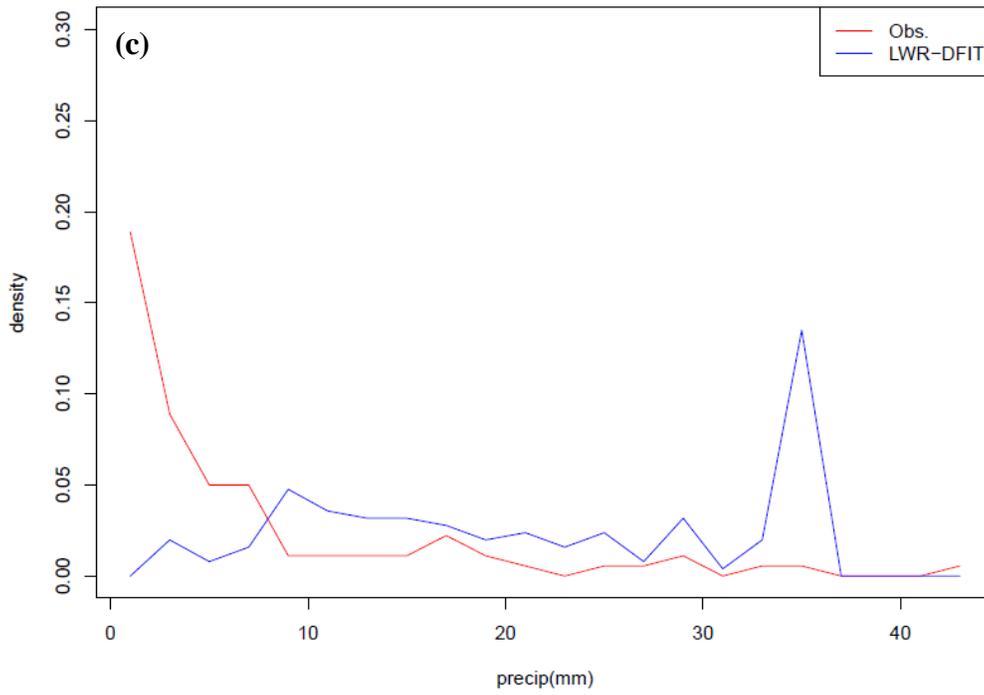
Comparison of Nonzero Rainfall PDF
Station: 286026 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 286026 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 286026 Month: 1



Comparison of Nonzero Rainfall PDF
Station: 286026 Month: 1

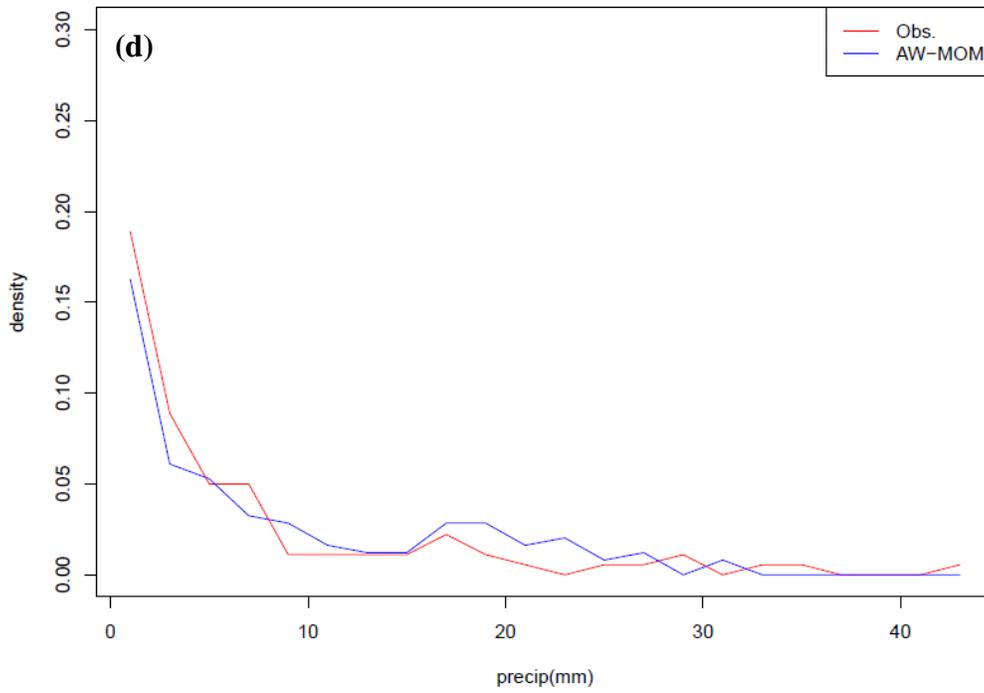
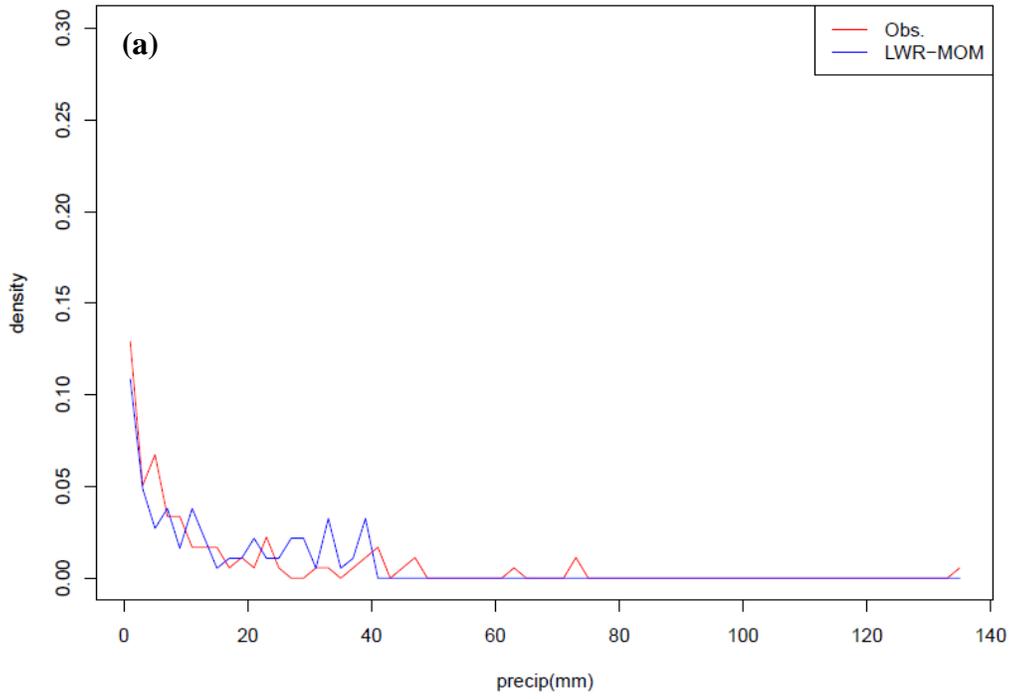
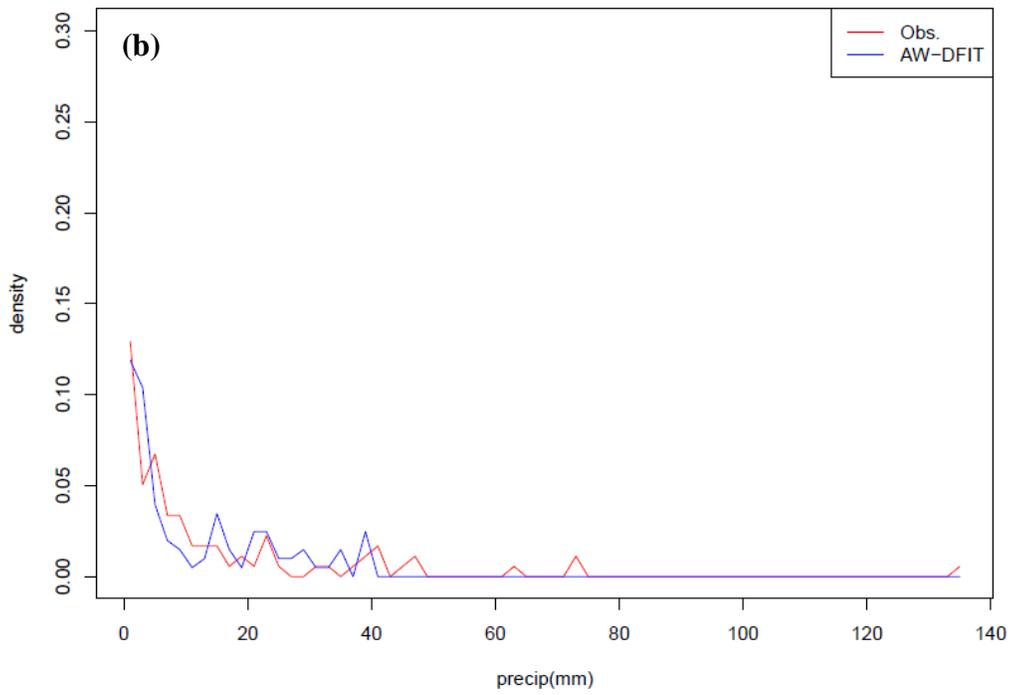


Figure 5.18. Observed vs. Generated PDFs of nonzero precipitation for station 090969 in Blairsville, GA for September for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).

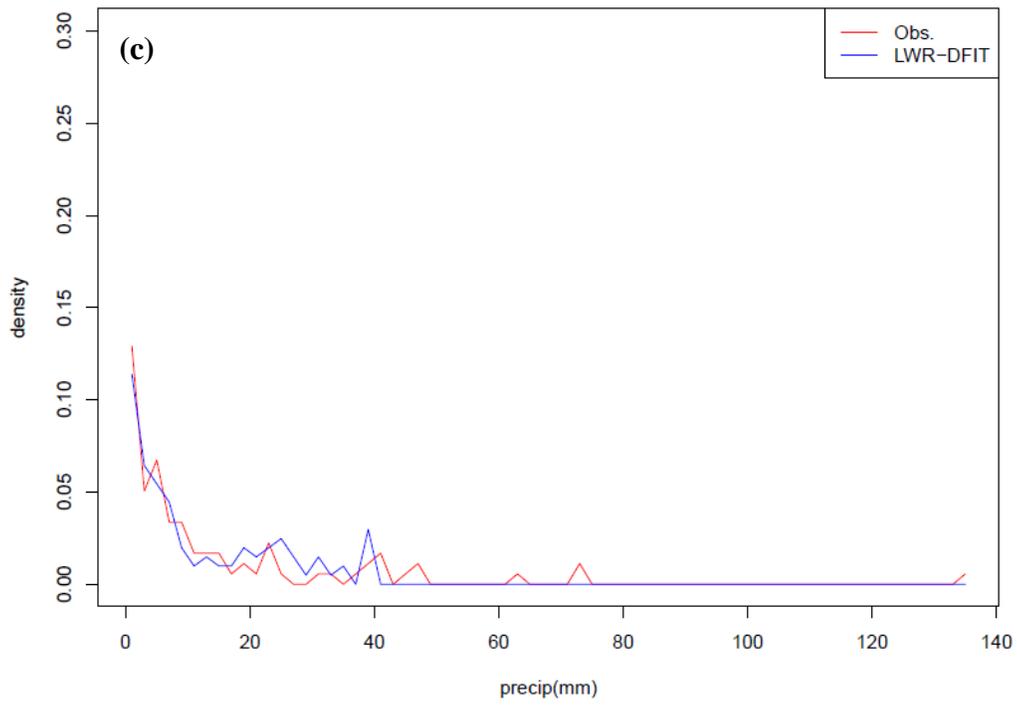
Comparison of Nonzero Rainfall PDF
Station: 090969 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 090969 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 090969 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 090969 Month: 9

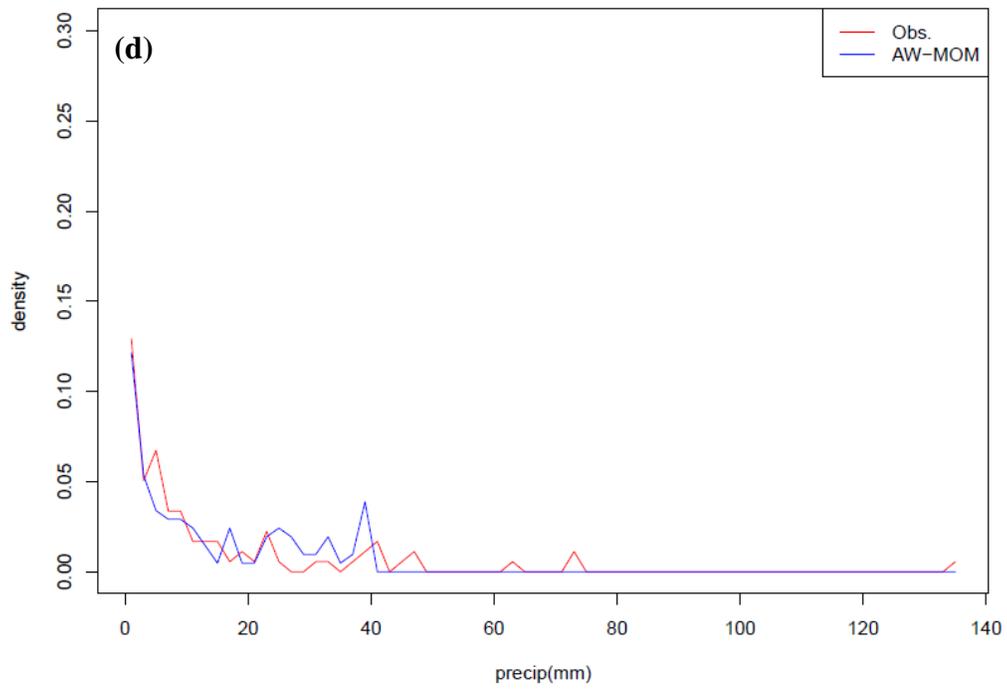
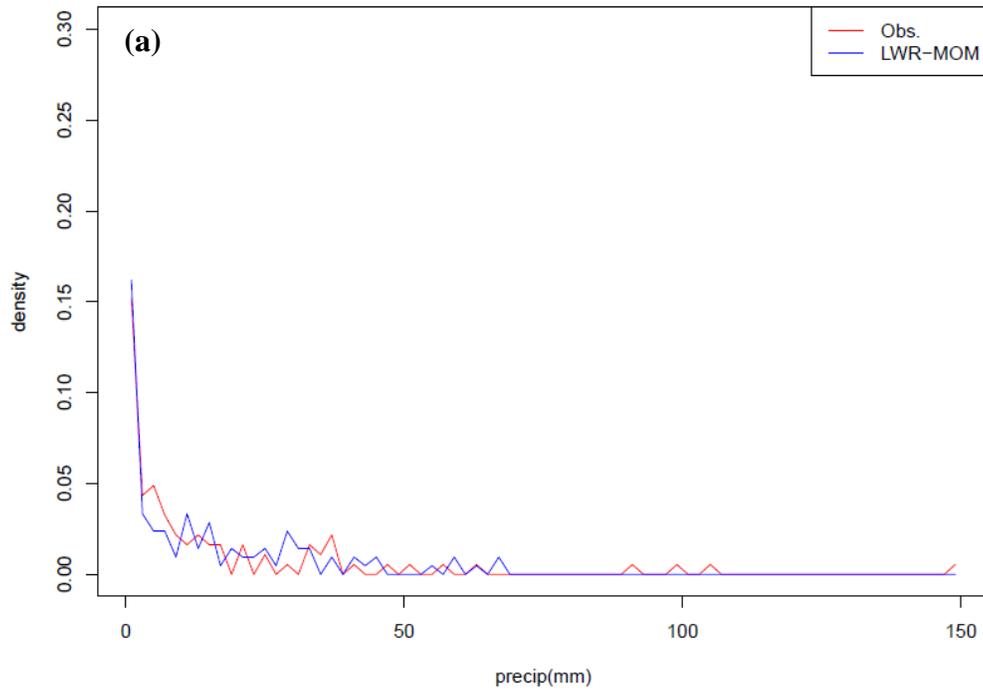
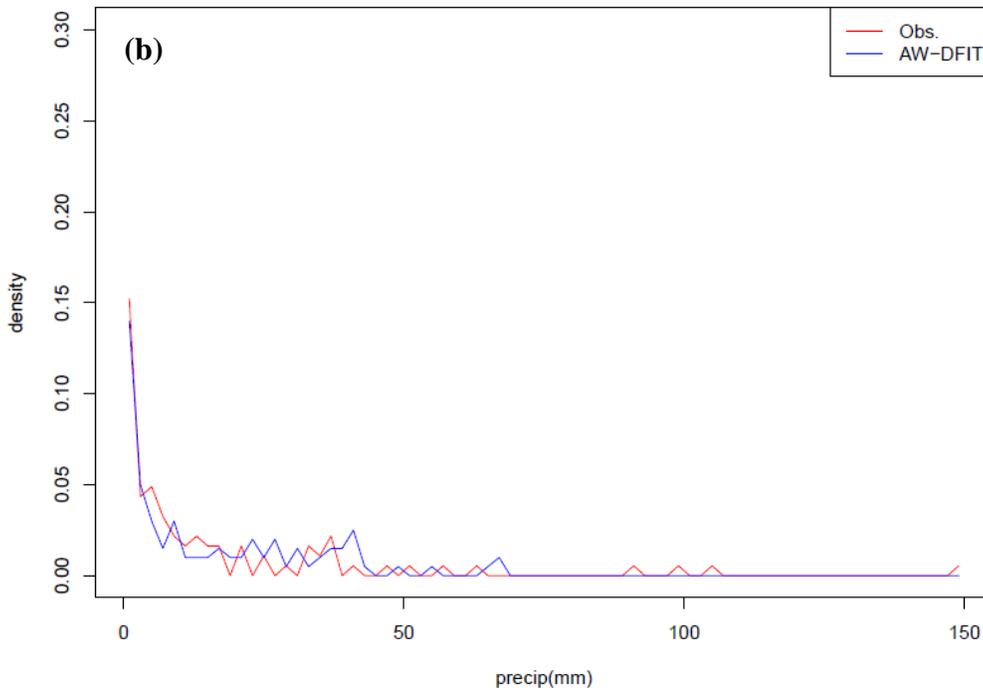


Figure 5.19. Observed vs. Generated PDFs of nonzero precipitation for station 319457 in Wilmington, NC for September for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).

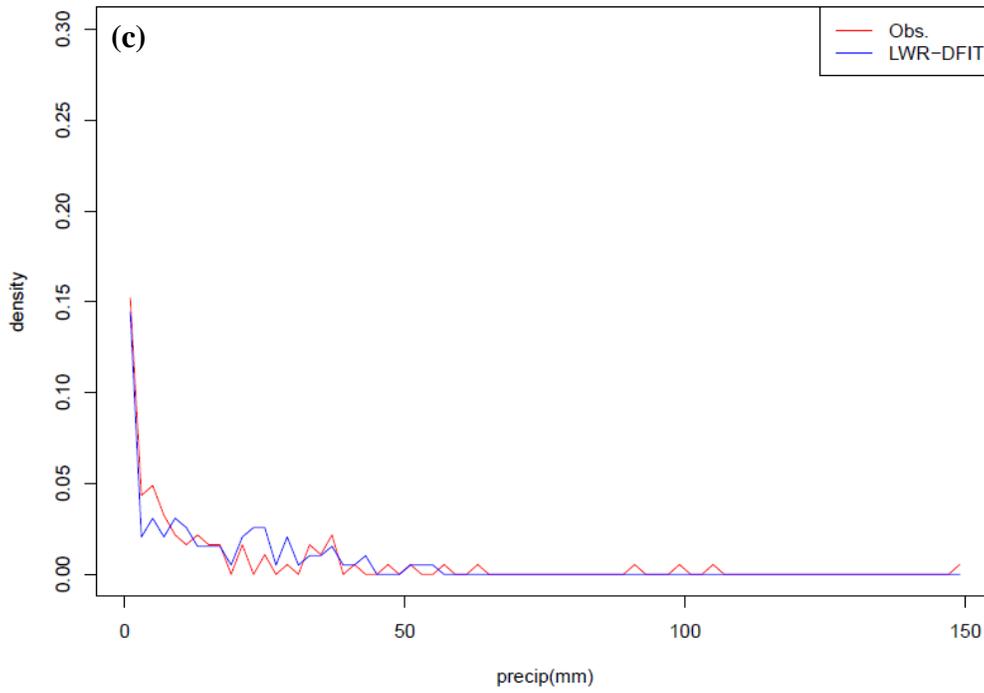
Comparison of Nonzero Rainfall PDF
Station: 319457 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 319457 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 319457 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 319457 Month: 9

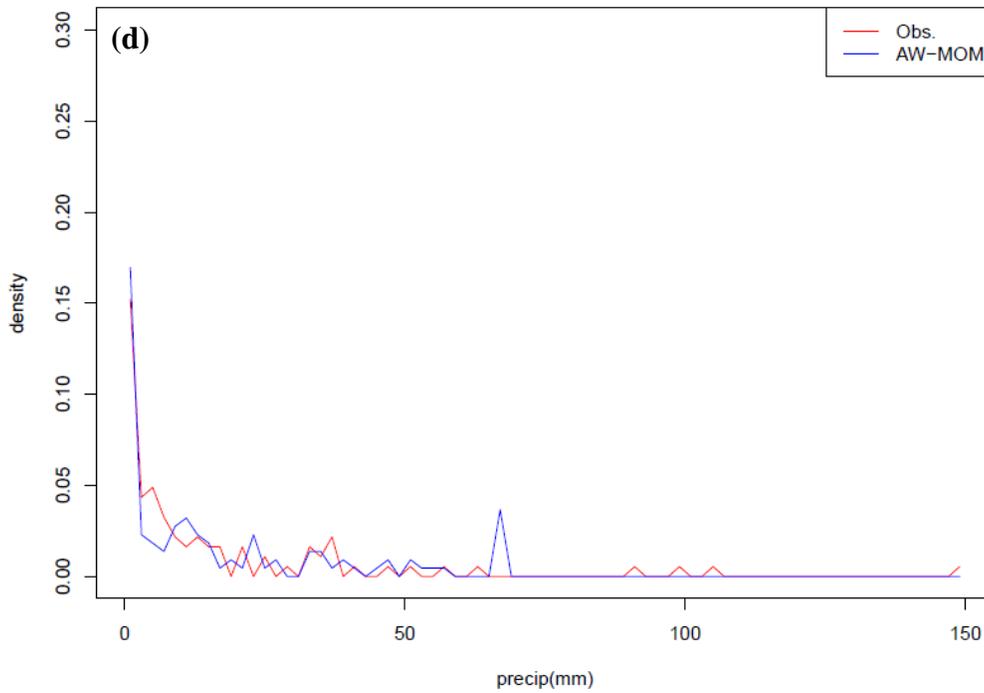
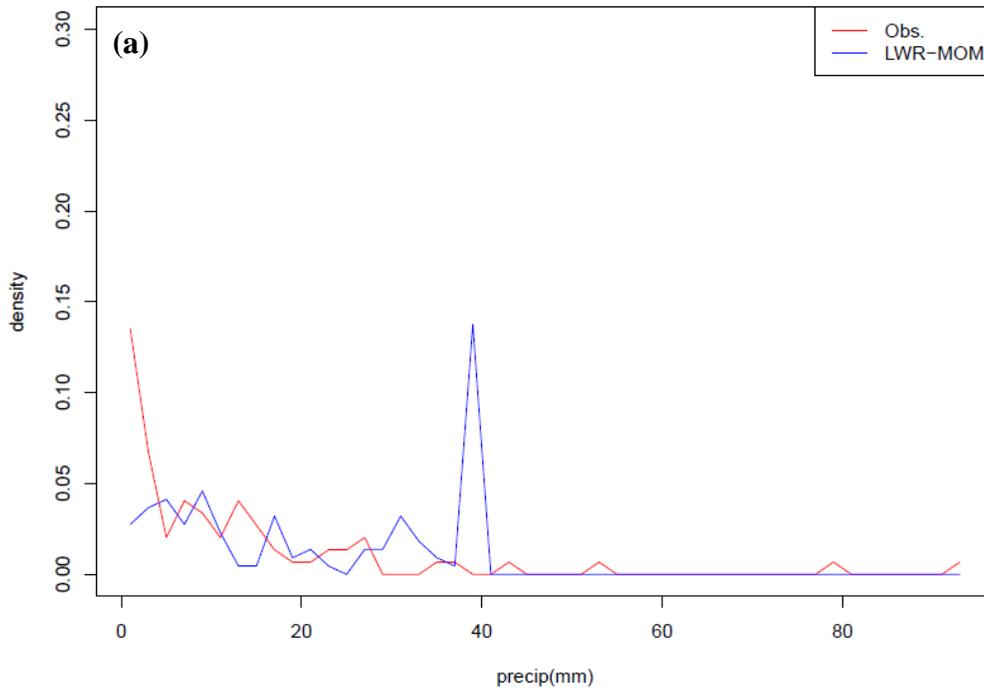
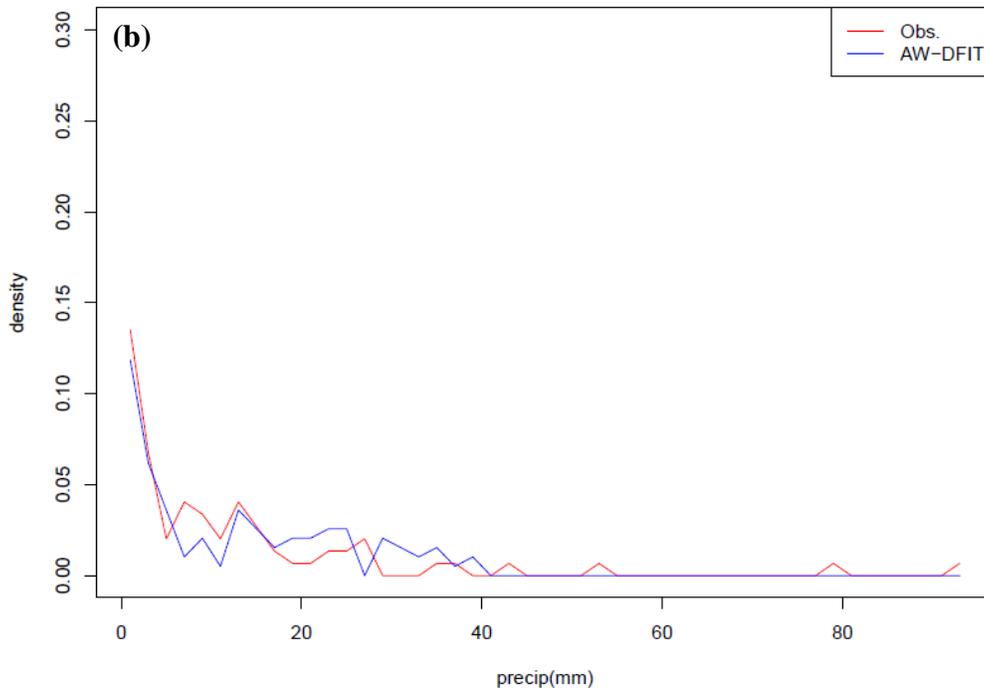


Figure 5.20. Observed vs. Generated PDFs of nonzero precipitation for station 286026 in Newark, NJ for September for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).

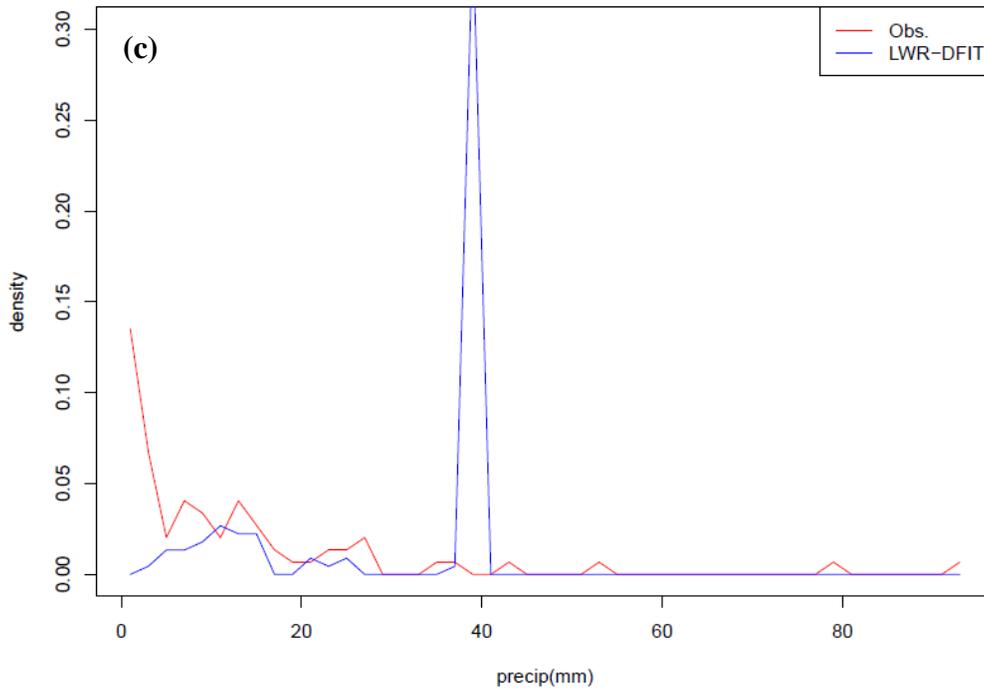
Comparison of Nonzero Rainfall PDF
Station: 286026 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 286026 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 286026 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 286026 Month: 9

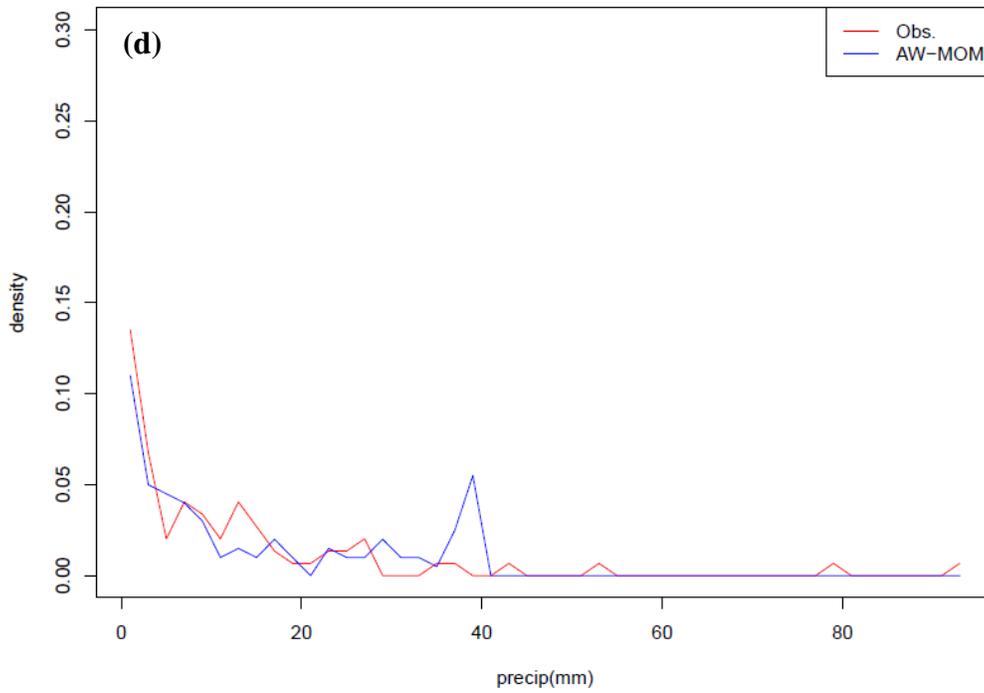
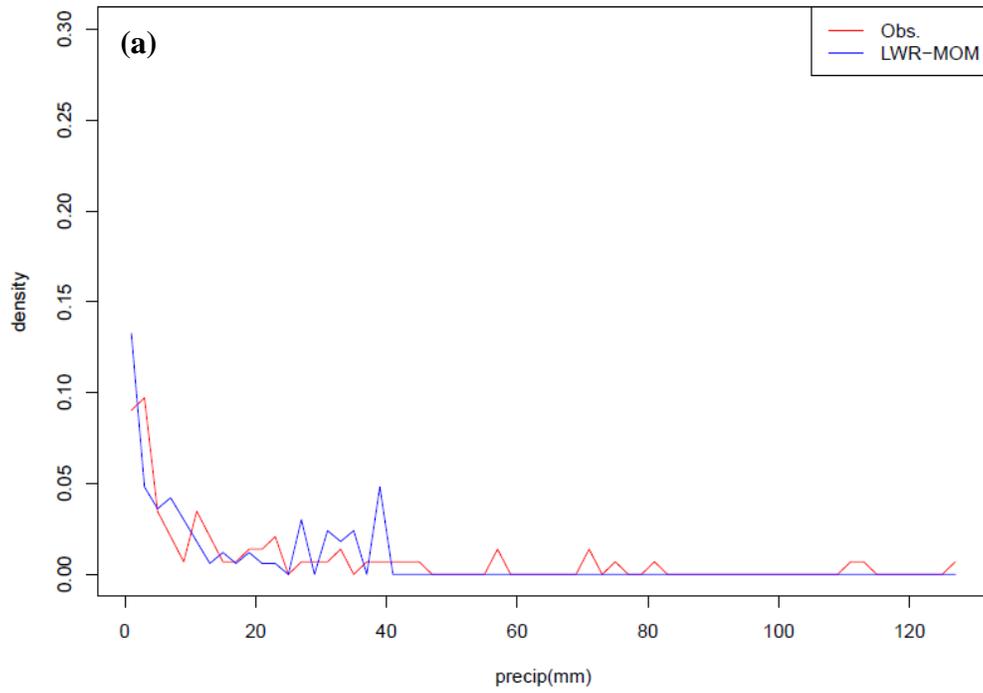
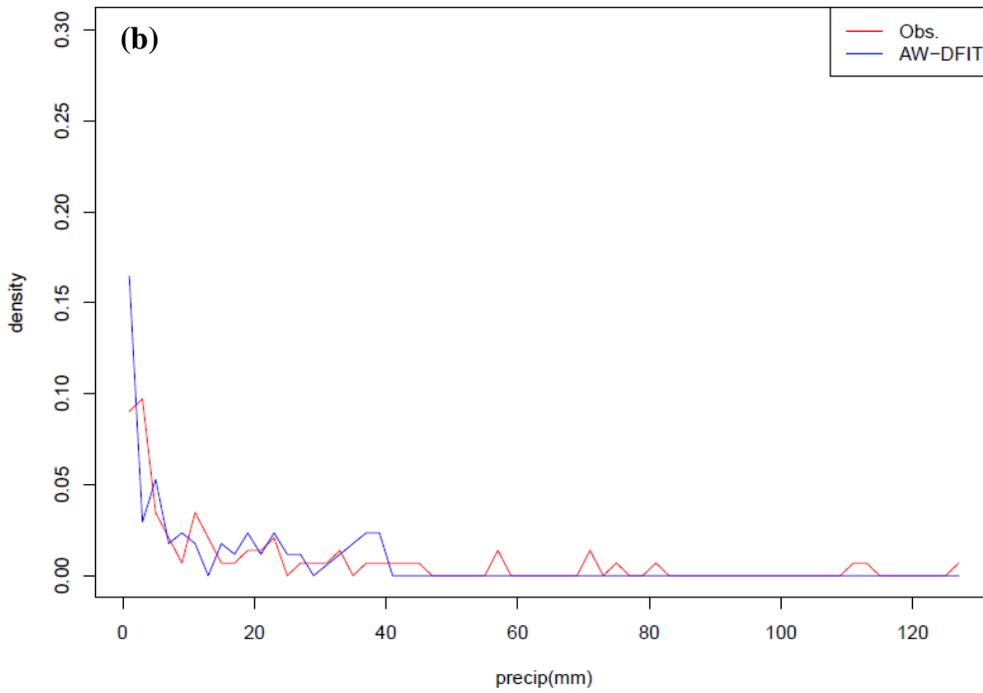


Figure 5.21. Observed vs. Generated PDFs of nonzero precipitation for station 092283 in Cornelia, GA for September for LWR-MOM (a), AW-DFIT (b), LWR-DFIT (c), and AW-MOM (d).

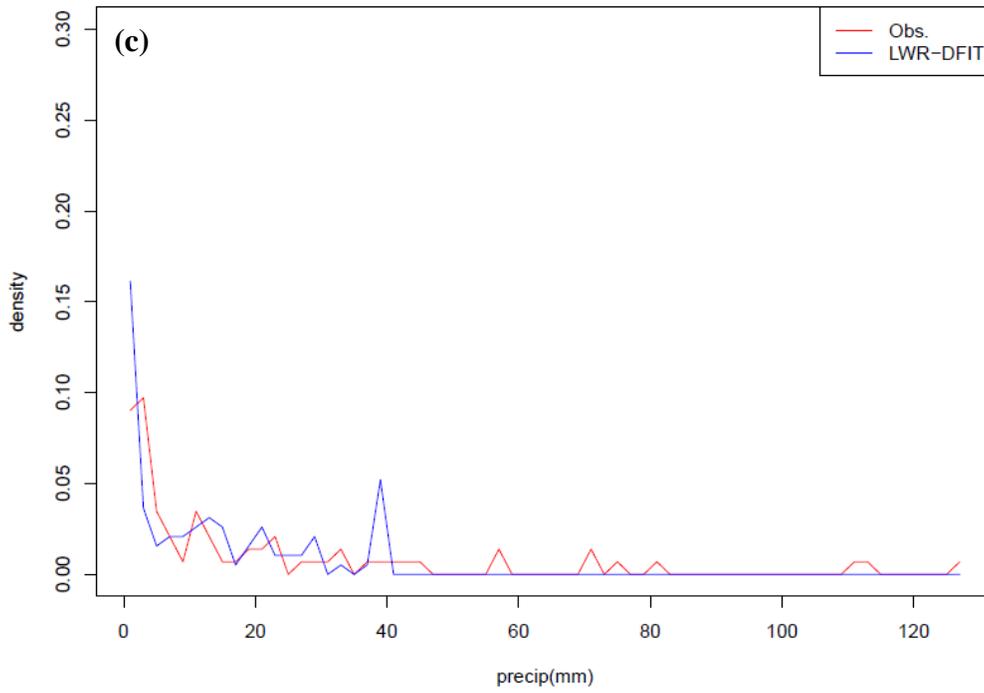
Comparison of Nonzero Rainfall PDF
Station: 092283 Month: 9



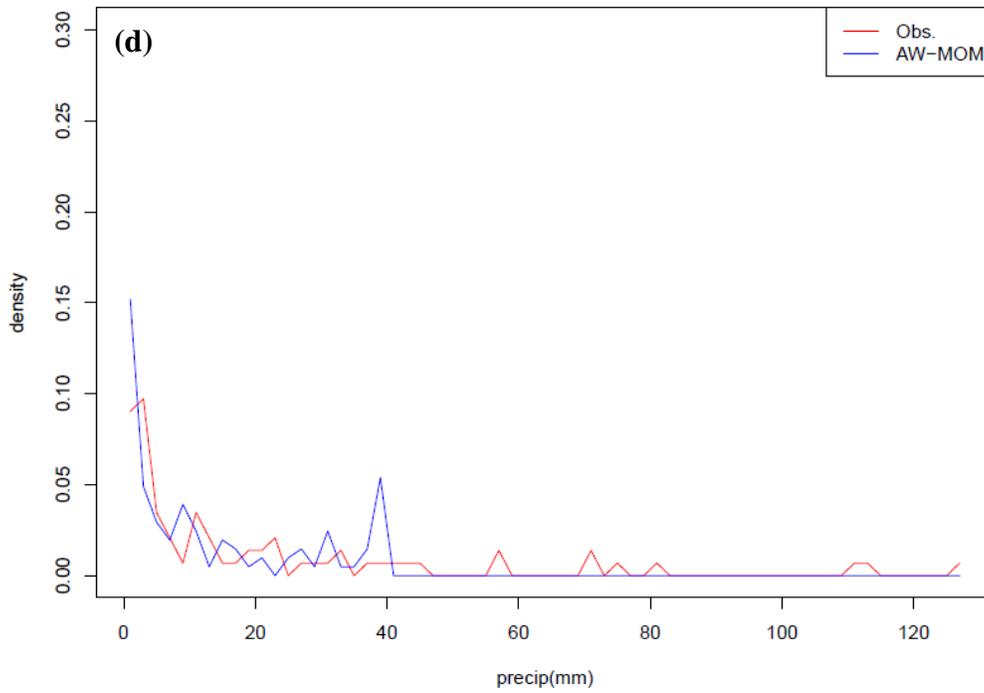
Comparison of Nonzero Rainfall PDF
Station: 092283 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 092283 Month: 9



Comparison of Nonzero Rainfall PDF
Station: 092283 Month: 9



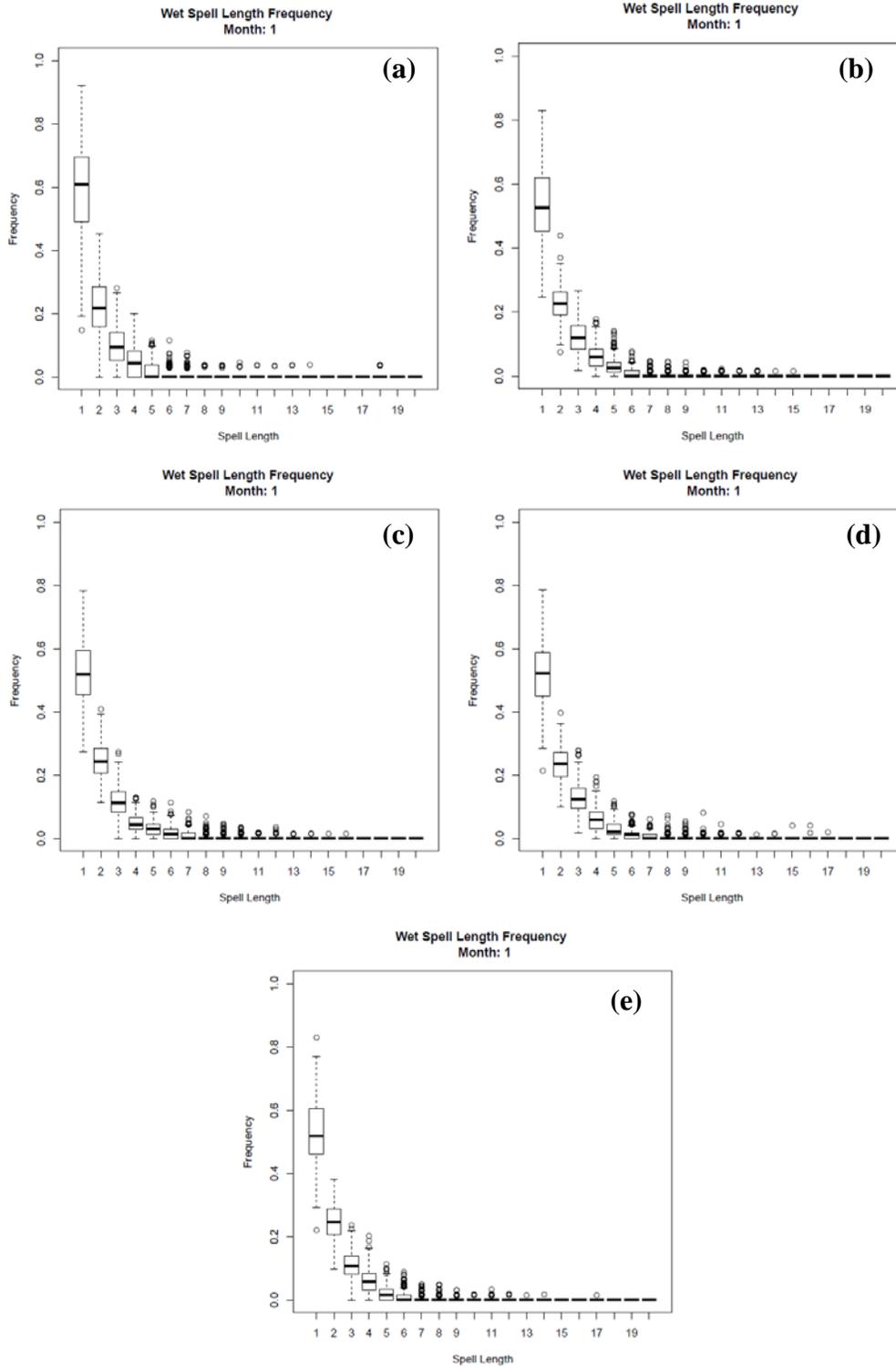


Figure 5.22. Boxplots of January Wet Spell Frequencies in 2001-2009 for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).

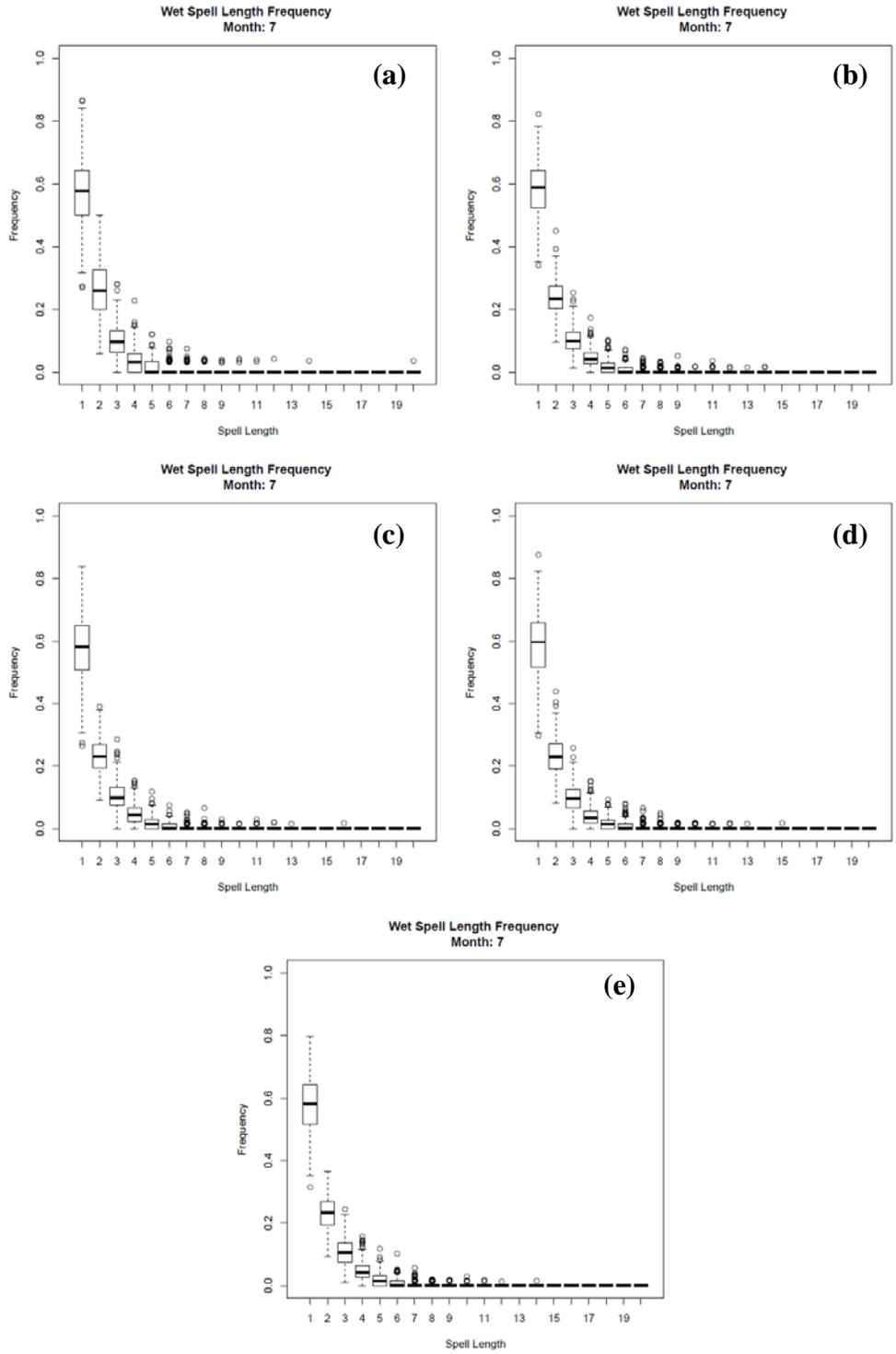


Figure 5.23. Boxplots of July Wet Spell Frequencies in 2001-2009 for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).

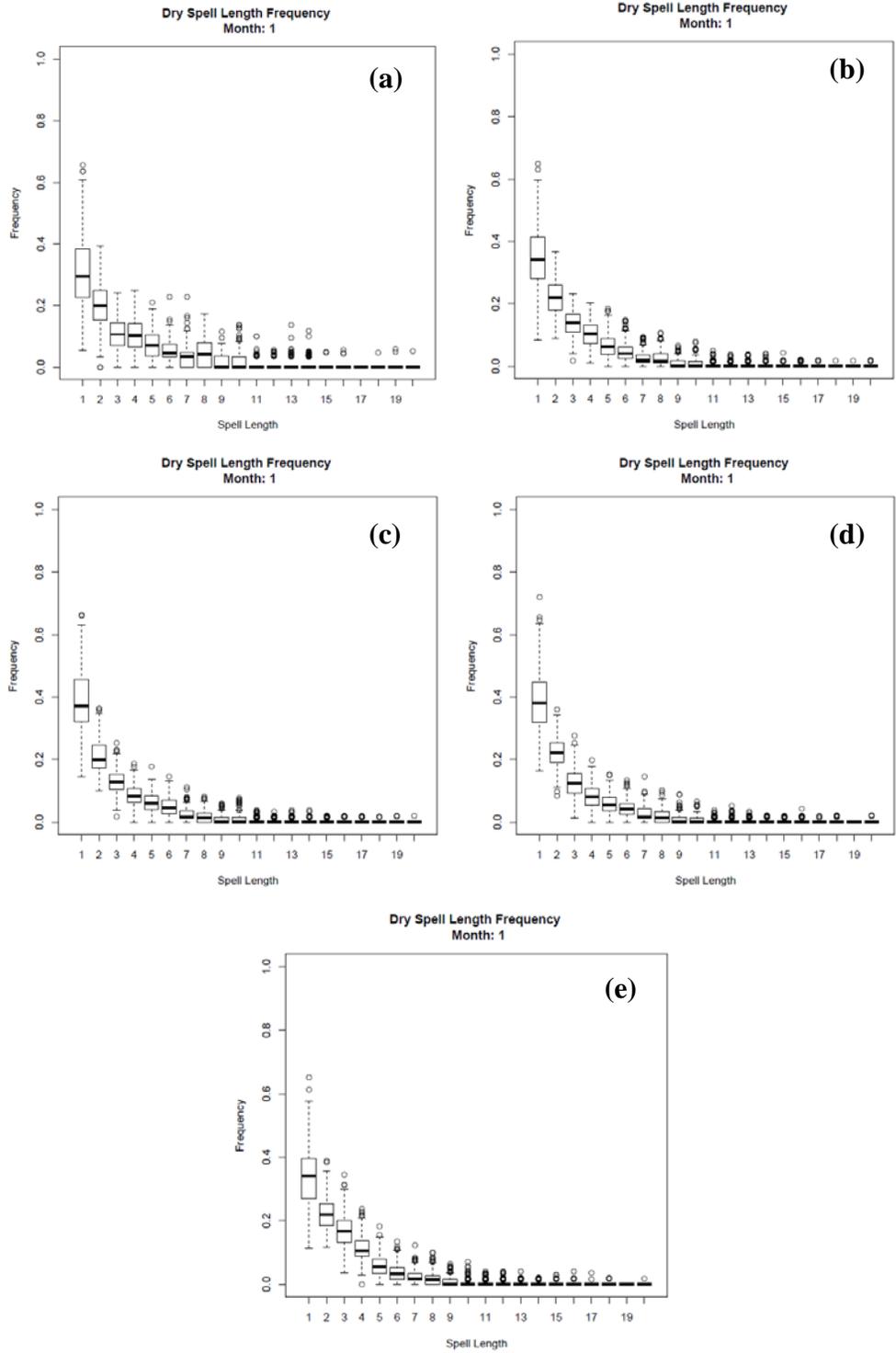


Figure 5.24. Boxplots of January Dry Spell Frequencies in 2001-2009 for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).

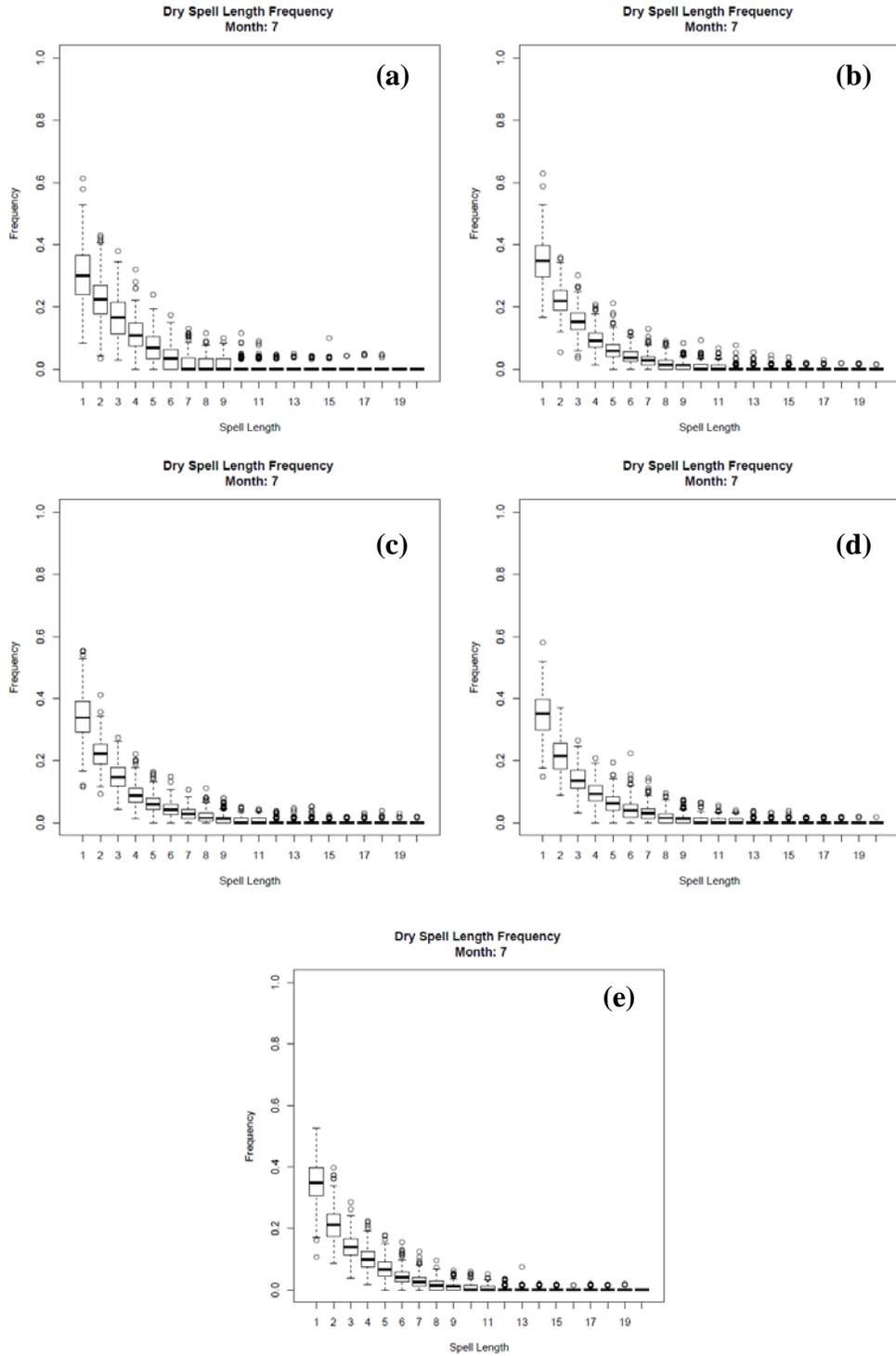


Figure 5.25. Boxplots of July Dry Spell Frequencies in 2001-2009 for observations (a), LWR-MOM (b), AW-DFIT (c), LWR-DFIT (d) and AW-MOM (e).

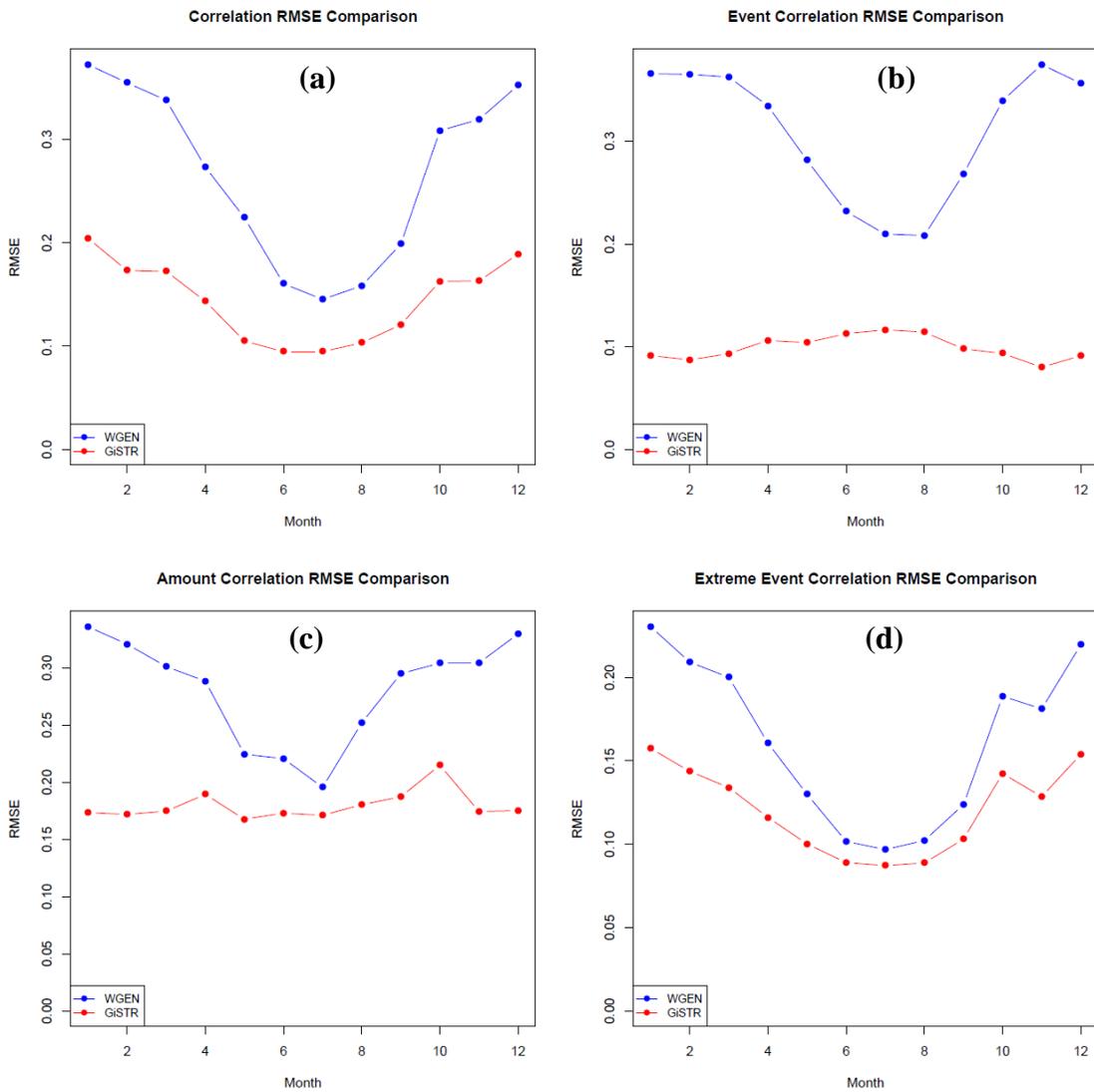


Figure 5.26. RMSE of the WGEN and GiSTR downscaling simulation by month for the correlation matrix of precipitation (ρ) (a), the correlation matrix of precipitation events (ρ_{ev}) (b), the correlation matrix of precipitation amounts (ρ_{am}) (c), and the correlation matrix of precipitation extreme events (ρ_{ex}) (d). Time period of 2001-2009.

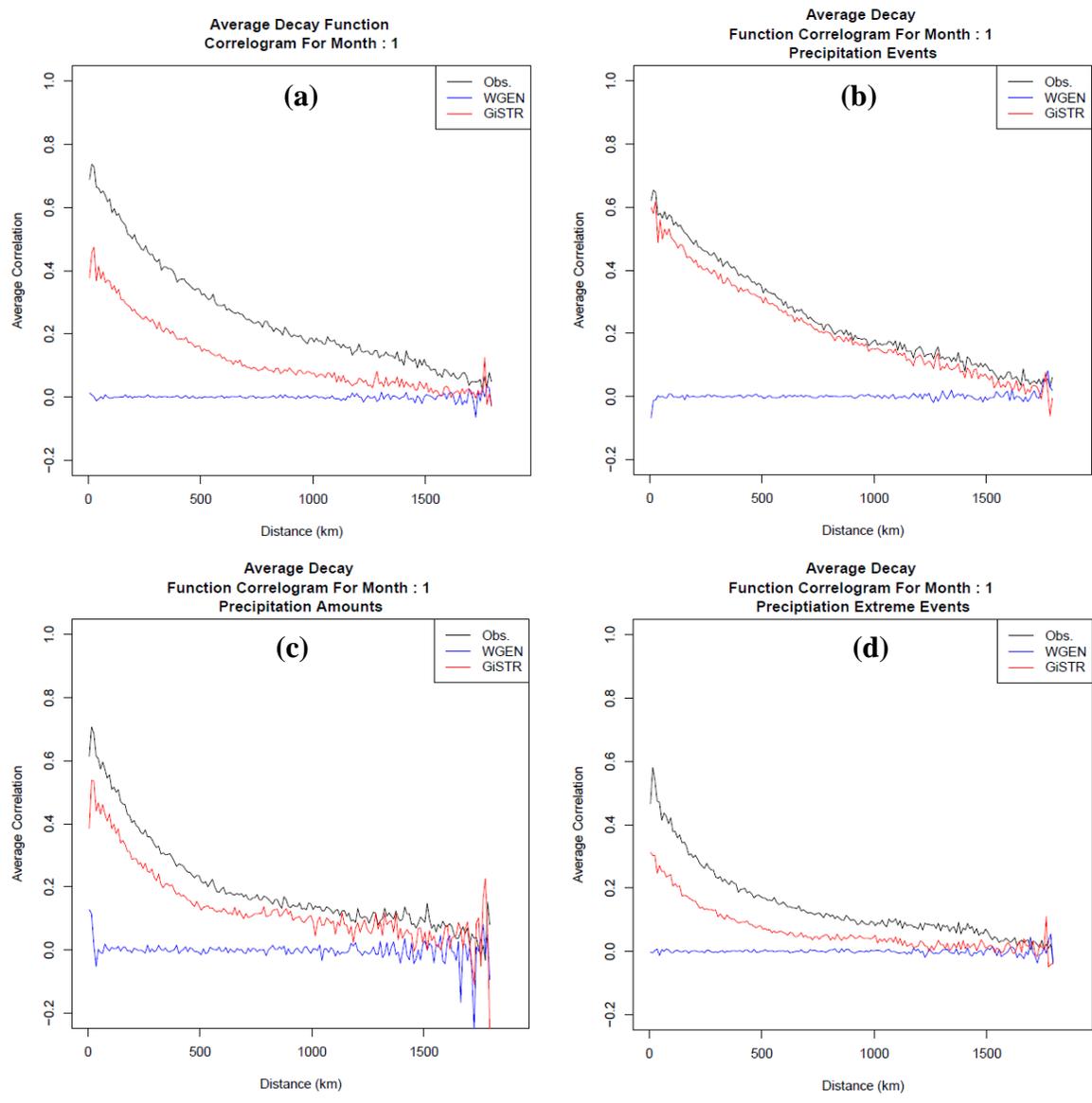


Figure 5.27. Average Decay Function Correlogram for WGEN and GiSTR compared to observations for January for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 2001-2009 time period.

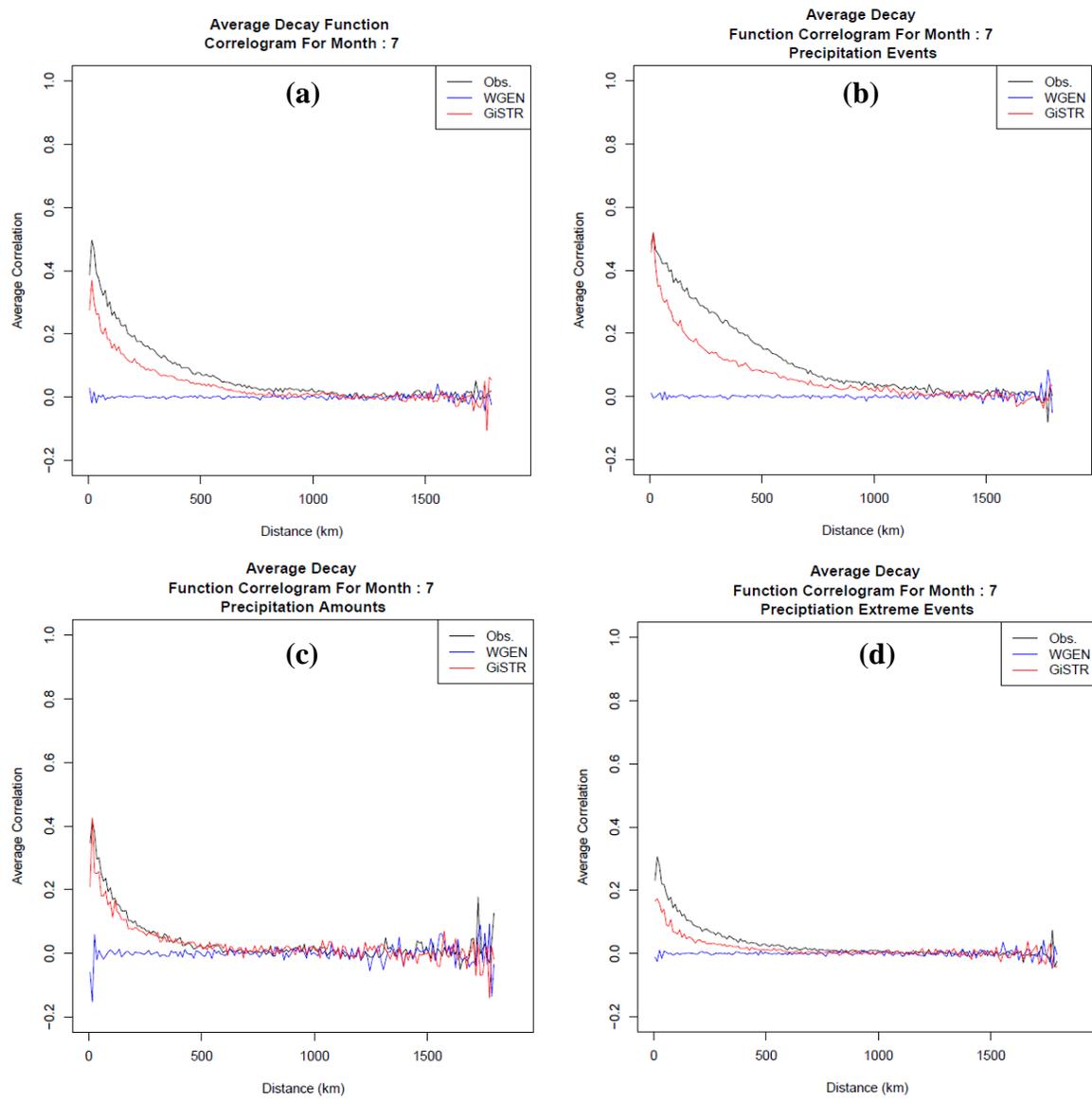


Figure 5.28. Average Decay Function Correlogram for WGEN and GiSTR compared to observations for July for precipitation (a), precipitation events (b), precipitation amounts (c), and precipitation extreme events (d). 2001-2009 time period.

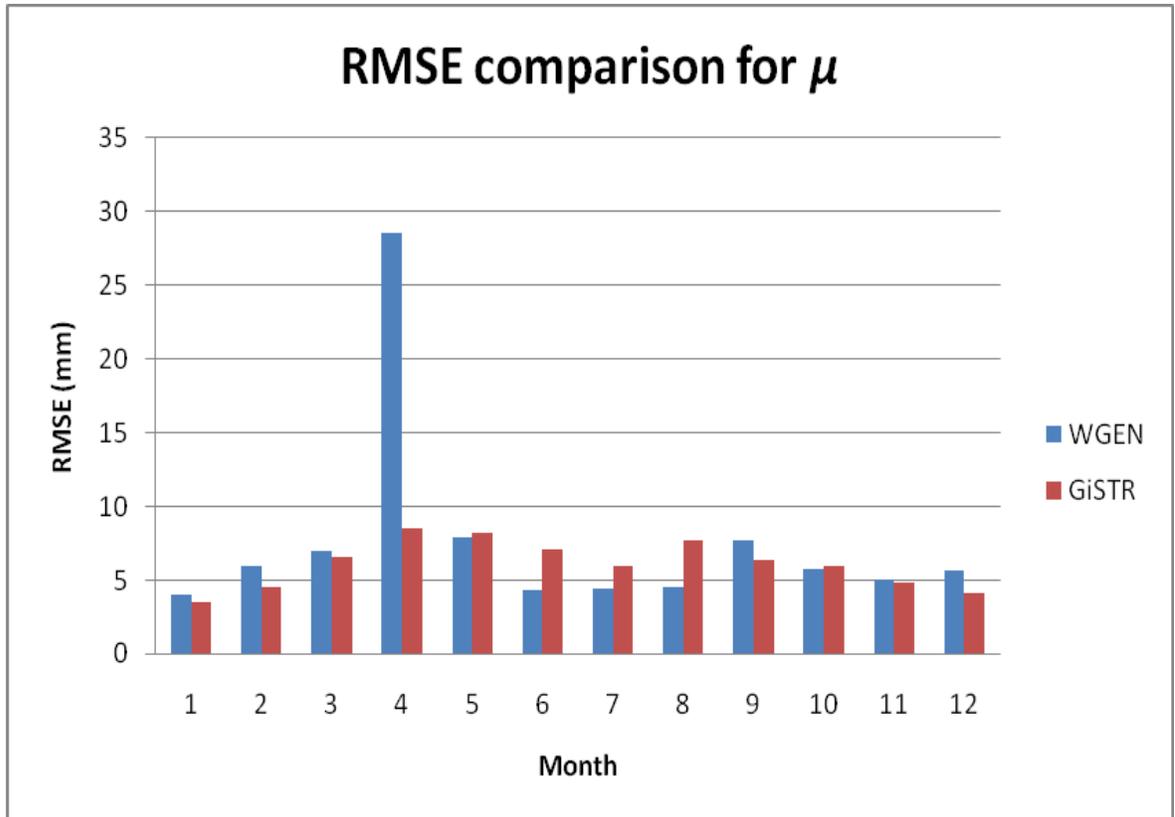
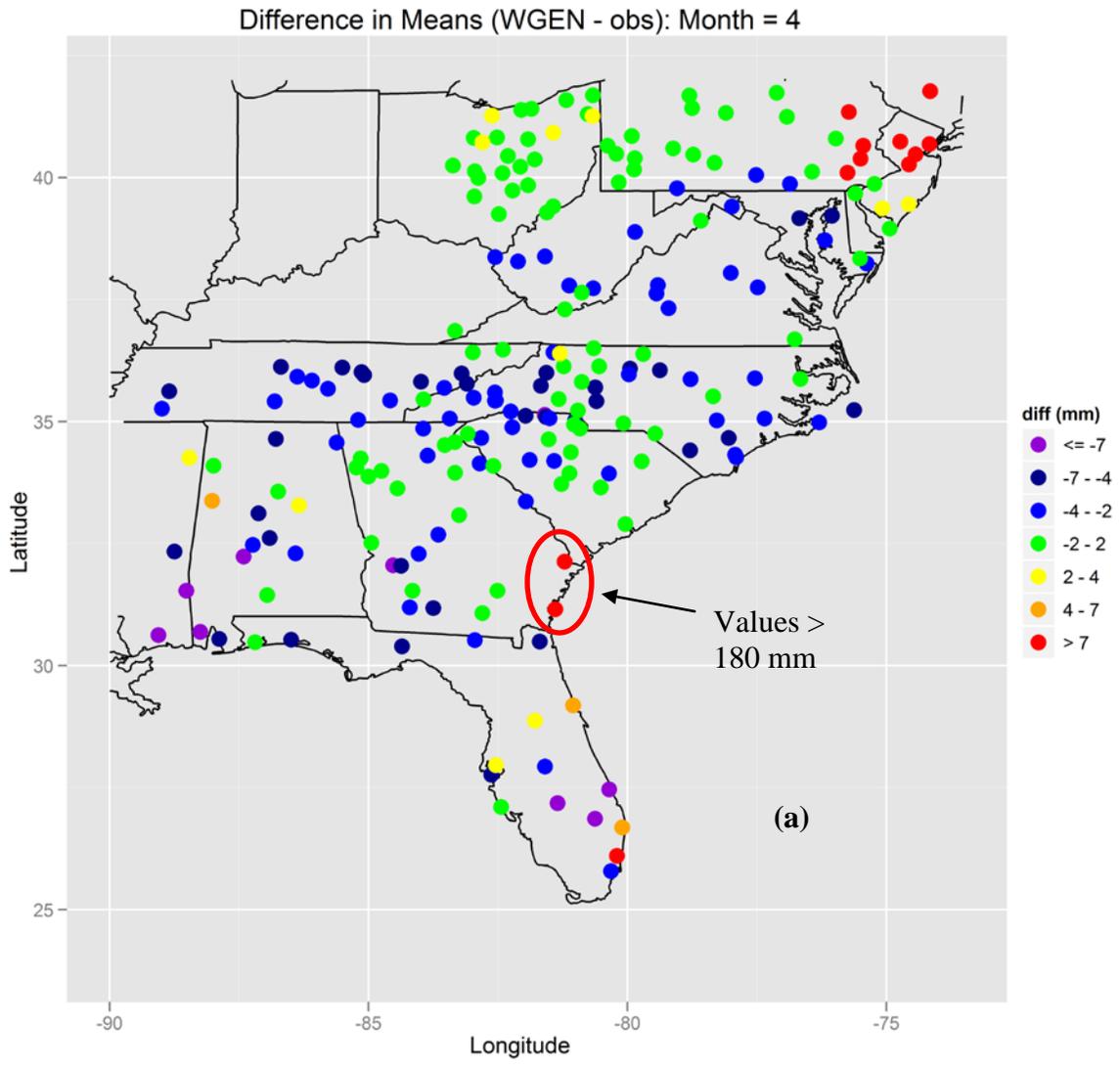
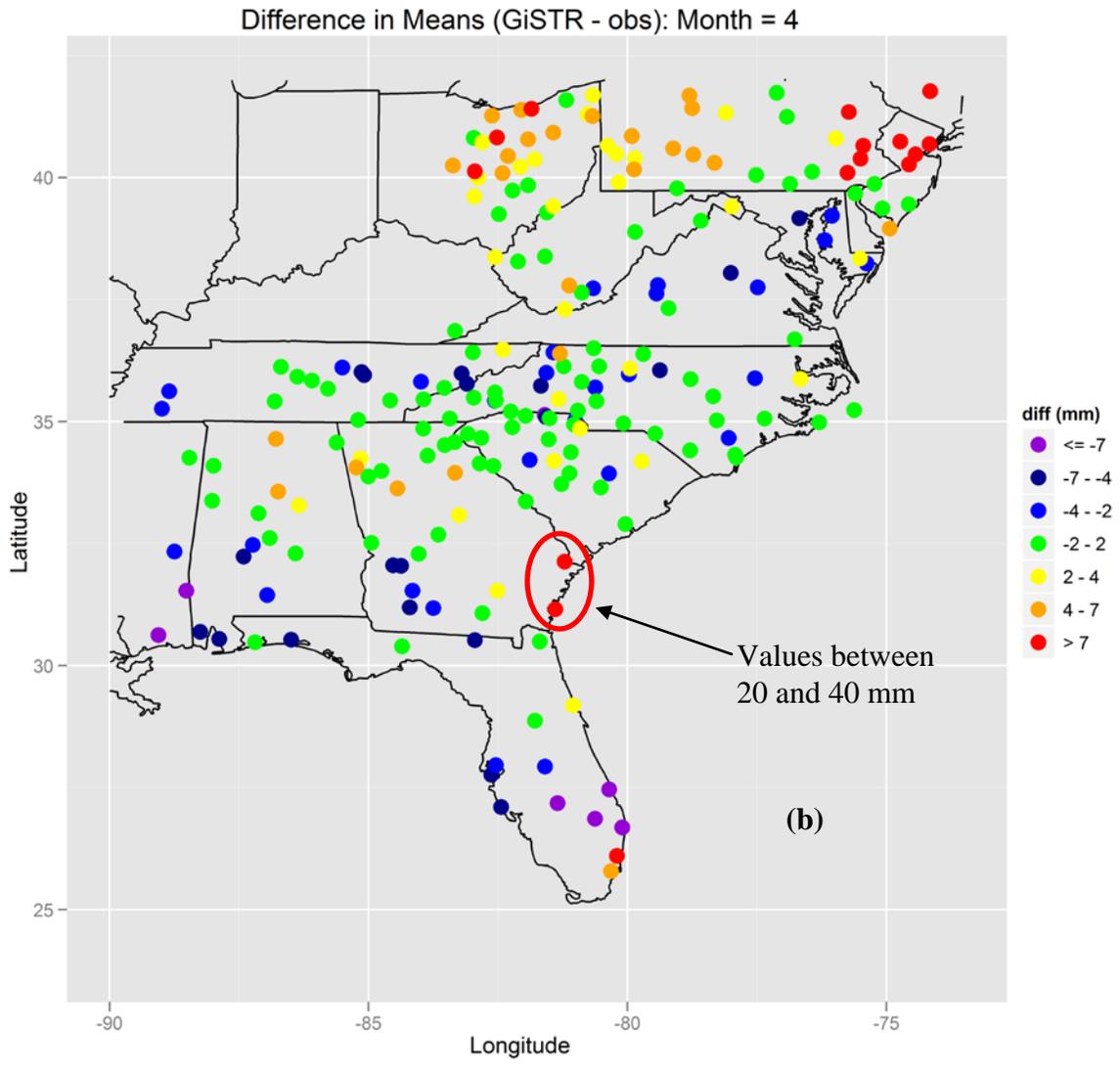


Figure 5.29. RMSE Comparison between downscaling simulations for mean daily nonzero precipitation (μ). 2001-2009 time period.

Figure 5.30. Difference between generated and observed values of mean daily nonzero precipitation (μ) for the WGEN simulation (a) and the GiSTR simulation (b) in April, 2001-2009 time period.





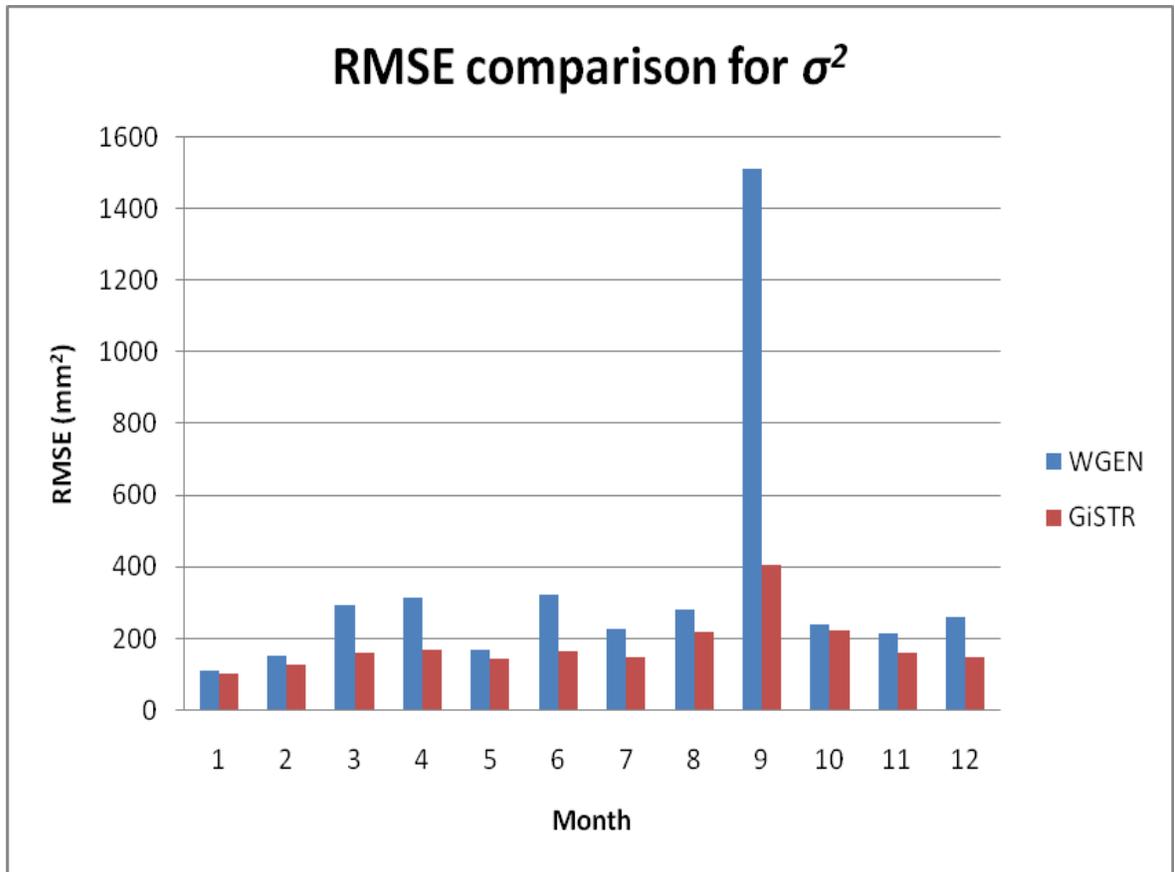
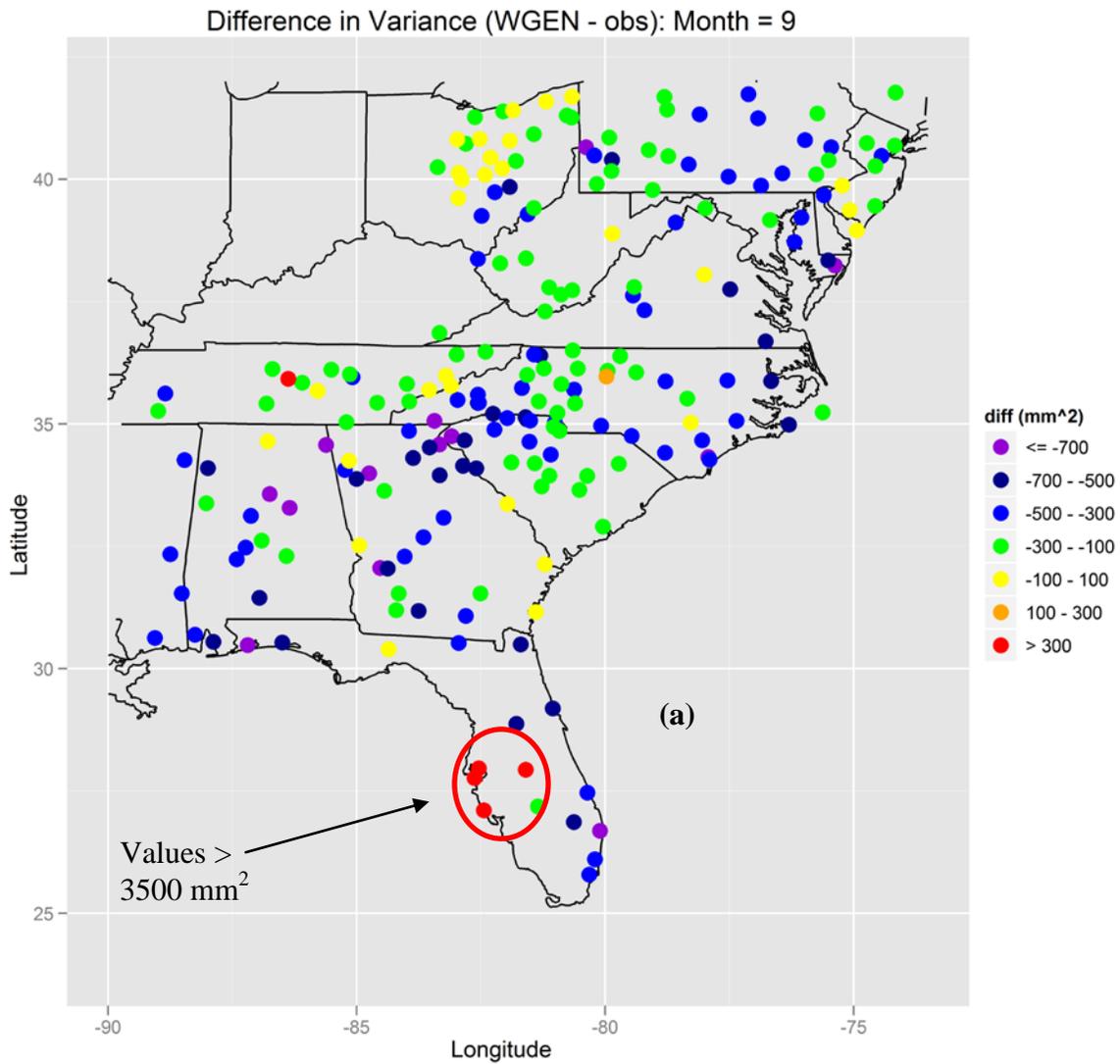
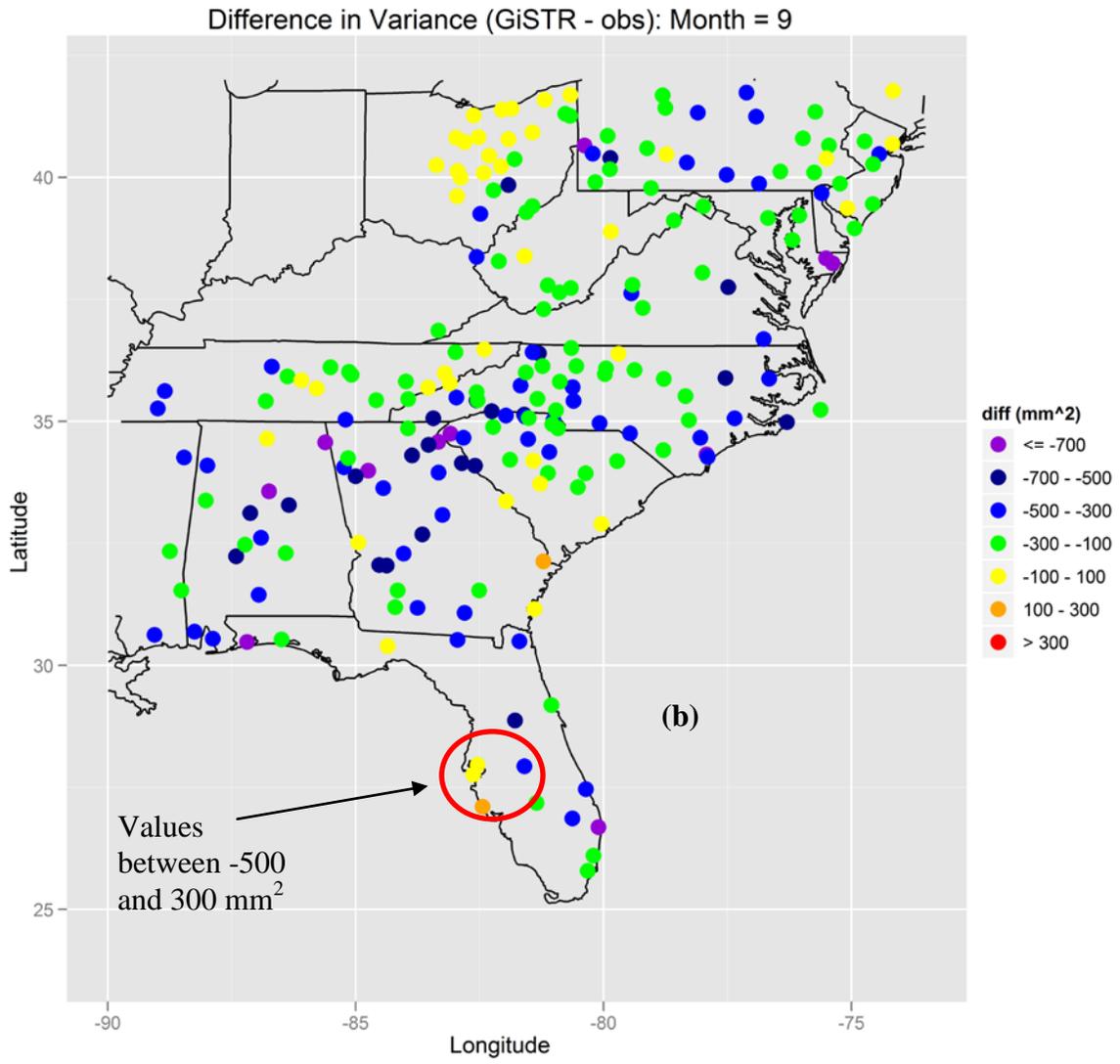


Figure 5.31. RMSE Comparison between downscaling simulations for the variance of daily nonzero precipitation (σ^2). 2001-2009 time period.

Figure 5.32. Difference between generated and observed values of the variance of daily nonzero precipitation (σ^2), for the WGEN simulation (a) and the GiSTR simulation (b) in September, 2001-2009 time period.





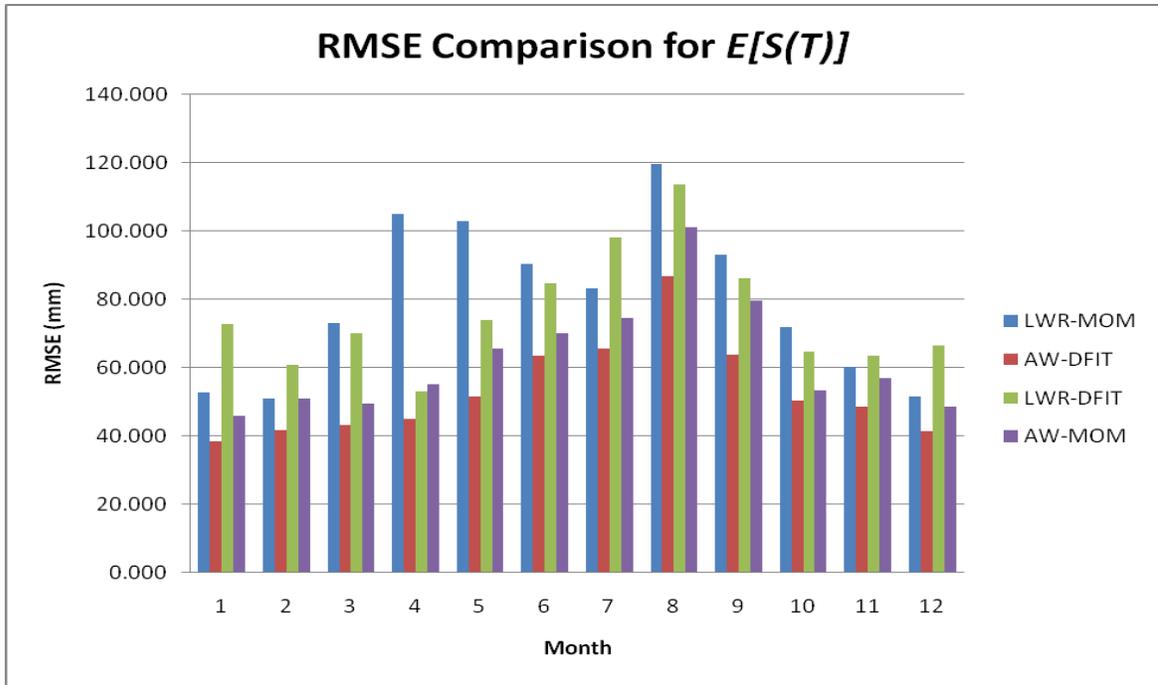


Figure 5.33. RMSE Comparison between downscaling simulations for the average total precipitation ($E[S(T)]$). 2001-2009 time period.

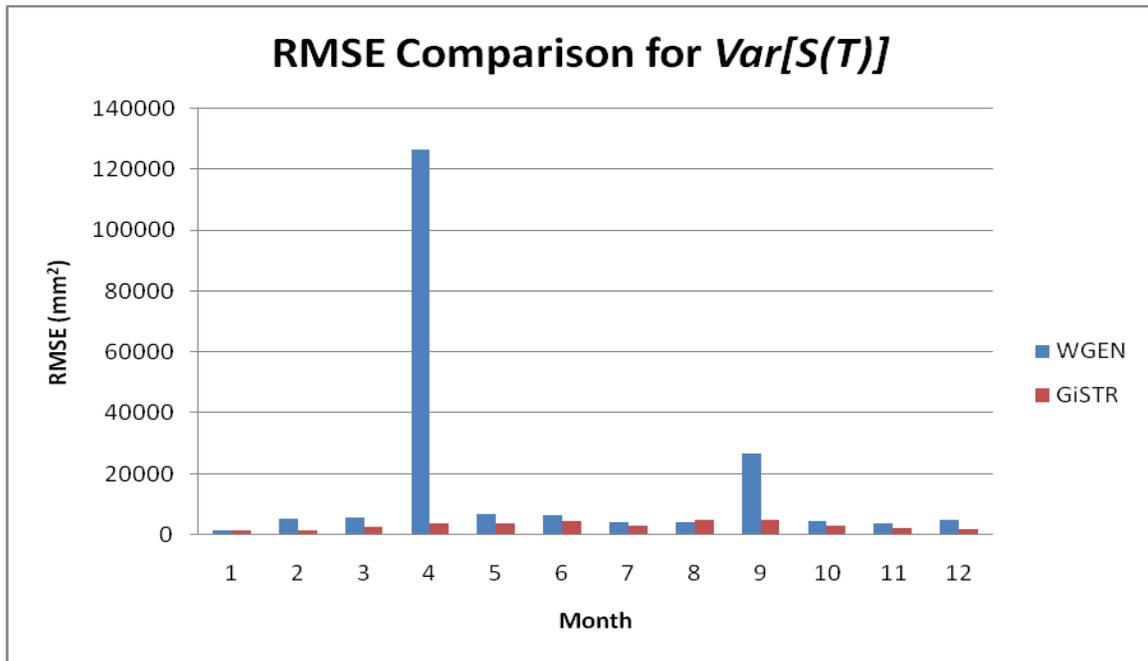
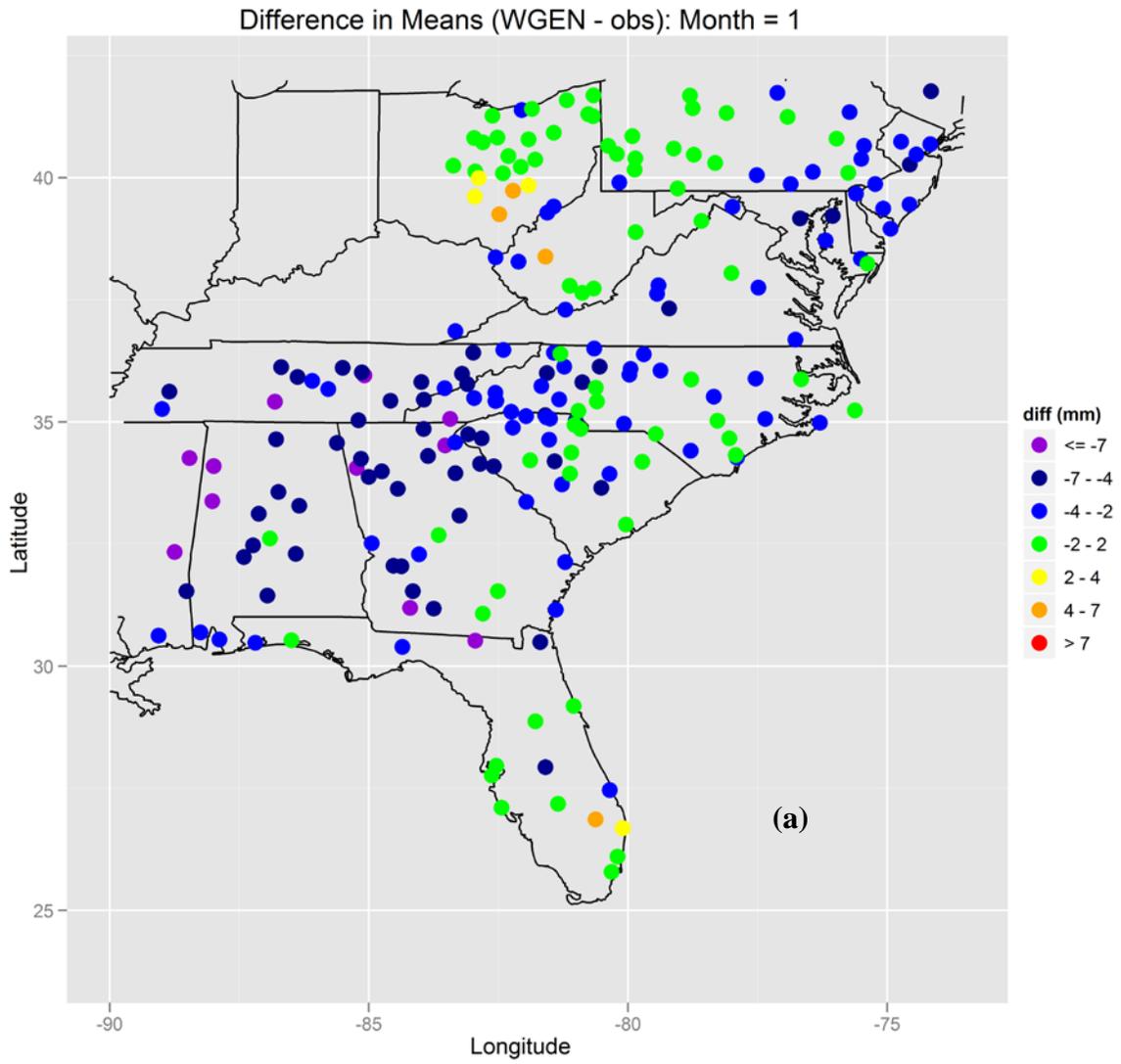
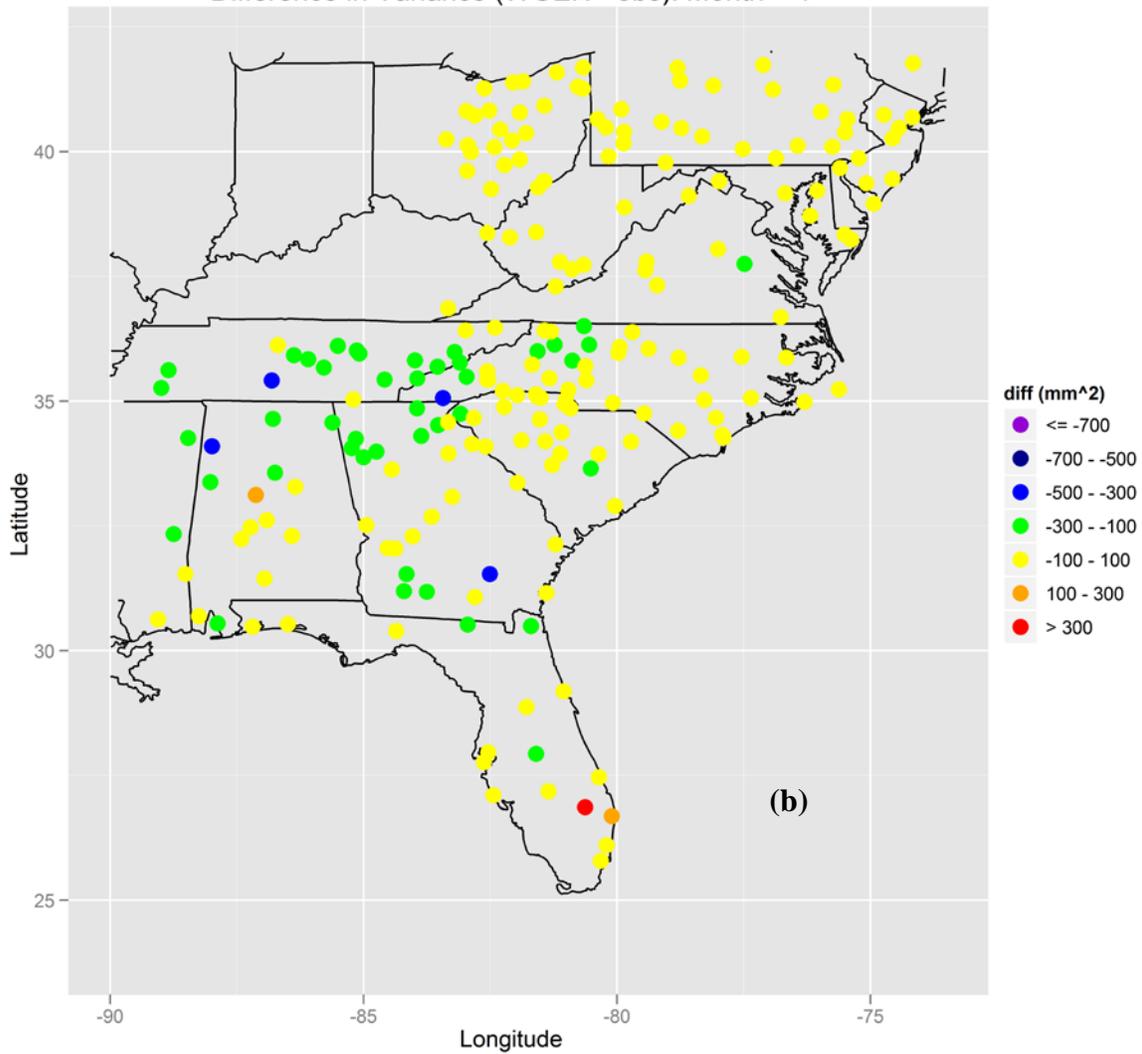


Figure 5.34. RMSE Comparison between downscaling simulations for the inter-annual variability ($Var[S(T)]$). 2001-2009 time period.

Figure 5.35. Difference between generated and observed values of the mean daily nonzero precipitation (μ) (a) and the variance of daily nonzero precipitation (σ^2). (b) in January for the WGEN downscale simulation. 2001-2009 time period.



Difference in Variance (WGEN - obs): Month = 1



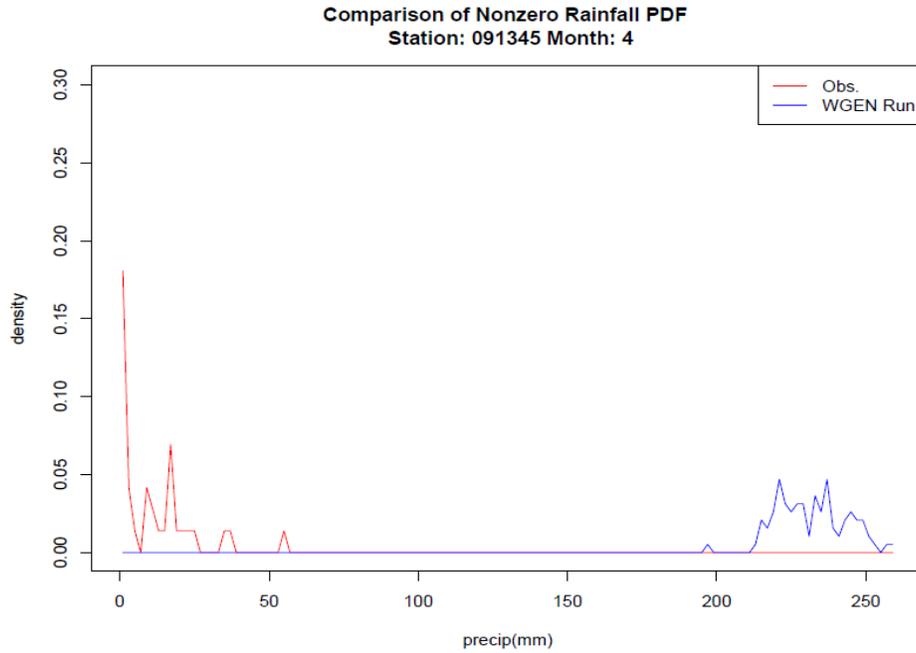


Figure 5.36. Difference between generated and observed PDF of nonzero precipitation for station 091345 in Brunswick, GA for the WGEN simulation in April.

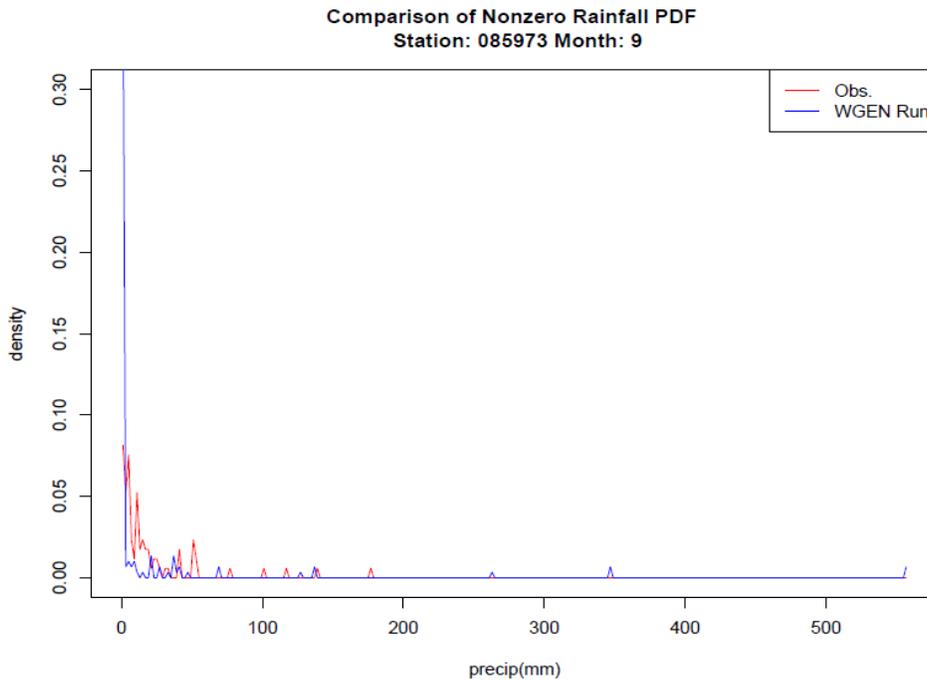


Figure 5.37. Difference between generated and observed PDF of nonzero precipitation for a station 092485 near Lake Wales, FL for the WGEN simulation in September.

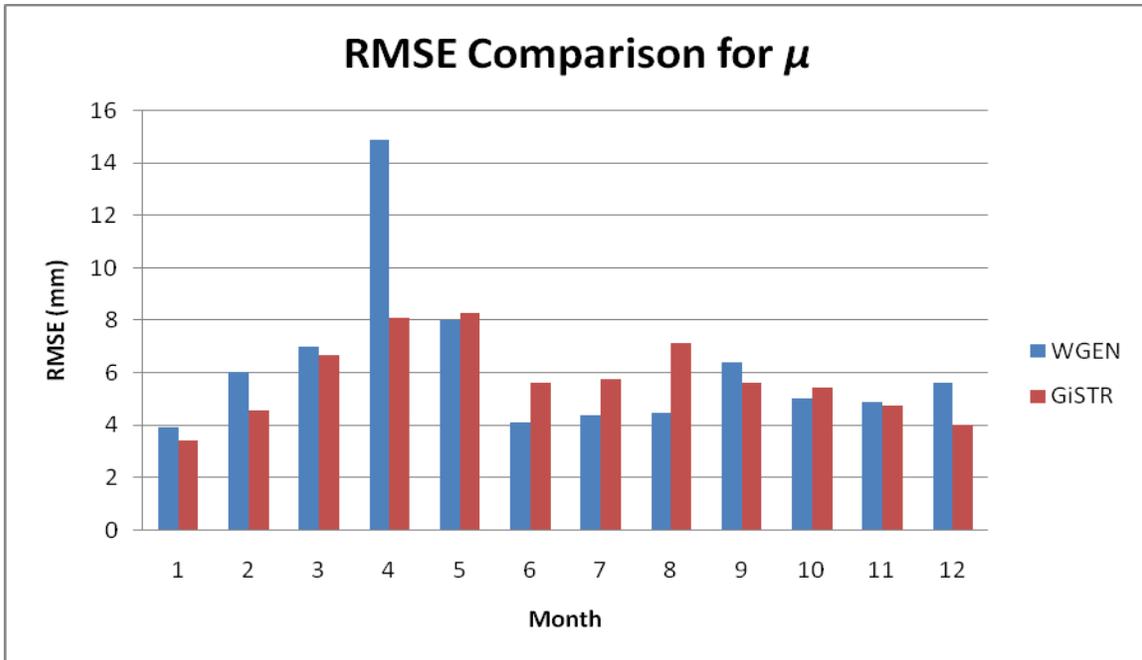


Figure 5.38. RMSE Comparison between downscaling simulations for the mean daily nonzero precipitation (μ). 2001-2009 time period. Outlier stations removed.

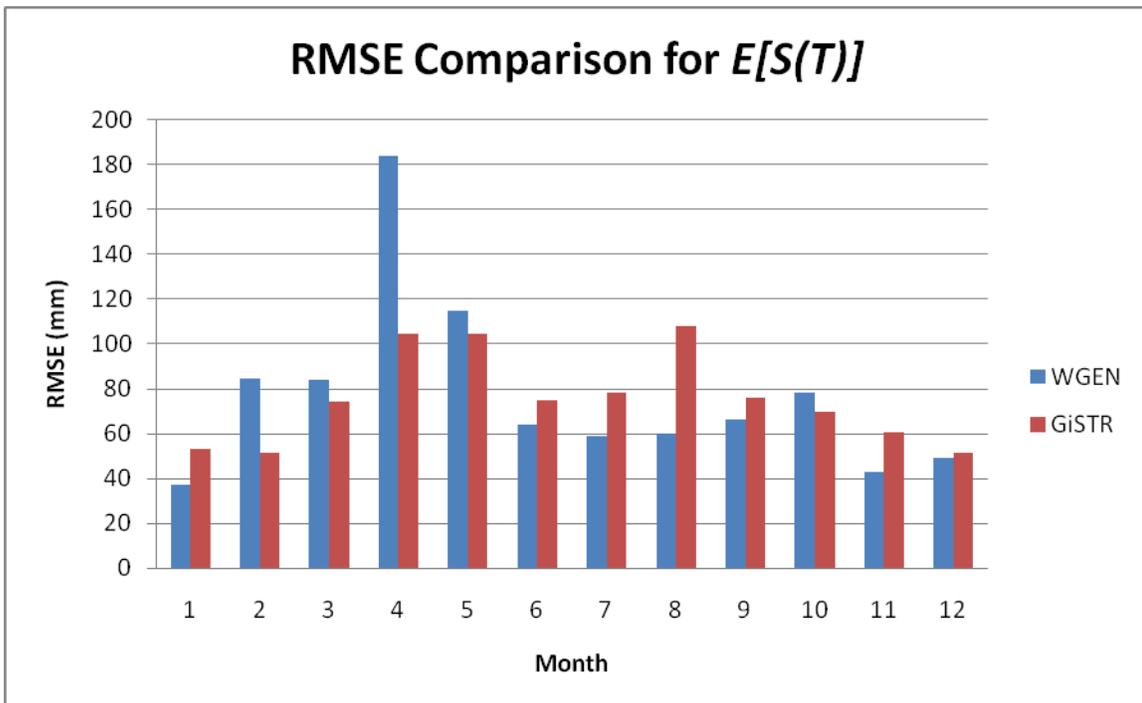


Figure 5.39. RMSE Comparison between downscaling simulations for the average total precipitation ($E[S(T)]$). 2001-2009 time period. Outlier stations removed.

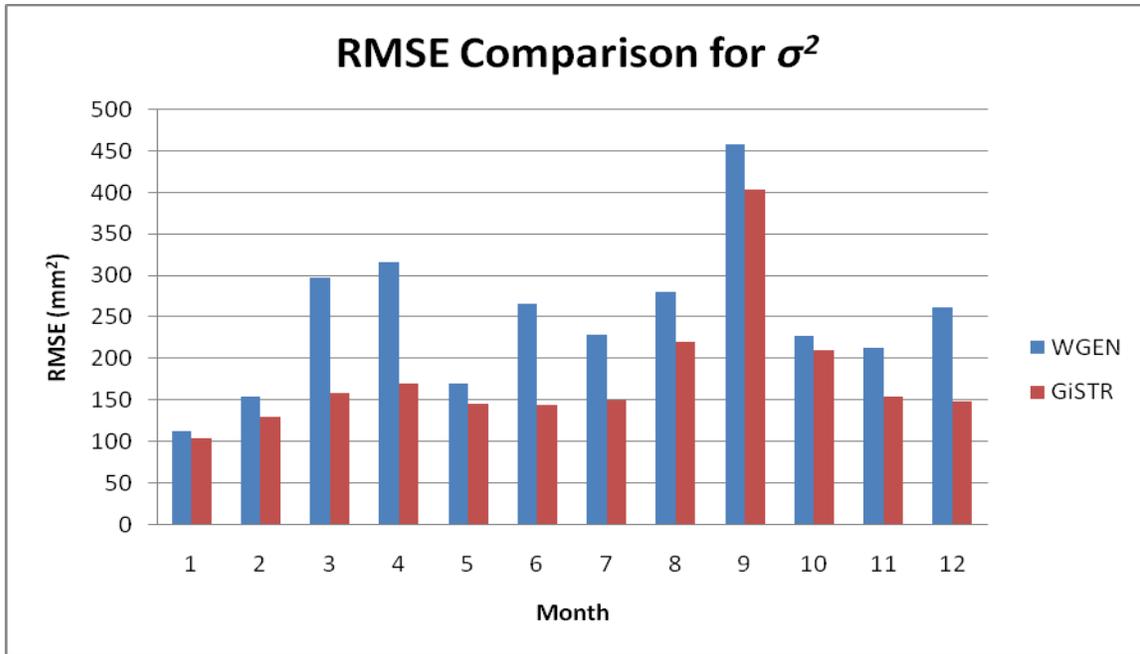


Figure 5.40. RMSE Comparison between downscaling simulations for the variance of daily nonzero precipitation (σ^2), 2001-2009 time period. Outlier stations removed.

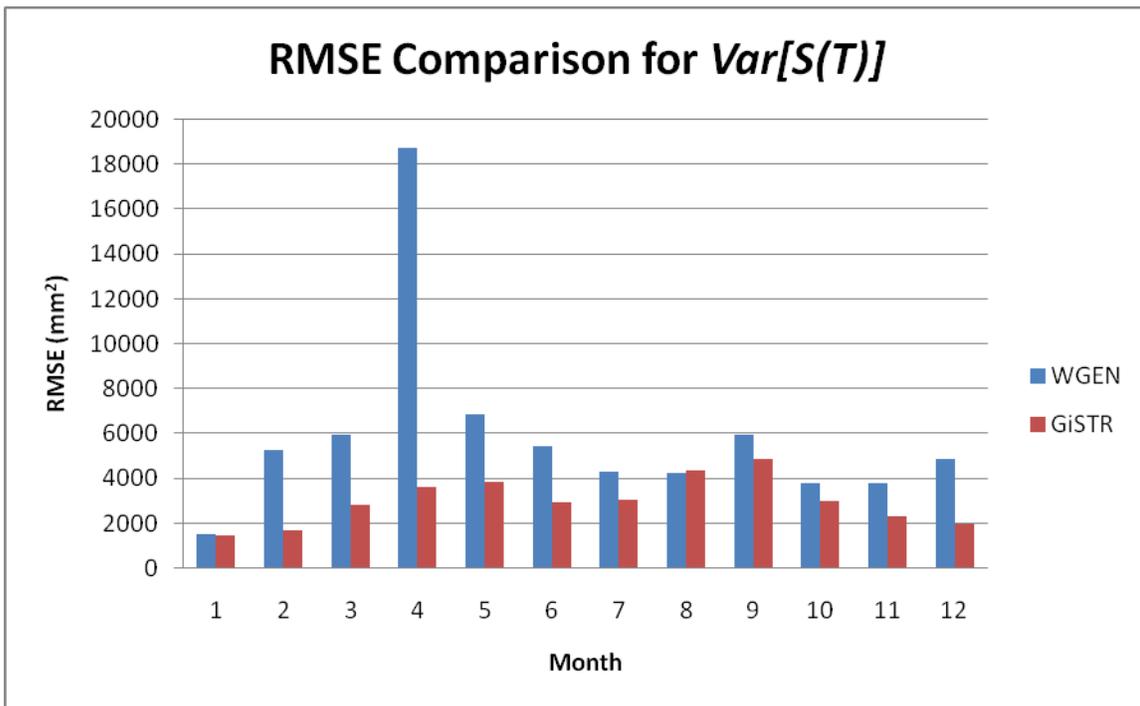


Figure 5.41. RMSE Comparison between downscaling simulations for the inter-annual variability ($Var[S(T)]$), 2001-2009 time period. Outlier stations removed.

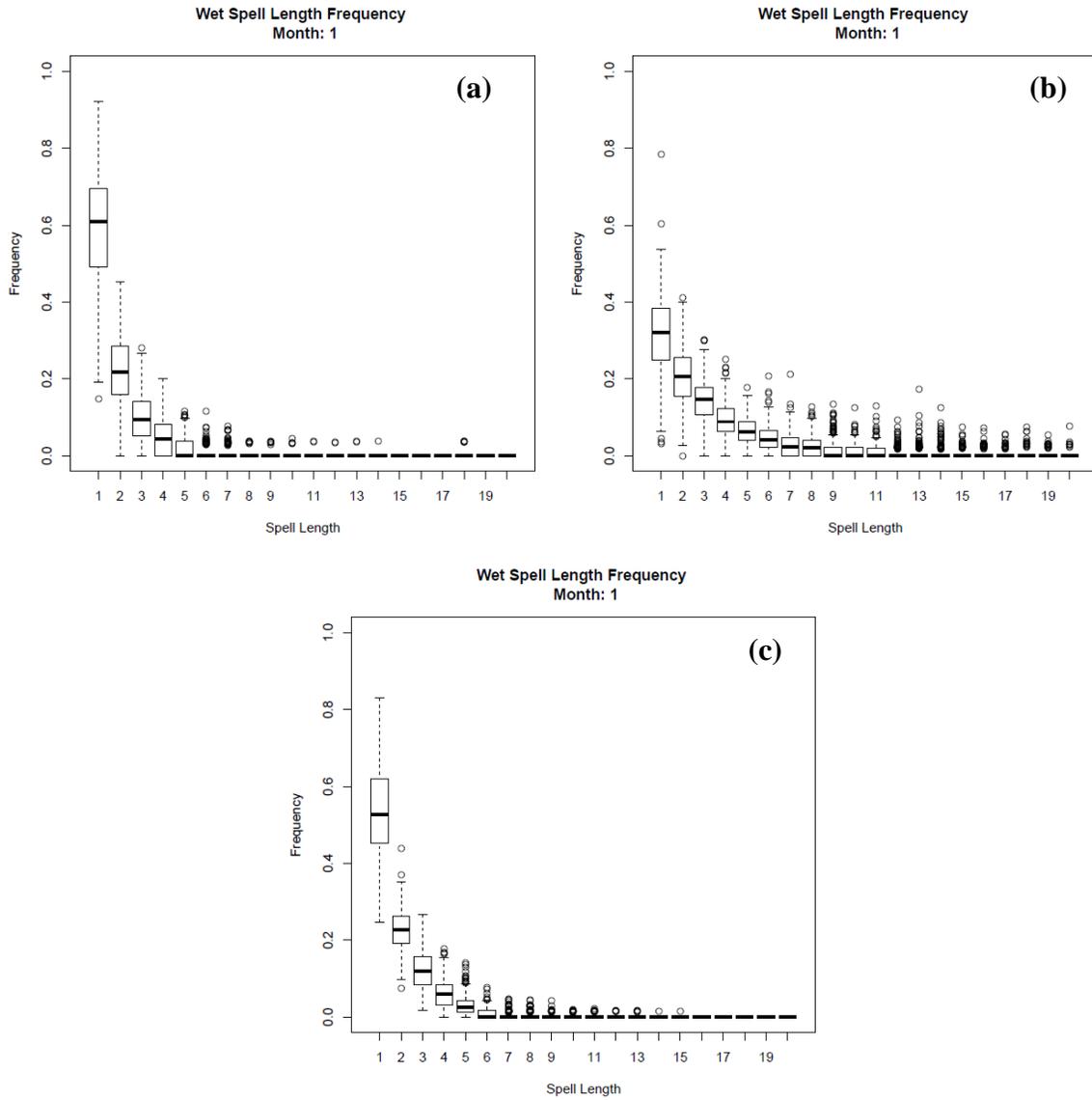


Figure 5.42. Boxplots of the frequency of wet spells of various lengths for January for observations (a), WGEN simulation (b), and GiSTR simulation (c). 2001-2009 time period.

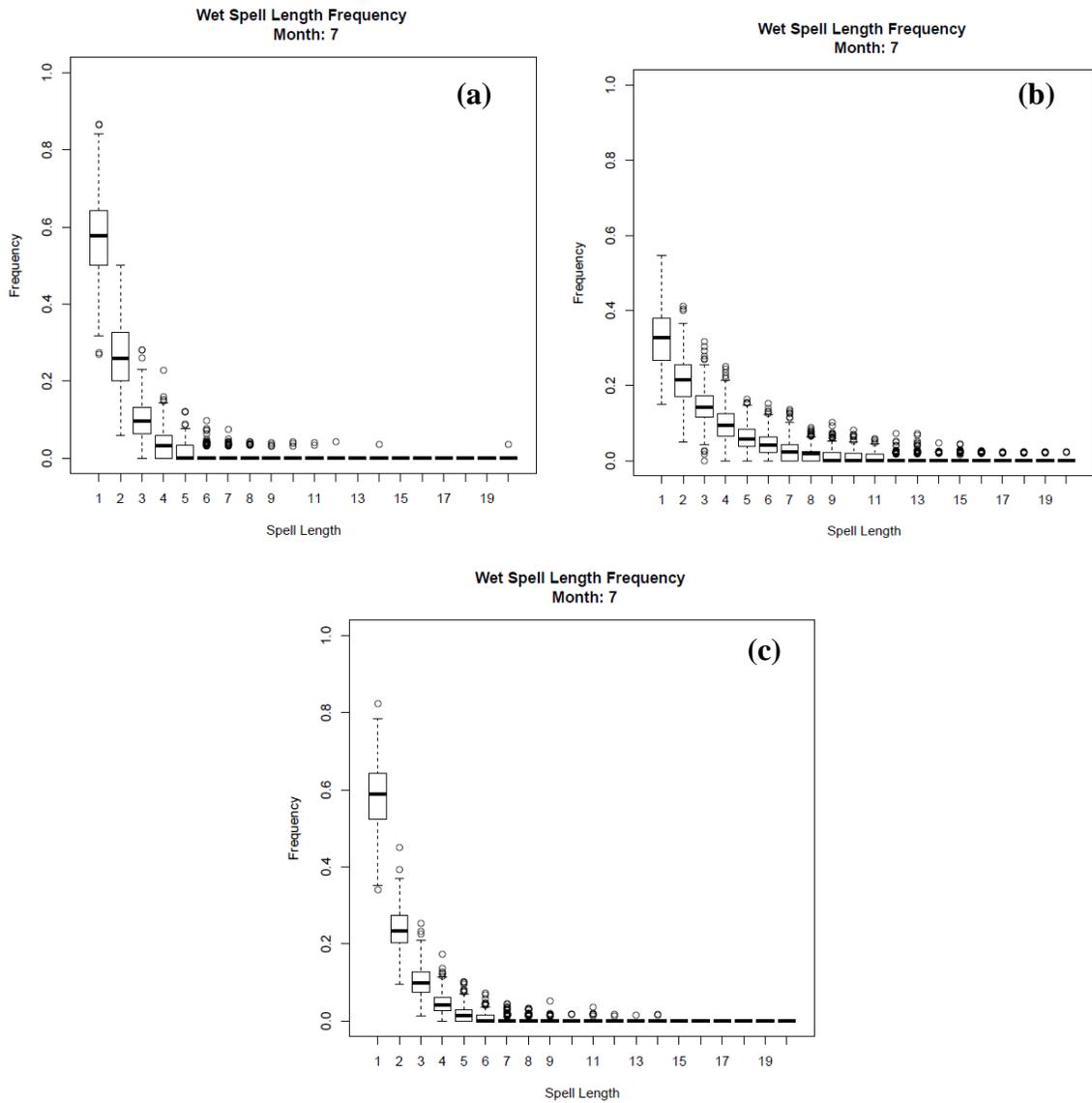


Figure 5.43. Boxplots of the frequency of wet spells of various lengths for July for observations (a), WGEN simulation (b), and GiSTR simulation (c). 2001-2009 time period.

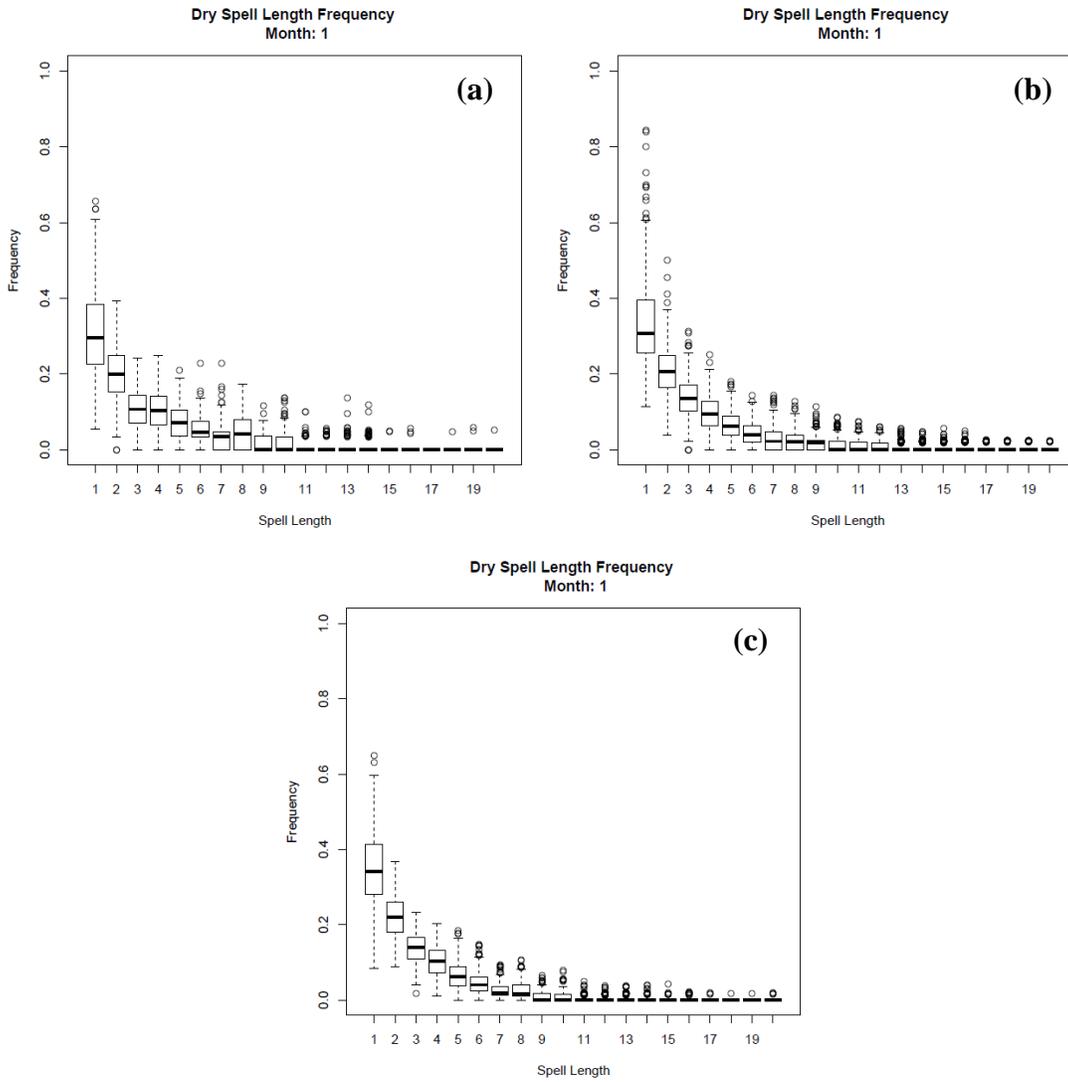


Figure 5.44. Boxplots of the frequency of dry spells of various lengths for January for observations (a), WGEN simulation (b), and GiSTR simulation (c). 2001-2009 time period.

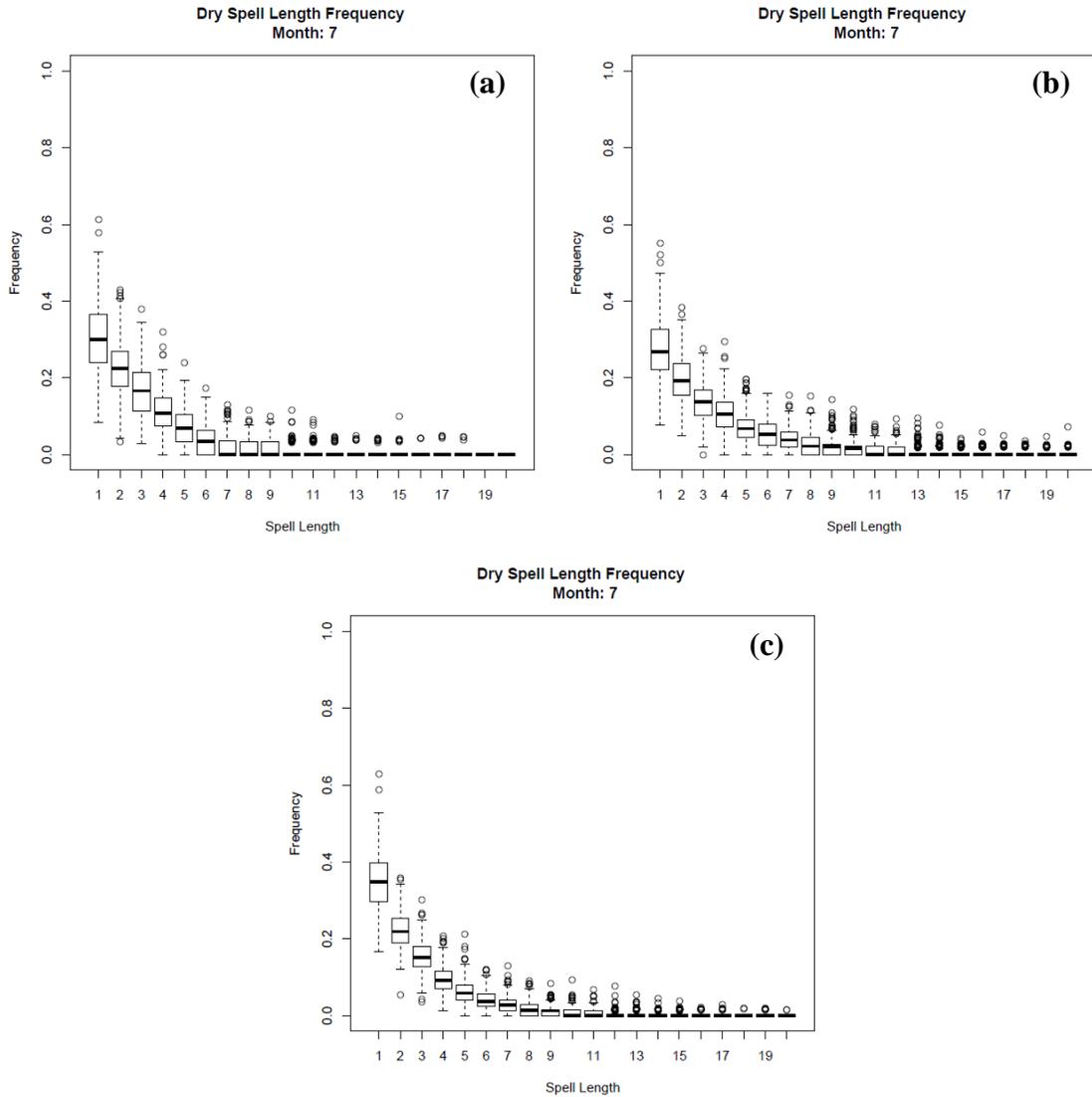


Figure 5.45. Boxplots of the frequency of dry spells of various lengths for July for observations (a), WGEN simulation (b), and GiSTR simulation (c). 2001-2009 time period.

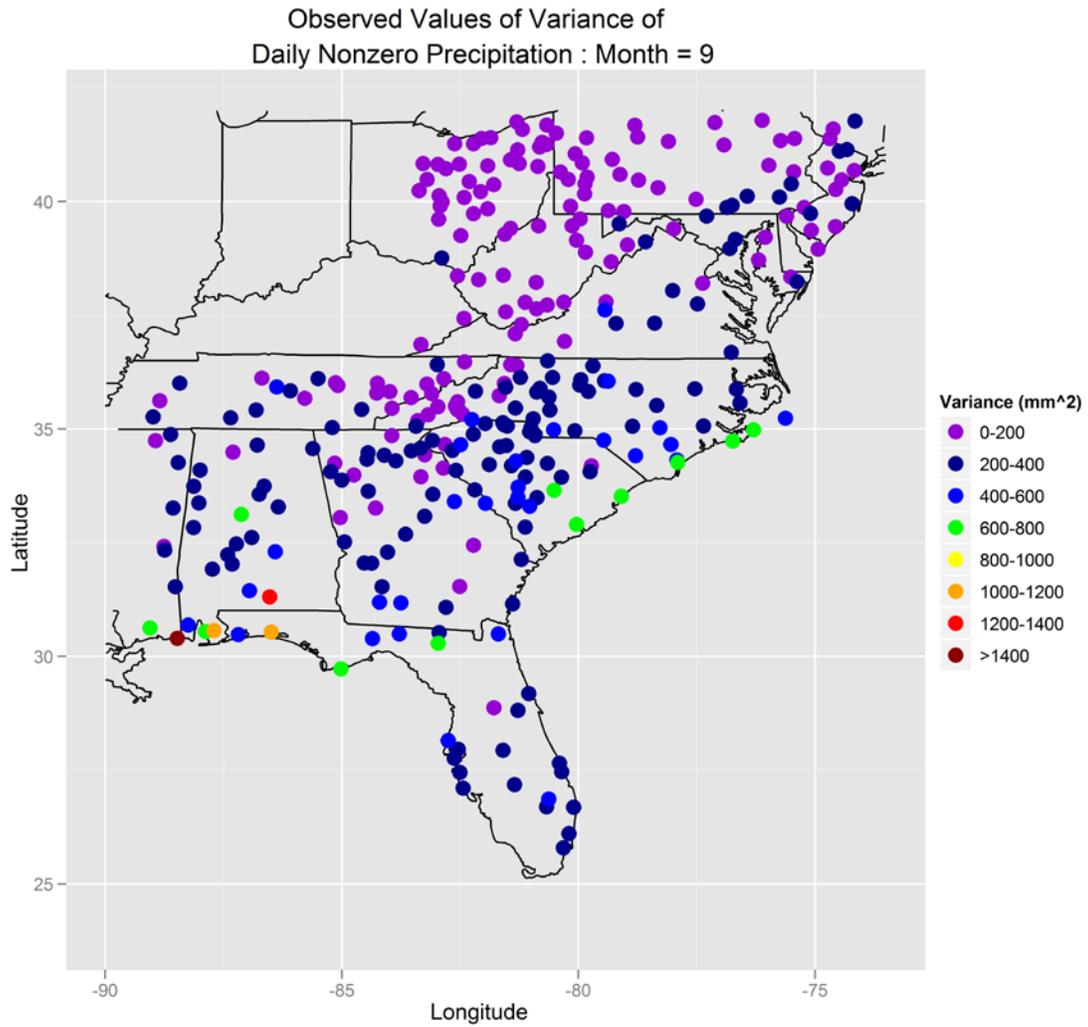


Figure 5.46. Observed values of the variance of daily nonzero precipitation (σ^2) for the 1979-2000 time period across the domain for September.

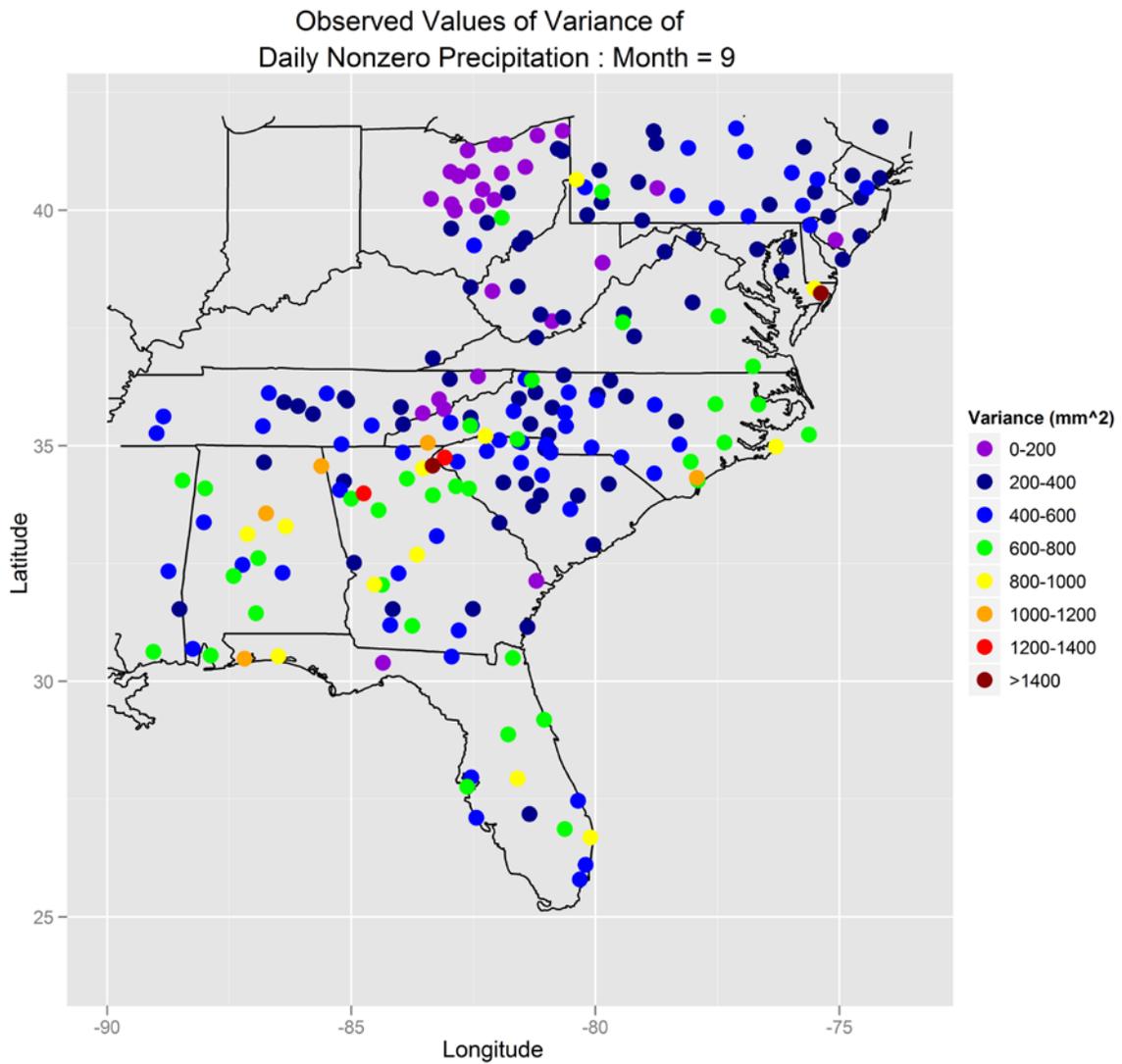


Figure 5.47. Observed values of the variance of daily nonzero precipitation (σ^2) for the 2001-2009 time period across the domain for September.

Chapter 6. Conclusions and Future Work

This study focuses on evaluating three weather generators which are used for downscaling precipitation across the Southeast U.S. The weather generators are evaluated for the ability to replicate the observed statistics of precipitation in the 1979-2000 time period. These weather generators are evaluated for two separate domains: a smaller domain covering only North Carolina, and a larger domain extending over the Southeast U.S. A single weather generator is then used to evaluate the error associated with four variations of the two assumptions in the scaling relationship used for downscaling with weather generators. This analysis was under the assumption that the error for each downscaling variation would be similar regardless of the generator used. Finally, two of the three generators, using the same variation of the downscaling assumptions, were compared to determine if the error patterns are similar to those shown when each generator is not used for downscaling. These two analyses were performed using CFSR data for the base period of 1979-2000 and a forecast period of 2001-2009. The conclusions resulting from each of these analyses are presented based on each of the main questions to be answered as part of this study. Recommendations for future work and the caveats of this research are also presented in this Chapter.

6.1. Summary and Conclusions

The previous three chapters provided full details regarding the analyses performed in this study. The major conclusions drawn from this research are presented based upon the major study questions described below (also in Chapter 1).

- What are the strengths and weaknesses of each generator in each domain?
- What seasonal/spatial trends (if any) are present in the error of each generator?
- What are the strengths and weaknesses of different approaches used to downscaling with weather generators?

- Are there seasonal/spatial trends in the error of each approach?
- What are the possible physical causes for this error? That is, what climatological phenomena may not be captured by each generator/downscaling simulation?

Each analysis in this study made use of a set of parameters which are common to each weather generator and commonly referred to in literature for downscaling with weather generators. For reference in this chapter the parameter descriptions from Chapter 2 are included in Table 6.1. This breakdown of the results by each major study question will be followed by some general conclusions from the study as a whole.

What are the strengths and weaknesses of each generator in each domain?

While there are some discrepancies regarding the error of each generator between domains, there are several patterns generally observed which are consistent between domains. For the temporal parameters reflecting the occurrence of precipitation events, WGEN has the lowest error when replicating these parameters across the domain. In contrast, GiST and GiSTR have larger error for replicating these parameters. However, while GiST and GiSTR have larger error for these parameters, the difference in error compared to WGEN is less than 0.08 on average. In addition, each generator is equally capable of reproducing dry and wet spells of various lengths across the entire domain. This suggests that although both GiST and GiSTR have a slightly higher error than WGEN for replicating these parameters the difference is negligible in practice. GiST and GiSTR are shown to have substantially lower error replicating the spatial structure. WGEN has the highest error when replicating spatial structure (by 0.09 to 0.24).

These results are consistent with the structure of each weather generator. By design, GiST and GiSTR consider both the spatial and temporal structure of events and amounts when simulating precipitation. In contrast, WGEN does not consider spatial structure during its generation process. This lack of spatial structure in either the event generation or amount

generation process leads to the increased error with regards to spatial structure. However, while GiST and GiSTR consider spatial structure, the joint consideration of both spatial and temporal structure is a potential cause for the increased error with regards to temporal parameters related to precipitation occurrence. Regardless, while WGEN is able to capture the occurrence of precipitation amounts it cannot replicate the dominant spatial structure of precipitation and how that structure changes during the year. That is, WGEN is not capable of distinguishing between different spatial structures of precipitation such as such as mid-latitude cyclones and frontal precipitation or isolated convective processes.

For parameters reflecting precipitation amounts, there are several consistent features between domains. First, GiST has the lowest error when replicating the mean daily nonzero precipitation amounts (0.46 mm for the North Carolina domain, 0.75 mm for the Southeast U.S. domain), while GiSTR has the largest error (4.83 mm for the North Carolina domain, 2.49 mm for the Southeast U.S. domain). Second, GiSTR has the lowest error (65.85 mm² for the North Carolina domain, 81.18 mm² for the Southeast U.S. domain) when replicating the variance of daily nonzero precipitation, while GiST has the largest error (122.49 mm² for the North Carolina domain, 124.08 mm² for the Southeast U.S. domain). GiSTR is shown to overestimate values of the mean daily nonzero precipitation for most stations in both domains. The impact of this error pattern is that the GiST generator is more able to capture normal precipitation amounts than both GiSTR and WGEN. However, the variance of daily nonzero precipitation is representative of the variability associated with extreme precipitation events, both extreme wet and dry spells. Therefore the impact of this shows that while GiST is capable of capturing normal precipitation amounts, GiST does not capture the variability associated with extreme precipitation events, including droughts and hurricanes. Recall that for the generation of precipitation amounts the only difference between each generator is that GiSTR generates extreme precipitation events with an extra gamma distribution. This extra gamma distribution re-samples the extreme events causing them to occur more frequently. The overestimation of the mean daily nonzero precipitation by GiSTR reflects both the increased sampling associated with the extra gamma distribution and the overlap of this extra

gamma distribution with the regular gamma distribution which reflects smaller precipitation amounts.

While the process used by GiSTR to re-sample precipitation extremes causes the mean daily nonzero precipitation to be overestimated, it also improves the ability of GiSTR to replicate the variance of the daily nonzero precipitation. While both GiST and WGEN make use of a single gamma distribution to represent all precipitation amounts, GiST was shown to have a significantly smaller error for the mean daily nonzero precipitation and larger error for the variance of the daily nonzero precipitation. WGEN slightly overestimates the mean daily nonzero precipitation across for most stations, and underestimates the variance of daily precipitation for coastal stations. GiST also underestimates the variance of daily nonzero precipitation for the same locations as WGEN with a larger magnitude.

For the amount generation process of each weather generator, the major difference is that GiST uses the spatial correlation of precipitation amounts in that process while WGEN does not. As such, it can be concluded that considering the spatial structure improves the ability of GiST to replicate mean daily nonzero precipitation amounts, but increases the error of GiST for replicating the variance of nonzero precipitation. GiST also has the lowest error (between 8.44 and 11.64 mm on average) for replicating values of the average total precipitation for both domains. However, the error for each generator for the inter-annual variability is not consistent between domains. WGEN has the smallest error (923.39 mm^2 on average) for the North Carolina domain, while GiSTR has the smallest error (836.89 mm^2 on average) in the Southeast U.S. domain. Recall that values for the inter-annual variability are affected by the error of both the mean and the variance of the daily nonzero precipitation. In addition, it was also shown in Chapter 3 that GiSTR has a larger error for values of the mean daily nonzero precipitation in the North Carolina domain (4.83mm) than in the Southeast U.S. domain (2.49mm). This difference is the likely cause for the change in which generator is best for replicating the inter-annual variability. The average total precipitation and the inter-annual variability indicate the total precipitation during a period and the variability of precipitation during that period. On a monthly or seasonal time scale, these represent the average total precipitation in a month or season, and the inter-annual variability is

representative of the climatic variability during that period. Historically, the inter-annual variability has been a major source of error for most weather generators and associated in the failure to capture extreme precipitation and the variance of daily nonzero precipitation (Wilks and Wilby, 1999; Furrer and Katz, 2008). Given that GiSTR has smaller error for the inter-annual variability than WGEN and GiST, it implies that GiSTR is more capable of capturing the precipitation extremes associated with climatic variability. However, given that the inter-annual variability is affected by the mean daily nonzero precipitation and the variance of the daily nonzero precipitation, applying different distributions to improve the ability of GiSTR to capture the mean daily nonzero precipitation would likely also improve the inter-annual variability produced by GiSTR more than shown in this analysis.

What seasonal/spatial trends (if any) are present in the error of each generator?

While there are seasonal trends in the error for several of the parameters evaluated for each generator, there is no seasonal trend present for any of the temporal parameters for the occurrence of precipitation events for either generator. For the spatial structure, WGEN has a tendency to better replicate the spatial structure of precipitation in summer months. The dominant spatial structure for precipitation during June, July and August is unstructured convection. From a purely statistical point of view, this translates into a more random spatial structure than the stratiform precipitation which dominant in winter months. That is, there is a weaker relationship between locations and that correlation between locations decreases more rapidly in summer months than in winter months. The weaker relationship between locations also WGEN, which does not consider spatial structure, to be more accurate in capturing this weak relationship in summer months than during other months. However, this improvement in summer still shows larger error (by up to 0.11 in summer, by up to 0.24 over the entire year) than GiST and GiSTR. This is indicative of the fact that while unstructured convection is a dominant form of precipitation in summer months, structured convection or large scale systems still contributes to precipitation. Therefore, unstructured convection does cause the spatial structure to be more random, but not entirely random. For the mean daily

nonzero precipitation, the peak error for each generator is in July through October. In addition, for the variance of the daily nonzero precipitation, each generator has two peaks for the error, in April and in September and October. The peak error in April may be related to the transition between the dominant structures of precipitation. In this transition between stratiform precipitation and unstructured convection during April there is a possibility that both unstructured convection and convection embedded in synoptic scale structures. This suggests an increase in the occurrence of extreme precipitation events from the combination of unstructured and embedded convection which causes this secondary peak in the error for the variance of the daily nonzero precipitation. Given that the error of this parameter in March and May also larger than January, February and June, it is also possible that the transition between dominant structures also has an impact on the error of each generator. That is, there may be more unstructured convection in May and less stratiform precipitation, and in March there may be more stratiform precipitation and less convection, while April has an equal amount of both types of precipitation. This transition may also impact the error of the variance of daily nonzero precipitation in August through October and be augmented by the effect of tropical precipitation during these months.

For the mean daily nonzero precipitation and the variance of the daily nonzero precipitation, each generator shows distinct spatial patterns and each domain. In the North Carolina domain, WGEN and GiST have similar error for the mean daily nonzero precipitation across the domain for all months. Aside from overestimating values of the mean daily nonzero precipitation for most stations, GiSTR has a general tendency to overestimate by larger amounts for coastal stations during July and August. Each generator also has a tendency to underestimate the variance of the daily nonzero precipitation more for coastal locations in August, September and October. The tendency of GiSTR to overestimate the mean daily nonzero precipitation and underestimate the variance across much of the domain is likely related to the capping value which is used by both GiST and GiSTR (Baigorria, 2010, personal communication). The re-sampling approach for extreme precipitation in GiSTR causes it to hit the capping value more frequently than GiST does in the same situation. Given that more extreme precipitation can happen in convective and

tropical processes, and that unstructured convective processes dominate in summer, and the peak of the Atlantic hurricane season is in September, GiSTR will produce amounts at this capping value more frequently than for other months. This increased frequency of precipitation at this capping value causes the mean daily nonzero precipitation to be overestimated throughout the domain, but also keeps values larger than the capping value from occurring, which causes the variance of daily nonzero precipitation to still be underestimated by GiSTR. The overestimation of the mean daily nonzero precipitation and underestimation of the variance of daily nonzero precipitation at coastal stations by larger amounts than elsewhere in the domain for July through October is also related to tropical precipitation. Precipitation from hurricanes have the largest impact on coastal stations given that these locations are subject to the most intense precipitation from landfall and from hurricanes which pass close to the coast. As such, the extreme events which are not captured by any of the three generators occur more frequently, causing the variance of daily nonzero precipitation for these coastal stations to underestimate more than other locations in the domain. While WGEN and GiST will capture the mean daily nonzero precipitation for these months, the re-sampling mechanism in GiSTR will cause it to hit the capping value more frequently than for inland stations. This in turn causes GiSTR to overestimate the mean daily precipitation by a larger magnitude for coastal stations than inland stations.

For the Southeast U.S. domain, WGEN and GiST have similar error across the domain for all months. However, aside from the tendency to overestimate the mean daily nonzero precipitation for all stations and months, GiSTR also has a tendency to overestimate this parameter by larger values in the summer months in Florida. While the error of each generator replicating the variance of the daily nonzero precipitation has a seasonal trend, there is also a spatial trend. Each generator underestimates the variance of the nonzero precipitation, but the largest magnitude for this underestimation is during September and October in the northern Gulf Coast and in southern Alabama. In addition, GiSTR also has a tendency to overestimate the variance of daily nonzero precipitation for summer months in Florida. Each generator has the largest error for the average total precipitation in August through October in both domains. Finally, each generator has the largest error for the inter-

annual variability in August and September in both domains. Between domains, GiST and WGEN show common trends in error for most parameters. However, GiSTR has a tendency to better replicate the mean daily nonzero precipitation and the average total precipitation in the Southeast domain, which is likely the result of the increased number of stations in the Southeast U.S. domain where values of both parameters are accurately reproduced by GiSTR.

While much of the seasonal and spatial error patterns are similar between domains, the tendency of GiSTR to overestimate both the mean daily nonzero precipitation and the variance of the daily nonzero precipitation in Florida during summer months may be related to several factors. In summer, convective and tropical processes, including sea breeze and unstructured convection, play a role in precipitation occurring in the state. Given that Florida has the largest number of thunderstorms compared to other areas of the Southeast U.S. These convective processes can produce extreme events, but the observed PDFs also show that small precipitation events occur more frequently than extreme events in summer. Therefore, the increased number of precipitation events in Florida lowers the observed variance of daily nonzero precipitation and mean daily nonzero precipitation. This possible cause, combined with the re-sampling done by GiSTR, may cause this generator to overestimate both the mean daily nonzero precipitation and the variance of the daily nonzero precipitation in Florida when the state receives most of its convective activity in summer. However, the affect of this combination of factors is speculation and further research is required to determine the exact cause of these errors in GiSTR.

What are the strengths and weaknesses of different approaches used to downscaling with weather generators?

The four variations of the two the downscaling assumptions have been used with GiSTR to downscale precipitation data from CFSR to the southeast U.S. These four variations used include the following:

- LWR-MOM – LWR for parameter interpolation to GCM grid points, and the MOM for the gamma distribution parameters
- AW-DFIT – AW for parameter interpolation to GCM grid points, and the DFIT for the gamma distribution parameters.
- LWR-DFIT – LWR for parameter interpolation to GCM grid points, and the DFIT for the gamma distribution parameters.
- AW-MOM – AW for parameter interpolation to GCM grid points, and the MOM for the gamma distribution parameters.

LWR refers to locally weighted regression (Wilks, 2008), AW refers to area weighted averaging, MOM refers to the method of moments estimation of gamma distribution parameter, and DFIT refers to the Greenwood and Durand (1960; Wilks, 2006) approximation for the gamma distribution parameters. Each downscaling simulation, regardless of time period increases the error of all parameters by some amount, including the spatial structure, which is not directly affected by the CFSR data. However, regardless of the downscaling simulation the spatial structure of precipitation is not significantly affected by the CFSR data and the error is not significantly different between downscaling simulations. There is no difference in error for the temporal parameters of precipitation occurrence between these downscaling simulations in either period. This is also reflected by the dry and wet spell frequencies produced by each downscaling simulation.

However, while there is no significant difference between simulations for the error of the spatial structure of precipitation or the temporal parameters of precipitation occurrence, there are significant differences between simulations for the mean daily nonzero precipitation and the average total precipitation. Table 6.2 shows the average RMSE for values of the mean daily nonzero precipitation and the average total precipitation in both time periods for each simulation. For the mean daily nonzero precipitation and the average total precipitation, AW-DFIT has the smallest error in both time periods and LWR-MOM has the largest error. In addition, for these two parameters, LWR-MOM and LWR-DFIT (which use LWR) have larger error than AW-DFIT and AW-MOM. This indicates that the parameter

interpolation using LWR increases error for the mean daily nonzero precipitation and the average total precipitation.

For the variance of the daily nonzero precipitation, LWR-MOM and AW-MOM have the smaller error (127 mm^2 on average for both) in the 1979-2000 period than the remaining simulations. The AW-MOM has the smallest error (177 mm^2 on average) and the LWR-MOM has error less than 4% larger on average (183 mm^2 on average) in the 2001-2009 time period. For the inter-annual variability, AW-MOM has the smallest error for the 1979-2000 time period and AW-DFIT has the smallest error for the 2001-2009 time period. However, while the difference in error between downscaling simulations for the mean daily nonzero precipitation and average total precipitation shows significance, the difference in error for the variance of daily nonzero precipitation and the inter-annual variability between downscaling simulations is not significant in most cases for both time periods.

While there is no significant difference in error between simulations for the variance of daily nonzero precipitation or the inter-annual variability, the difference between simulations with regards to the mean daily nonzero precipitation and the average total precipitation may contribute to the subtle differences between time periods. While the LWR-MOM and the AW-MOM have comparable error for the variance of daily nonzero precipitation in both periods, the LWR interpolation makes use of the station elevation and GCM grid point elevation. Therefore, the coarse land surface elevation in the CFSR may have contributed to the slight increase in error between time periods for this value. In addition, a shorter time period can cause observed values of the variance to be larger. This fact coupled with the potential error associated with coarse land surface resolution, may contribute to the slight difference between AW-MOM and LWR-MOM in each time period. Given that the only difference between the AW-MOM and AW-DFIT is the method used to estimate the parameters of the gamma distribution, the DFIT technique may have produced slightly higher error for the inter-annual variability in 1979-2000 through error in multiple parameters. Regardless of the exact causes, the error for the variance of daily nonzero precipitation and inter-annual variability is not significant between simulations. In each time

period, the results indicated that the commonly used AW-MOM variation has higher error for most parameters compared to the remaining three downscaling simulations.

Given that only the error for mean daily nonzero precipitation and the average total precipitation show significant difference, the best downscaling variation is the AW-DFIT variation. However, if the error for the variance of nonzero precipitation and the inter-annual variability is also considered, then the best choices for the downscaling variation are the LWR-MOM variation and the AW-DFIT variation, since these variations have the lower error for most parameters across the domain.

Aside from holding the weather generator used constant and comparing all variations, the downscaling variation used was fixed and the WGEN and GiSTR weather generators were evaluated in a downscaling context. The results from this analysis in both time periods indicates that when used for downscaling GiSTR has lower error for most parameters than WGEN, and better replicates the observed dry and wet spell frequencies in the domain. This is contradictory to the results shown in Chapter 3, where WGEN had lower error for μ and $E[S(T)]$. The increased error from WGEN in both time periods was caused by errors that are much larger than the average error specifically in April and September. However, removing these outlier stations did not substantially change the results in either time period. This suggests that WGEN may be more sensitive to the choice of downscaling variation than GiSTR, and that the use of a different variation might produce different results, particularly with regards to the outlier stations where the LWR-MOM variation may have produced erroneous inputs for both GiSTR and WGEN.

Are there seasonal/spatial trends in the error of each approach?

Each downscaling variation shows specific trends for the error of the mean daily nonzero precipitation for both time periods. First, the peak error for each simulation in each time period is most often during the summer months or in April. Second, the peak error for each simulation for the average total precipitation is also in the summer months for each time period. This seasonal pattern is also reflected in the spatial pattern of the error of each

downscaling simulation. During the summer months, each downscaling simulation overestimates the mean daily nonzero precipitation in Florida for the 1979-2000 time period. In addition, during the same time period the LWR-MOM and LWR-DFIT simulations both have a tendency to overestimate the mean daily nonzero precipitation in the northeast corner of the domain, while having similar error to the remaining simulations for the rest of the domain. In this time period, this is caused by the failure of the LWR-DFIT and LWR-MOM downscaling simulations to capture the probability distribution function (PDF) of the daily nonzero precipitation in this region.

The 2001-2009 forecast time period shows similar seasonal patterns in the error of the mean daily nonzero precipitation and similar sources of error. While the patterns are similar, each simulation shows a tendency to underestimate the mean daily nonzero precipitation in the Carolinas during this time period, with the largest magnitude in the southern Appalachians. This is in contrast to the errors shown by the 1979-2000 simulation, which has similar underestimates along the coast of the Carolinas. In addition it is also shown that the mean daily nonzero precipitation and the variance of the daily nonzero precipitation produced by each simulation in the 2001-2009 time period in Florida is more accurate than in the 1979-2000 time period for summer months. While several of these errors are related to the same physical reasons that are sources of error for GiSTR (the tropical influence in summer and the transition between dominant structures in April), there are several sources of error which may be related to the downscaling variations and the CFSR data itself.

The downscaling simulations using LWR overestimate the mean daily nonzero precipitation in specific regions for the 1979-2000 time period, and individual stations in small regions for the 2001-2009 time period. There are several possible causes for the errors with these downscaling simulations. The LWR interpolation approach makes use of the elevation at the each station as well as the elevation at each CFSR grid point. The two degree resolution of the CFSR model suggests that coarse land surface elevation has an impact on the results from the LWR interpolation of present day station data to the CFSR grid points. Another possible explanation is an edge effect, particularly for those stations and CFSR grid points on the edge of the domain. No stations are included beyond the edges of

the domain in this analysis, which could also have an impact on the results produced by the LWR downscaling simulations. However, further analysis should be done to determine the exact cause of this increased error for the LWR downscaling simulations.

For the variance of daily nonzero precipitation and the inter-annual variability, the peak error for each downscaling simulation from July through October in each time period. While this seasonal pattern is consistent between the base and forecast time periods, the sources of error are different between domains. For the 1979-2000 period the sources of highest error are in the northern Gulf Coast, southern Alabama (both underestimates), and in Florida (overestimates). For the 2001-2009 time period, the sources of highest error include southern Alabama, the southern Appalachians, and eastern Maryland (all underestimates). However, it is important to note that each downscaling simulation had similar error patterns in both time periods, which is consistent with the lack of significant difference between downscaling simulations for the error of the variance of nonzero precipitation and the inter-annual variability.

In addition, it is noted that the variance of daily nonzero precipitation for August and September is overestimated by GiSTR and the 1979-2000 downscaling simulations in Florida, but underestimated by the 2001-2009 downscaling simulations. This period is subject to large variability in much of Florida from several events, including the 2004 and 2005 Atlantic hurricane seasons and the drought during 2007. Figure 6.1a shows that the variance of daily nonzero precipitation in during the 1979-2000 time period, while the variance of the daily nonzero precipitation in 2001-2009 is shown in Figure 6.1b in September. However, for Florida there is no period during the 1979-2000 period where the variance of daily nonzero precipitation is comparable to values in 2001-2009, shown in Figure 6.2 also for September. However, there are other portions of the domain where the value of the variance of daily nonzero precipitation is comparable between time periods. Comparing 1991-2000 and 2001-2009 there are several areas along the Gulf Coast and East Coast where values of the variance of daily nonzero precipitation is comparable. However, it is also important to note that in northern and western areas of the domain the variance of daily nonzero precipitation is larger in the 2001-2009 time period compared to the 1979-2000

time period. This increase cannot be matched by the variance of daily nonzero precipitation from different ten year time periods within the 1979-2000 time period. The difference in the observed values of the variance of daily nonzero precipitation between time periods may be related to the size of the time periods. The value of the variance of daily nonzero precipitation is also subject to the size of the time period. Smaller time periods can cause the variance to increase compared to a larger time period with the same percentage of extreme events. In addition, tropical precipitation from the 2004 and 2005 hurricane seasons did track into the northern and western portions of the study domain. This may have caused the increase in the observed variance of nonzero precipitation between time periods in these areas of the domain. However, the increase in the northern and western portions of the domain is less than in Florida because of the decrease in the strength of a hurricane as it travels inland from the Gulf Coast and East Coast. Regardless, the change in error of the variance of the daily nonzero precipitation for the downscaling simulations with GiSTR between the 1979-2000 and 2001-2009 is likely the result of a set of extreme events which affected Florida in 2001-2009 and cause the variance to increase to the point where it did not appear to be matched in by any ten year period in the 1979-2000 period. This could contribute to the subsequent decrease in the error for the variance of daily nonzero precipitation in Florida for all the 2001-2009 downscaling simulations. During this time period, the mean daily nonzero precipitation would also be larger in Florida, primarily from the land falling hurricanes in Florida during this period. Given how GiSTR generates precipitation amounts, it can be concluded that GiSTR used in downscaling, regardless of the variation used, improves the simulation of precipitation in Florida for periods with above average hurricane seasons. However, to determine if GiSTR in downscaling is best for periods with active hurricane seasons this analysis should be repeated using all three weather generators and all four downscaling variations.

Additional Conclusions and Caveats

In addition to the conclusions and answers to the main questions of this study, there are several conclusions and caveats which are also important to note. In several chapters it is noted that GiST has a capping value in order to prevent unrealistic values of precipitation from occurring (Baigorria, 2010, personal communication). Since the focus of this study is on seasonal or near-term precipitation estimates, the capping value has been retained for both GiST and GiSTR. Given that GiSTR re-samples heavy precipitation the capping value for each station is hit more frequently, causing the GiSTR to produce unrealistic PDFs for the observed precipitation at several stations. Adjusting or removing the capping value when using GiSTR might allow for more realistic PDFs to be produced potentially improving the error for precipitation amounts.

It is also noted that the spatial structure is not directly influenced by the CFSR data. This has been done since CFSR cannot resolve the differences between individual stations, and following common practice for downscaling with weather generators. However, since GiSTR makes use of the spatial and temporal structure as it generates precipitation the influence of the CFSR data has been shown to influence the spatial structure output by GiSTR. While this does have an impact, the error for the spatial structure between GiSTR and the GiSTR downscaling simulations was not significantly different. However, given that this study focuses on near term downscaling, this conclusion may not be valid for climate change downscaling studies.

6.2. Recommended Usage

There are several potential uses and applications for downscaling with weather generators, but what is recommended for use does depend on the ultimate application of downscaling in question. Given the ability of GiST to replicate the dominant spatial structure, frequencies of dry and wet spells and the average total precipitation for a given time period, GiST is recommended over WGEN and GiSTR for use in situations involving

generating synthetic weather data as replacements for missing or poor quality data or for seasonal forecasting of precipitation in winter months or when the variability is not expected to be strongly influenced by climatic variability. However, in situations where climatic variability or changes in climatic variability are concerns, GiSTR is recommended since it is better able to capture this variability. These situations include forecasts of drought and extreme precipitation associated with above normal hurricane season activity.

For downscaling, there are separate recommendations for the variations described in this study. The error of each variation is similar for the temporal and spatial structure parameters. However, given the tendency of the LWR-MOM and AW-DFIT to have lower error for most parameters the other variations, the LWR-MOM or AW-DFIT variations are recommended for use in most situations. As with the comparison between generators LWR-MOM and AW-DFIT have similar error for most parameters. However, the LWR-MOM better replicates the variability of daily precipitation than the AW-DFIT. In addition, the AW-DFIT better replicates the mean daily nonzero precipitation than the LWR-MOM. Therefore, AW-DFIT is recommended for use when the variability of daily precipitation is not a concern in the application, and LWR-MOM is recommended for when daily variability is a concern.

An interesting note regarding this recommendation is that each simulation has one new assumption and one commonly used assumption. Both of these have smaller error for most parameters than the traditionally used AW-MOM, which shows that each the AW-DFIT and LWR-MOM improve upon the commonly used AW-MOM. However, the LWR-DFIT has higher error than the LWR-MOM and AW-DFIT while making use of both new assumptions. It is possible that using the DFIT with the LWR interpolation causes any errors in the LWR interpolation to be exaggerated, causing the error to be larger for most parameters than LWR-MOM and AW-DFIT. Therefore, it is likely that the one new assumption compensates for the potential errors in the old and more commonly used assumption, which allows both LWR-MOM and AW-DFIT to have the smallest error for most parameters. While these variations are recommended, it is important to note that the

error in downscaling in any situation is also subject to the weather generator chosen. For instance, at a given location one variation may provide better estimates than others, such as for the outlier stations described in this study. As such, the end use of the downscaling must also be considered to determine the appropriate generator to use of the generators used in this study.

6.3. Potential Applications

Historically, weather generators have been used as inputs for crop modeling for the determination of crop yields in a region (Wilks and Wilby, 1999). In recent years, the focus has shifted to applications of weather generators in seasonal forecasting. This includes daily realizations of seasonal forecasts (Wilks, 2002), downscaled probabilistic guidance (Kim et al. 2009), seasonal forecasting for crop yields (Semenov and Doblas-Reyes, 2007), and forecasts for stream flow modeling (Caraway and Rajagopalan, 2011). Each of these studies demonstrates ways that GiST and GiSTR may be applied to seasonal forecasting and those forecasts in turn could be applied to multiple fields. For the basic user, the potential probabilistic guidance developed from GiST or GiSTR could be applied to provide point probabilities of having more or less than a specific amount of precipitation in an area. Built into a user friendly web product, the ensemble members produced by either generator could be used to provide this information at seasonal and monthly time scales, potentially providing the end user with local guidance regarding a required or dangerous amount of precipitation for their interests while maintaining more realism than other weather generation techniques and operating more efficiently than most dynamic techniques. One of the advantages of all weather generators is the ability to produce a large number of ensembles in a short time period. The number of ensembles provides an equal number of scenarios which can be applied to crop modeling and hydrology modeling. Given the advantages of GiST and GiSTR demonstrated by this study, each of these generators, providing realistic precipitation and temperature scenarios can allow for producing a range of forecasts for streamflow and

crop yields. This range of forecasts can allow for determining potential “best case” and “worst case” scenarios for both crop yields and streamflow in a given forecast period.

In addition to the application to seasonal forecasting, the applied focus of weather generators has also included determining the local impacts of climate change in regions (e.g. Kilsby et al, 2007; Qian et al, 2005; Mavromatis and Jones, 1998). Given that GiST and GiSTR incorporate the spatial structure directly in their processes, the mean, variability and spatial structure could be perturbed to reflect the potential effect of climate change on the local structure and frequency of precipitation in a region. Given an ensemble from each generator, the likelihood for different conditions could be assessed compared to current time on a local scale while preserving a realistic spatial and temporal structure. Finally, the ensemble members can again provide temperature and precipitation scenarios at daily timescales under a changed climate. These scenarios can be fed to crop models and hydrology models to determine the potential impacts of climate change on local crop yields (as suggested by Apipattanavis et al, 2010), stream flow, and reservoir levels. While there is potential for both seasonal forecasting and climate change predictions for precipitation, temperature, crop yields, and stream flow through the use of these weather generators, it is important to note that while these generators and downscaling variations have been tested against each other and not yet evaluated against other techniques for these purposes, which is a recommendation for future work in this area.

6.4. Recommendations for Future Work

Section 6.1 presented the conclusions of this study based on the main study questions presented in Chapter 1, along with some additional conclusions and caveats of this research. Sections 6.2 and 6.3 provided information on the recommended usage and potential applications. The conclusions and caveats in Section 6.1 and the potential applications in Section 6.3 suggest multiple possibilities for future work for downscaling with GiST and GiSTR.

While GiSTR is shown to capture the variability associated with extreme precipitation events, it is also shown to overestimate the mean values of precipitation amounts. Therefore, a possible exploration for future work involves determining a better method for replicating extreme precipitation events in the domain. This may include using a hybrid precipitation distribution (Furrer and Katz, 2008), a similar approach to GiSTR using a mixture of distributions (Wilks and Wilby, 1999), or a single distribution different from the two parameter gamma distribution which is commonly used.

Given that the differences between GiSTR and WGEN change when downscaling is applied, an additional recommendation for future work is apply all the downscaling variations used in this study to all three generators and both time periods. This would allow for a complete picture of the differences between each generator when each downscaling variation is applied.

As mentioned several times, the CFSR data has an indirect impact on the spatial structure produced by the GiSTR weather generator, although the spatial structure is not directly perturbed. Given that both GiST and GiSTR make use of the correlations between individual stations as part of their generation processes, there will always be some impact of the GCM data on the spatial structure produced. While the downscaling simulations have no significant difference in error for the spatial structure compared to GiSTR used without downscaling applied, this is likely because the spatial structure of precipitation did change significantly between the two time periods used in this study. However, if either GiST or GiSTR are to be used for downscaling for climate change scenarios it is likely that the spatial structure of precipitation may change in a changed climate and that the parameters reflecting this information in GiST and GiSTR will need to be perturbed to reflect this. Most literature involving weather generator downscaling for climate change scenarios either does not alter the spatial structure or influences the other parameters of the weather generator based on the change in the average inter-station correlation (Wilks 1999; Osborn, 1997), or uses the correlation between individual stations to influence other parameters of the weather generator (Wilks, 2009; Mehrotra et al. 2006; Wilks, 1998). However, no investigators have considered the spatial structure as part of the basic process in the same fashion that GiST and

GiSTR consider this. That is, other studies focus on translating spatial structure through the conditioning of other parameters such as the Markov chains or the inter-annual variability, while GiST and GiSTR consider the spatial structure directly through the orthogonal Markov chain in the event generation process. Therefore, a potential source of future work in this field involves the development of a technique to infer the change in the relationship between individual stations from global climate model information, which could also be used in downscaling with GiST and GiSTR. In addition, the CFSR data was used for downscaling both time periods. Therefore, the results for these simulations in the 2001-2009 time period reflects the potential errors associated with using each of these techniques for forecasting or climate change predictions. In order to determine the full errors of using each of these techniques in a true forecast, each of these techniques and generators should be re-evaluated using the CFS Reforecast or GCM of interest in climate change predictions.

Finally, the major focus of this study is on the different variations of the two assumptions of the scaling relationship used for downscaling with weather generators. These variations consider the general approach to downscaling with weather generators. However, there are several variations of this general approach (described in Chapter 2) which blend weather generator downscaling with other types of statistical downscaling. In addition, there are several other dynamic downscaling, ESD, statistical-dynamic downscaling techniques which could be or have been applied to downscaling in the Southeast U.S. In order to determine the best downscaling technique for the Southeast U.S. the generator techniques in this study should be compared with other weather generators, other statistical techniques and dynamic techniques. While the weather generator downscaling techniques shown in this study are applicable to seasonal forecasting and potentially to downscaling climate change scenarios, each downscaling approach will have its own strengths and weaknesses for producing precipitation estimates in the Southeast U.S. These strengths and weakness should be studied and documented to provide the potential users of this climate information in the Southeast U.S. with details about the uncertainty associated with each technique under a variety of situations.

Table 6.1. Description of parameters used to evaluate the each weather generator.

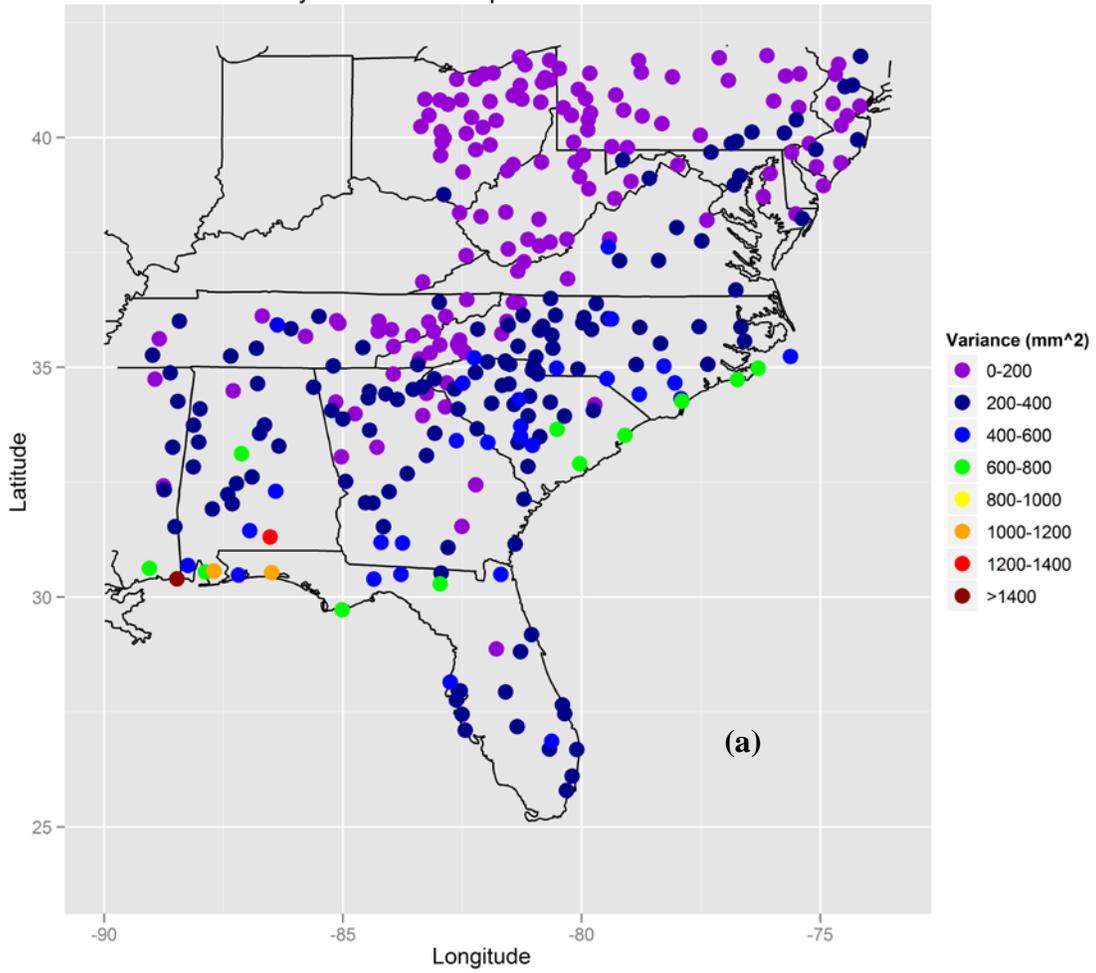
Parameter	Description
μ	Mean Daily Nonzero Precipitation
σ^2	Variance of the Daily Nonzero Precipitation
$E[S(T)]$	Average total precipitation
$Var[S(T)]$	Inter-annual variability
P_{01}, P_{11}	1 st Order Markov transition probabilities
π	Unconditional probability of rain
γ	Persistence (lag 1 autocorrelation)
ρ	Correlation matrix of precipitation
ρ_{ev}	Correlation matrix of precipitation events
ρ_{am}	Correlation matrix of precipitation amounts
ρ_{ex}	Correlation matrix of precipitation extreme events

Table 6.2. Average RMSE Comparison for values of the mean daily nonzero precipitation (μ) and the average total precipitation ($E[S(T)]$) between time periods and downscaling simulations.

Simulation	μ		$E[S(T)]$	
	1979-2000	2001-2009	1979-2000	2001-2009
LWR-MOM	5.35	6.1	65.73	79.61
AW-DFIT	3.78	3.99	44.47	53.45
LWR-DFIT	5.25	5.95	64.61	75.74
AW-MOM	4.21	4.5	52.5	62.66

Figure 6.1. Observed values of the variance of daily nonzero precipitation (σ^2) for the 1979-2000 time period (a) and the 2001-2009 time period (b).

Observed Values of Variance of
Daily Nonzero Precipitation : Month = 9



Observed Values of Variance of
Daily Nonzero Precipitation : Month = 9

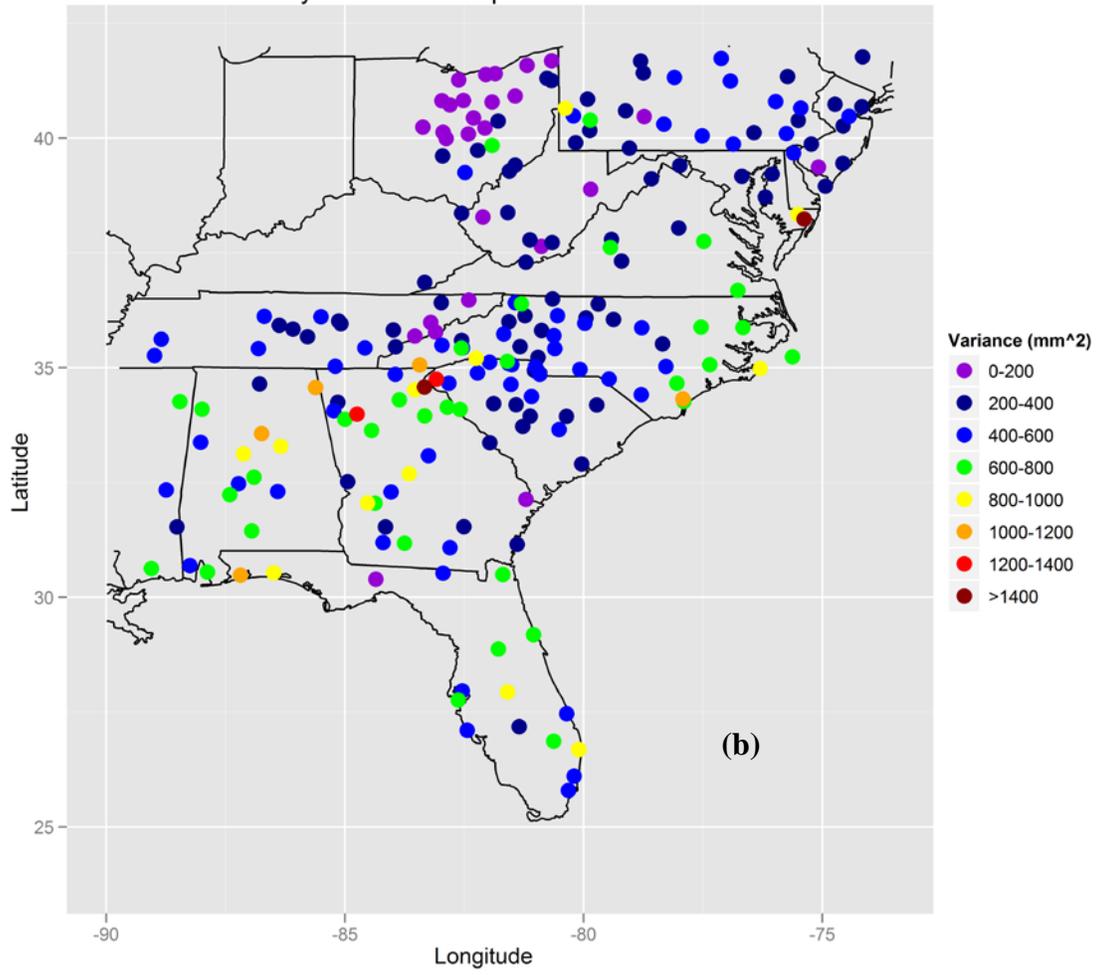
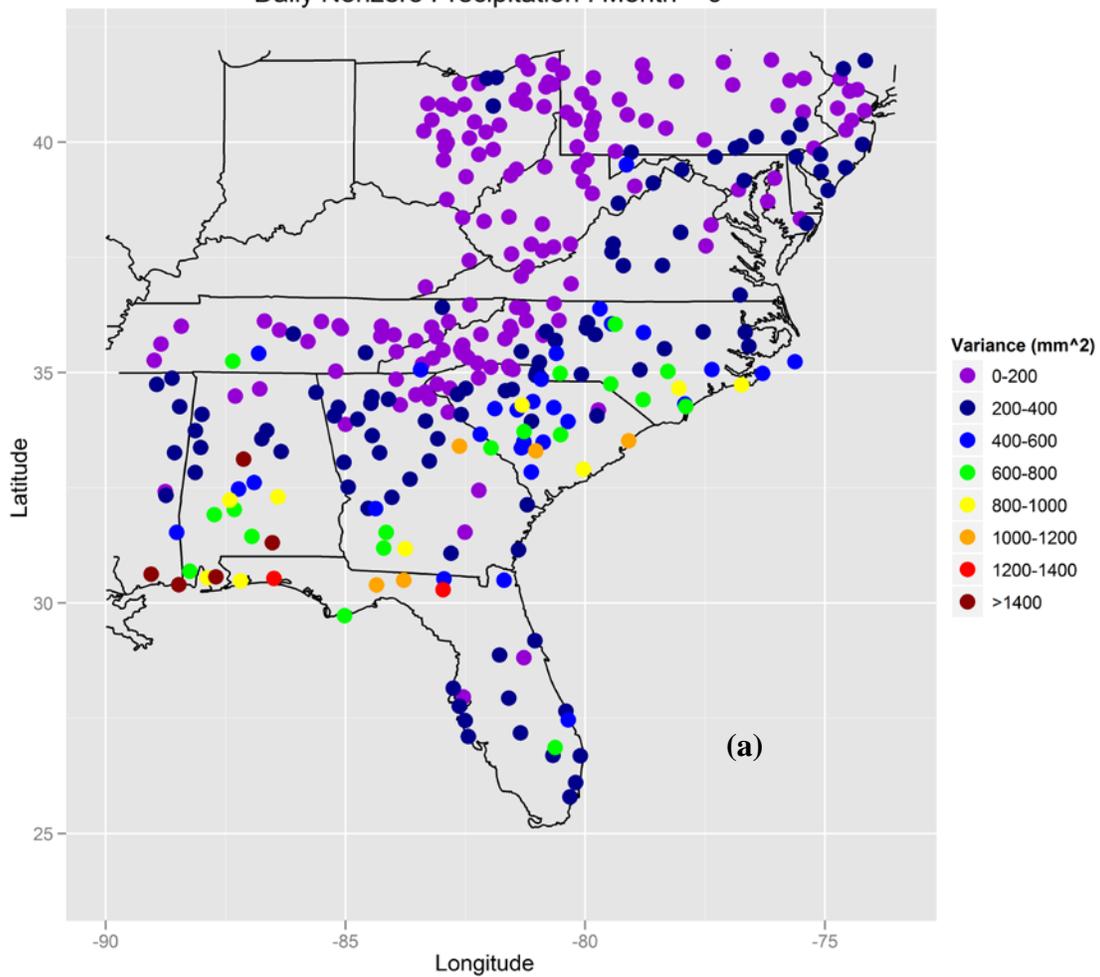
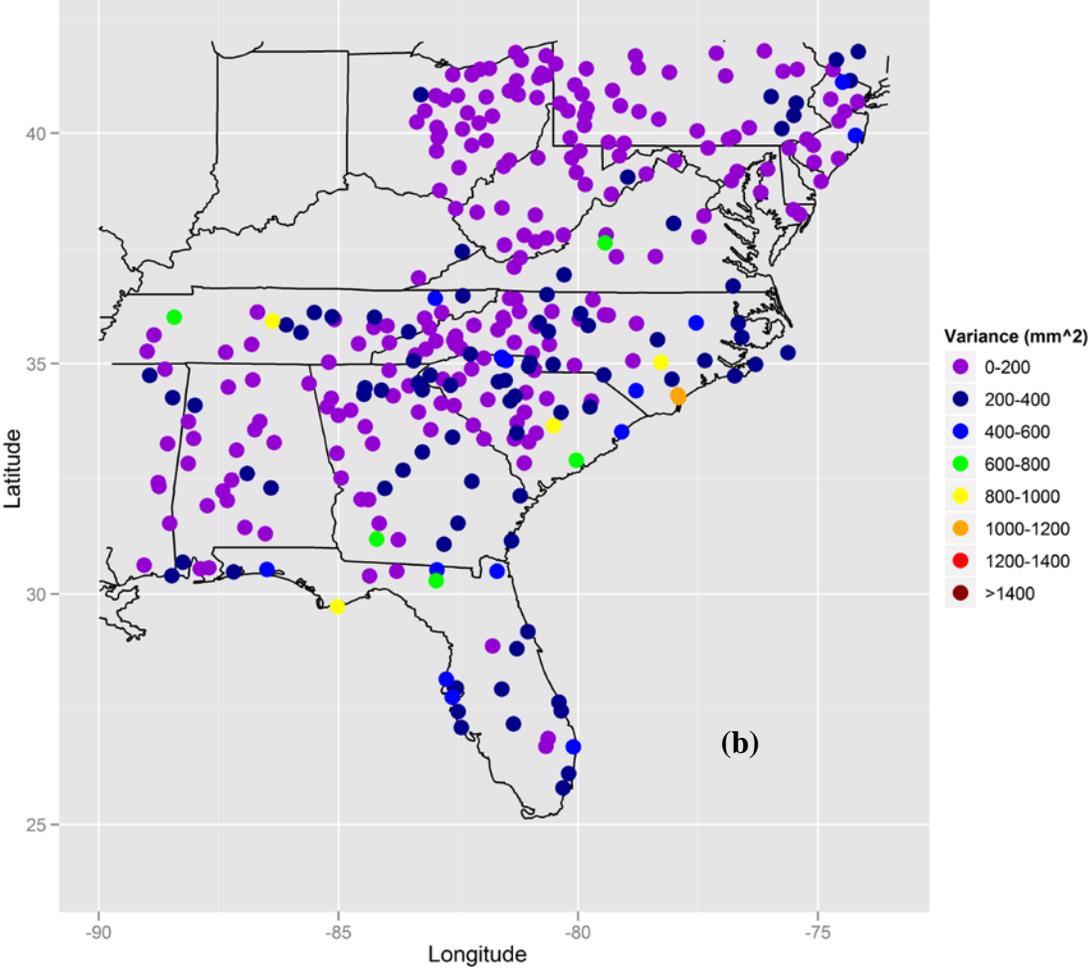


Figure 6.2. Observed values of the variance of daily nonzero precipitation (σ^2) for the 1991-2000 time period (a) and the 1981-1990 time period (b).

1991-2000 Observed Values of Variance of
Daily Nonzero Precipitation : Month = 9



1981-1990 Observed Values of Variance of
Daily Nonzero Precipitation : Month = 9



REFERENCES

Apipattanavis, S., F. Bert, G. Podestá, and B. Rajagopalan, 2010: Linking weather generators and crop models for assessment of climate forecast outcomes. *Agricultural and Forest Meteorology*, **150**(2), 166–174.

Baigorria, G. A., 2010: . Personal Communication.

Baigorria, G. A., and J. W. Jones, 2010: GiST: A Stochastic Model for Generating Spatially and Temporally Correlated Daily Rainfall Data. *J. Climate*, **23**(22), 5990–6008.

Baigorria, G. A., J. W. Jones, and J. J. O'Brien Understanding rainfall spatial variability in southeast USA at different timescales. *International Journal of Climatology*, **27**(6), 749–760.

Baigorria, G. A., J. W. Jones, D. Shin, A. Mishra, and J. J. O'Brien, 2007: Assessing uncertainties in crop model simulations using daily bias-corrected Regional Circulation Model outputs. *Climate Research*, **34**, 211–222.

Bardossy, A., and H. J. Caspary, 1990: Detection of Climate Change in Europe by Analyzing European Atmospheric Circulation Patterns from 1881 to 1989. *Theoretical and Applied Climatology*, **42**, 155–167.

Barrow, E., 2002: Downscaling: An Introduction. *Nov. 2002 Climate Scenarios Training Workshop*, Canadian Climate Impacts and Scenarios Project.

Barstad, I., A. Sorteberg, F. Flatøy, and M. Déqué, 2009: Precipitation, temperature and wind in Norway: dynamical downscaling of ERA40. *Climate Dynamics*, **33**(6), 769–776.

Benestad, R. E., 2007: Novel methods for inferring future changes in extreme rainfall over Northern Europe. *Climate Research*, **34**, 195–210.

Benestad, R. E., I. Hanssen-Bauer, and D. Chen, 2008: *Empirical-Statistical Downscaling*. World Scientific Publishing Co.

Busuioc, A., D. Chen, and C. Hellstrom Performance of statistical downscaling models in GCM validation and regional climate change estimates: application for Swedish precipitation. *International Journal of Climatology*, **21**(5), 557–578.

Caraway, N., and B. Rajagopalan, 2011: . A Multi-Site Stochastic Weather Generator for Improved Streamflow Forecast.

Chen, Y. D., X. Chen, C. Y. Xu, and Q. Shao, 2006: Downscaling of daily precipitation with a stochastic weather generator for the subtropical region in South China. *Hydrology and Earth System Sciences Discussions*, **3**, 1145–1183.

Christensen, J., M. Hulme, H. Von Storch, P. Whetton, R. Jones, L. Mearns, and C. Fu, 2001: *Regional Climate Information - Evaluation and Projections*, pp. 583–638, Cambridge University Press.

Christensen, J. H., and O. B. Christensen, 2001: *Regional Climate Simulations - A Study on Precipitation*, pp. 151–166, Danish Meteorological Institute/Danish Climate Centre.

Conway, D., and P. D. Jones, 1998: The Use of Weather Types and Air Flow Indices for GCM downscaling. *Journal of Hydrology*, **212-213**, 348–361.

Corte-Real, J., B. Qian, and H. Xu, 1999: Circulation patterns, daily precipitation in Portugal and implications for climate change simulated by the second Hadley Centre GCM. *Climate Dynamics*, **15**, 921–935.

Ding, Y., K. H. Smith, and B. Fuchs, 2008: Economic Impacts of the 2007 Drought. *Drought-Scape*, **Winter**.

Dubrovsky, M., 1997: Creating Daily Weather Series with Use of a Weather Generator. *Environmetrics*, **8**, 409–424.

Fox-Rabinovitz, M., J. Côté, B. Dugas, M. Déqué, and J. L. McGregor, 2006: Variable resolution general circulation models: Stretched-grid model intercomparison project (SGMIP). *Journal of Geophysical Research*, **111**(D16).

Fox-Rabinovitz, M. S., L. L. Takacs, R. C. Govindaraju, and M. J. Suarez, 2001: A Variable-Resolution Stretched-Grid General Circulation Model: Regional Climate Simulation. *Monthly Weather Review*, **129**, 453–469.

Frei, C., 2003: Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps. *Journal of Geophysical Research*, **108**(D3).

Frogner, I.-L., H. Haakenstad, and T. Iversen, 2006: Limited-area ensemble predictions at the Norwegian Meteorological Institute. *Quarterly Journal of the Royal Meteorological Society*, **132**(621), 2785–2808.

Furrer, E. M., and R. W. Katz, 2008: Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resources Research*, **44**(12), W12,439+.

Geng, S., F. W. T. Penning de Vries, and I. Supit, 1986: A Simple Method for Generating Daily Rainfall Data. *Agricultural and Forest Meteorology*, **36**, 363–376.

Giorgi, F., 1990: Simulation of Regional Climate Using a Limited Area Model Nested in a General Circulation Model. *Journal of Climate*, **3**, 941–963.

Goodess, C., M. Hulme, and T. Osborn, 2001: The Identification and Evaluation of Suitable Scenario Development Methods for the Estimation of Future Probabilities of Extreme Weather Events. Tyndall Centre for Climate Change Research, Tech. rep.

Greenwood, J. A., and D. Durand, 1960: Aids for Fitting the Gamma Distribution by Maximum Likelihood. *Technometrics*, **2**(1), 55–65.

Hanssen-Bauer, I., E. J. Forland, J. E. Haugen, and O. E. Tveito, 2003: Temperature and Precipitation scenarios for Norway: comparison of results from dynamical and empirical downscaling. *Climate Research*, **25**, 15–27.

Harmel, R. D., C. W. Richardson, C. L. Hanson, and G. L. Johnson, 2002: Evaluating the Adequacy of Simulating Maximum and Minimum Daily Air Temperature with the Normal Distribution. *Journal of Applied Meteorology*, **41**, 744–753.

Hashmi, M. Z., A. Y. Shamseldin, and B. W. Melville, 2009: Downscaling of future rainfall extreme events: a weather generator based approach. *The 18th World IMACS Congress and MODSIM09 International Congress on Modeling and Simulation*, The Modeling and Simulation Society of Australia and New Zealand Inc. and the International Association for Mathematics and Computers in Simulation.

Hellstrom, C., D. Chen, C. Achberger, and J. Raisanen, 2001: Comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation. *Climate Research*, **19**, 45–55.

Hewitson, B. C., and R. G. Crane Consensus between GCM climate change projections with empirical downscaling: precipitation downscaling over South Africa. *International Journal of Climatology*, **26**(10), 1315–1337.

Hewitson, B. C., and R. G. Crane, 2002: Self-organizing maps: applications to synoptic climatology. *Climate Research*, **22**, 13–26.

Heyen, H., E. Zorita, and H. Von Storch, 1996: Statistical Downscaling of Monthly Mean North Atlantic Air-Pressure to Sea Level Anomalies in the Baltic Sea. *Tellus*, **48A**, 312–323.

Hoar, T., and D. Nychka, 2008: Statistical Downscaling of the Community Climate System Model (CCSM) monthly temperature and precipitation projections.

Hughes, J. P., D. P. Lettenmaier, and P. Guttorp, 1993: A Stochastic Approach for Assessing the Effect of Changes in Synoptic Circulation Patterns on Gauge Precipitation. *Water Resources Research*, **29**(10), 3303–3315.

Huth, R., 2000: A Circulation Classification Scheme applicable in GCM studies. *Theoretical and Applied Climatology*, **67**, 1–18.

Jones, P. D., M. Hulme, and K. R. Briffa, 1993: A comparison of Lamb Circulation Types with an Objective Classification Scheme. *International Journal of Climatology*, **13**, 655–663.

Jones, P. G., and P. K. Thornton, 2000: MarkSim: Software to Generate Daily Weather Data for Latin America and Africa. *Agronomy Journal*, **92**, 445–453.

Kidson, J. W., 2000: An Analysis of New Zealand Synoptic Types and Their use in Defining Weather Regimes. *International Journal of Climatology*, **20**, 299–316.

Kidson, J. W., and C. S. Thompson, 1998: A Comparison of Statistical and Model-Based Downscaling Techniques for Estimating Local Climate Variations. *Journal of Climate*, **11**, 735–753.

Kilsby, C. G., P. D. Jones, A. Burton, A. C. Ford, H. J. Fowler, C. Harpham, P. James, A. Smith, and R. L. Wilby, 2007: A daily weather generator for use in climate change studies. *Environmental Modeling & Software*, **22**(12), 1705–1719.

Kim, Y., R. W. Katz, B. Rajagopalan, and G. Podesta, 2009: Reduced Overdispersion in Stochastic Weather Generators for Statistical Downscaling of Seasonal Forecasts and Climate Change Scenarios.

Mavromatis, T., and P. D. Jones, 1998: Comparison of climate change scenario construction methodologies for impact assessment studies. *Agricultural and Forest Meteorology*, **91**, 51–67.

May, W., and E. Roeckner, 2001: A time-slice experiment with the ECHAM4 AGCM at high resolution: the impact of horizontal resolution on annual mean climate change. *Climate Dynamics*, **17**, 407–420.

Mehrotra, R., R. Srikanthan, and A. Sharma, 2006: A comparison of three stochastic multi-site precipitation occurrence generators. *Journal of Hydrology*, **331**(1-2), 280–292.

Nicks, A. D., and G. A. Gander, 1994: Cligen: a weather generator for climate inputs to water resources and other models. *5th International Conference on Computer in Agriculture*, pp. 3–94.

Osborn, T. J. Areal and point precipitation intensity changes: Implications for the application of climate models. *Geophysical Research Letters*, **24**(22), null+.

Palmer, T. N., 1996: Predictability of the atmosphere and oceans: From days to decades. *Decadal Variability*, D. L. T. Anderson and J. Willebrand, Eds., vol. 44.

Pesquero, J., S. Chou, C. Nobre, and J. Marengo, 2010: Climate downscaling over South America for 1961-1970 using the Eta Model. *Theoretical and Applied Climatology*, **99**(1), 75–93.

Qian, B., H. Hayhoe, and S. Gameda, 2005: Evaluation of the stochastic weather generators LARS-WG and AAFC-WG for climate change impact studies. *Climate Research*, **29**, 3–21.

Qian, Y., S. J. Ghan, and L. R. Leung, 2009: Downscaling hydroclimatic changes over the Western US based on CAM subgrid scheme and WRF regional climate simulations. *International Journal of Climatology*, p. n/a.

Racsko, P., L. Szeidl, and M. Semenov, 1991: A serial approach to local stochastic weather models. *Ecological Modeling*, **57**, 27–41.

Raman, S., A. Sims, R. Ellis, and R. Boyles, 2005: Numerical Simulation of Mesoscale Circulations in a Region of Contrasting Soil Types. *Pure and Applied Geophysics*, **162**(8), 1689–1714.

Richardson, C. W., and D. A. Wright, 1984: WGEN: A Model for Generating Daily Weather Variables. U.S. Department of Agriculture, Agriculture Research Service, Tech. rep.

Rummukainen, M., J. Raisanen, B. Bringfelt, A. Ullerstig, A. Omstedt, U. Willen, U. Hansson, and C. Jones, 2001: A Regional Climate Model for Northern Europe: Model Description and Results from the Downscaling of two GCM Control Simulations. *Climate Dynamics*, **17**, 339–359.

Saha, S., et al., 2010: The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, **91**(8), 1015–1057.

Schoof, J., 2008: Application of the multivariate spectral weather generator to the contiguous United States. *Agricultural and Forest Meteorology*, **148**(3), 517–521.

Schoof, J., A. Arguez, J. Brolley, and J. Obrien, 2005: A new weather generator based on spectral properties of surface air temperatures. *Agricultural and Forest Meteorology*, **135**(1-4), 241–251.

Semenov, M. A., and E. M. Barrow, 1997: Use of a Stochastic Weather Generator in the Development of Climate Change Scenarios. *Climatic Change*, **35**, 397–414.

Semenov, M. A., and F. J. Doblas-Reyes, 2007: Utility of dynamical seasonal forecasts in predicting crop yield. *Climate Research*, **34**, 71–81.

Thom, H. C. S., 1958: A note on the gamma distribution. *Mon. Wea. Rev.*, **86**(4), 117–122.

Timbal, B., and B. J. McAvaney, 2001: An Analogue-Based Method to Downscale Surface Air Temperature: Application for Australia. *Climate Dynamics*, **17**, 947–963.

Tsvetsinskaya, E. A., L. O. Mearns, T. Mavromatis, W. Gao, L. McDaniel, and M. W. Downton, 2003: The Effect of Spatial Scale of Climatic Change Scenarios on Simulated Maize, Winter Wheat, and Rice Production in the Southeastern United States. *Climatic Change*, **90**, 37–71.

Tudor, M., and P. Termonia, 2010: Alternative Formulations for Incorporating Lateral Boundary Data into Limited-Area Models. *Mon. Wea. Rev.*, **138**(7), 2867–2882.

Varis, O., T. Kajander, and R. Lemmela, 2004: Climate and Water: From Climate Models to Water Resources Management and Vice Versa. *Climatic Change*, **66**, 321–344.

Von Storch, H., B. Hewitson, and L. Mearns, 2000: Review of Empirical Downscaling Techniques. *Conf. Proceedings RegClim Spring Meeting*, T. Iversen and B. A. K. Hoiskar, Eds., Norwegian Institute for Air Research, pp. 29–46.

White, P. S., et al., 1998: Regional Trends in Biological Resources - Southeast. *Status and Trends of the Nation's Biological Resources*, vol. 1, U.S. Geological Survey, pp. 255–314.

Widmann, M., C. S. Bretherton, and E. P. Salathe, 2003: Statistical Precipitation Downscaling over the Northwestern United States Using Numerically Simulated Precipitation as a Predictor. *Journal of Climate*, **16**, 799–816.

Wilby, R. L., T. M. L. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main, and D. S. Wilks, 1998: Statistical downscaling of general circulation model output: A comparison of methods. *Water Resources Research*, **34**(11), 2995–3008.

Wilby, R. L., S. P. Charles, E. Zorita, B. Timbal, P. Whetton, and L. O. Mearns, 2004: Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods. Tech. rep.

Wilks, D., 2008: High-resolution spatial interpolation of weather generator parameters using local weighted regressions. *Agricultural and Forest Meteorology*, **148**(1), 111–120.

Wilks, D. S., 1990: Maximum Likelihood Estimation for the Gamma Distribution Using Data Containing Zeros. *J. Climate*, **3**(12), 1495–1501.

Wilks, D. S., 1992: Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change*, **22**, 67–84.

Wilks, D. S., 1998: Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, **210**(1-4), 178–191.

Wilks, D. S., 1999: Multisite downscaling of daily precipitation with a stochastic weather generator. *Climate Research*, **11**, 125–136.

Wilks, D. S., 2002: Realizations of Daily Weather in Forecast Seasonal Climate. *Journal of Hydrometeorology*, **3**, 195–207.

Wilks, D. S., 2006: *Statistical methods in the atmospheric sciences*. Academic Press.

Wilks, D. S., 2009: A gridded multisite weather generator and synchronization to observed weather data. *Water Resources Research*, **45**(10), W10,419+.

Wilks, D. S., and R. L. Wilby, 1999: The weather generation game: a review of stochastic weather models. *Progress in Physical Geography*, **23**(3), 329–357.

Wooten, A., S. Raman, and A. Sims, 2010: Diurnal variation of precipitation over the Carolina Sandhills region. *Journal of Earth System Science*, **119**(5), 579–596.

Zorita, E., and H. Von Storch, 1999: The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods. *Journal of Climate*, **12**, 2472–2489.