Modeling Precipitation in the Northeastern United States

Lori Fomenko

Introduction

Rises in global temperature can lead to increased convective precipitation, assuming a sufficient amount of moisture availability in the atmosphere. The changes in precipitation patterns can vary in both intensity and movement across spatial domains (Dore, 2005). Changes in patterns of global circulation can also have local and regional impacts on precipitation distribution. However, regional and global connectivity in precipitation variability especially in regions such as the Northeast U.S. is still unclear (Guilbert et. al., 2015). The Northeast region of the U.S. has experienced an increase of nearly 10 mm/decade for precipitation, as well as the top increases in extreme precipitation of any region in the U.S. (Horton et. al., 2014). Variability in precipitation is important to understand, at the annual, monthly, daily, and sub-daily time scales. Increases in average annual precipitation may be spread out if there are just increases in wet days with minimal rainfall, but precipitation amounts may also come in shorter time periods with more intense and frequent storms. An overall increase in inter-annual precipitation may not fully explain what is happening at the daily and sub-daily scales, but it is a vital starting point. The Northeast region has seen increased amounts of precipitation over short time intervals leading to higher frequency of severe flooding over time (Collins, 2009). This flooding can have even larger impacts in urban areas with increased impervious surfaces due to land use changes, leading to increased runoff during events. If average annual precipitation is increasing, and there is variation even at the local-scale in cities, it might show that precipitation patterns and magnitude is increasing.

This study aims to look at a few objectives. First off, we want to test if there are any significant differences in observed precipitation at the small-scale. Precipitation can vary across a domain, such as a city due to urbanization impacts. Land use changes can lead to decreased permeability as well as decreased albedo. Darker colors absorb more heat from the sun, and reflect less radiation back to the atmosphere. This can lead to an increase in surface heating known as the “urban heat island” effect. This increased heat, especially that released in the evening can lead to more intense precipitation in urban areas known as the “urban rain island” effect. By looking at the differences between stations we are looking at the difference between low density areas and high density areas as well as, in this case, proximity to the ocean. Oceanic effects can also play a major role in precipitation. We also want to determine if there are significant differences between the individual station observations and the area-averaged precipitation amount across each city domain. We then compare the urban city domains (BOS and NYC) to the larger domains in which they lie to see if there is a major difference between average annual precipitation amounts in the cities as compared to the larger regions overall. We also wanted to see how downscaled GCMs perform in representing inter-annual precipitation patterns as compared to actual observations for a chosen historical time period. We then wanted to see the potential future projections for precipitation in each study domain based on bias correction from the historical study period applied to projections in a future GCM run. Lastly, we wanted to compare RCM and GCM output to observed precipitation for a selected event both spatially and numerically. Understanding how different models capture precipitation patterns and at what resolution can be essential to understanding both large-scale and small-scale processes and how they might impact overall precipitation amounts annually and at the daily scale during an extreme precipitation event.

Data and Methodology

This study looks at four domains in the Northeast region of the United States for including the entire Northeast region, and the Tri-State, Boston, and NYC areas. The boundary conditions (latitude and longitude) for each domain are summarized in Table 1. These boundary conditions are used throughout the various analyses to define the areas over which the precipitation is averaged.

Land surface stations were selected from the Global Historical Climatological Network (GHCN) database to compare precipitation distribution and variation at various points throughout the metropolitan areas of NYC and Boston. Nine stations in the Boston (BOS) area were evaluated, with five stations within the city boundaries, and four outside of the boundaries (Figure S1). Eleven stations in the New York City (NYC) area were also selected, with seven stations located inside the city boundaries, and four outside of the NYC bounds (Figure S2). Station information is summarized in Table S1. Quality control was done to determine the start and end date of data for each station and consistency of data. If there was missing data for any days during the study period, that station was not used for comparison with the area-averaged MetData for that grid cell.

Gridded surface meteorological data (MetData) from the University of Idaho (U of I) was obtained (http://metdata.northwestknowledge.net/) to compare individual GHCN stations to their corresponding grid cells represented in the MetData. The MetData is bias-corrected and derived from a combination of mesoscale reanalysis and assimilation data from PRISM (http://prism.oregonstate.edu/) gridded climate data and NLDAS-2 gauge-based data (https://ldas.gsfc.nasa.gov/nldas/NLDAS2forcing.php) and has a 4-km (1/24 degree) spatial resolution and daily temporal resolution. It is validated against a station of networks including; RAWS, AgriMet, AgWeatherNet, and USHCN-2.

Each GHCN station is located within one grid cell, with no two stations in the same 4 km x4 km grid cell. The average of each grid cell in the MetData dataset containing a GHCN station was taken so that the grid represented one point. The MetData average was then compared to the corresponding GHCN station data for a 20-year time period (1997-2016). The representation of each station with its corresponding grid-cell averaged MetData is shown by the addition of "Met" to the station identifier.

One year was chosen (2014) to isolate two stations in BOS (Figure S3) and two stations in NYC (Figure S4) and evaluate the ability of the MetData to represent the GHCN data at the daily time scale over the course of the year. The stations in each respective area were chosen based on continuous data during 1997-2016 period, length of data availability/record, and reliability. If the MetData was effectively able to consistently capture the precipitation trend throughout the chosen year, it could be used as a substitute for the GHCN observational data for further analysis.

To determine if differences in average annual precipitation among the stations were significant, ANOVA analysis was conducted between the stations themselves, as well as between the stations and the area-averaged MetData annual precipitation for both cities, Boston and NYC. The area-averaged annual precipitation is also calculated over 1991-2005 from the MetData observational dataset for all four study domains for comparison. To test for significant differences in annual average precipitation between the areas, ANOVA analysis was again used.

Downscaled GCM output from selected Coupled Model Inter-Comparison Project 5 (CMIP5) models was downloaded from the MACA Data Portal (http://maca.northwestknowledge.net/data_portal.php) at the daily timescale for a historical run from 1990-2005 and a future run under the Representative Concentration Pathway 8.5 (RCP 8.5) A2 emission scenario during the mid-century (2041-2060).

The MACA method of downscaling first interpolates the GCM and observation data to a 1 degree grid, then removes the seasonal and yearly trends at each grid point. A non-parametric quantile-mapping method is then used to correct the GCM coarse bias, and then analogs are constructed. The epochs adjusted in the second step are then replaced to maintain GCM data consistency, and the constructed analogs are then bias corrected using the same method to ensure observational data compatibility. The r1i1p1 is used for all downscaled GCM ensemble runs, except for CCSM4 which used the r6i1p1.

Historical runs (1991-2005) for the following CMIP5 models were used: CCSM4, CNRM-CR5, GFDL-ESM2M, HadGEM2-ES365, NorESM1m, MIROC5, and MIROC-ESM. Average annual precipitation averaged over each study domain was then compared to the corresponding area-average MetData observations to determine which models, if any, could successfully capture and represent the observed inter-annual precipitation variability.

Next, one CMIP5 model (CCSM4) was chosen for bias correction and a future model run. The bias was calculated by finding the difference between the model predicted average annual precipitation and the actual observed annual precipitation. This bias was then subtracted from the future run (2041-2060) to better represent what the actual inter-annual precipitation could look like in the later part of the century. This was done for each study area.

Lastly, an RCM3 run was completed to look at one selected precipitation event in 2005. The parameters were used in the domain.param file to run the model are summarized in Table S2. The spatial patterns of precipitation were compared at the daily scale between the MetData observations over the time period (10/07/2005 to 10/12/2005), and two GCM runs (CCSm4 and NorESM1m) and at the sub-daily 6 hourly scale for the RCM3 run over the Northeast domain. The magnitude and peak of the precipitation was also compared for the area-average MetData and both the GCM and RCM output.

Results and Discussion

Table 1. The boundaries of the four study domains for analyses, as well as modified boundary domain conditions for the RCM3 run.

Area	Lat Bounds(min, max)	Lon Bounds (min, max)	RegCM B.C.s (lat,lon)
Northeast Region	37, 48	-79,-67	Same
Tri-State Area	38.7,43.5	-75.3,-70.3	Same
Boston	42.18,42.57	-71.19,-70.98	(41.9,42.9), (-71.6,-70.6)
NYC	40.47,40.94	-74.11,-73.68	(40.4,41.4), (-74.5,-73.5)

Figure 1. Station Locations in Boston (left) and NYC (right)

Figure 2. Area-averaged precipitation for each station based on its respective grid cell in the MetData observational dataset compared with the domain average for Boston (left) and NYC (right) from 1997-2016.

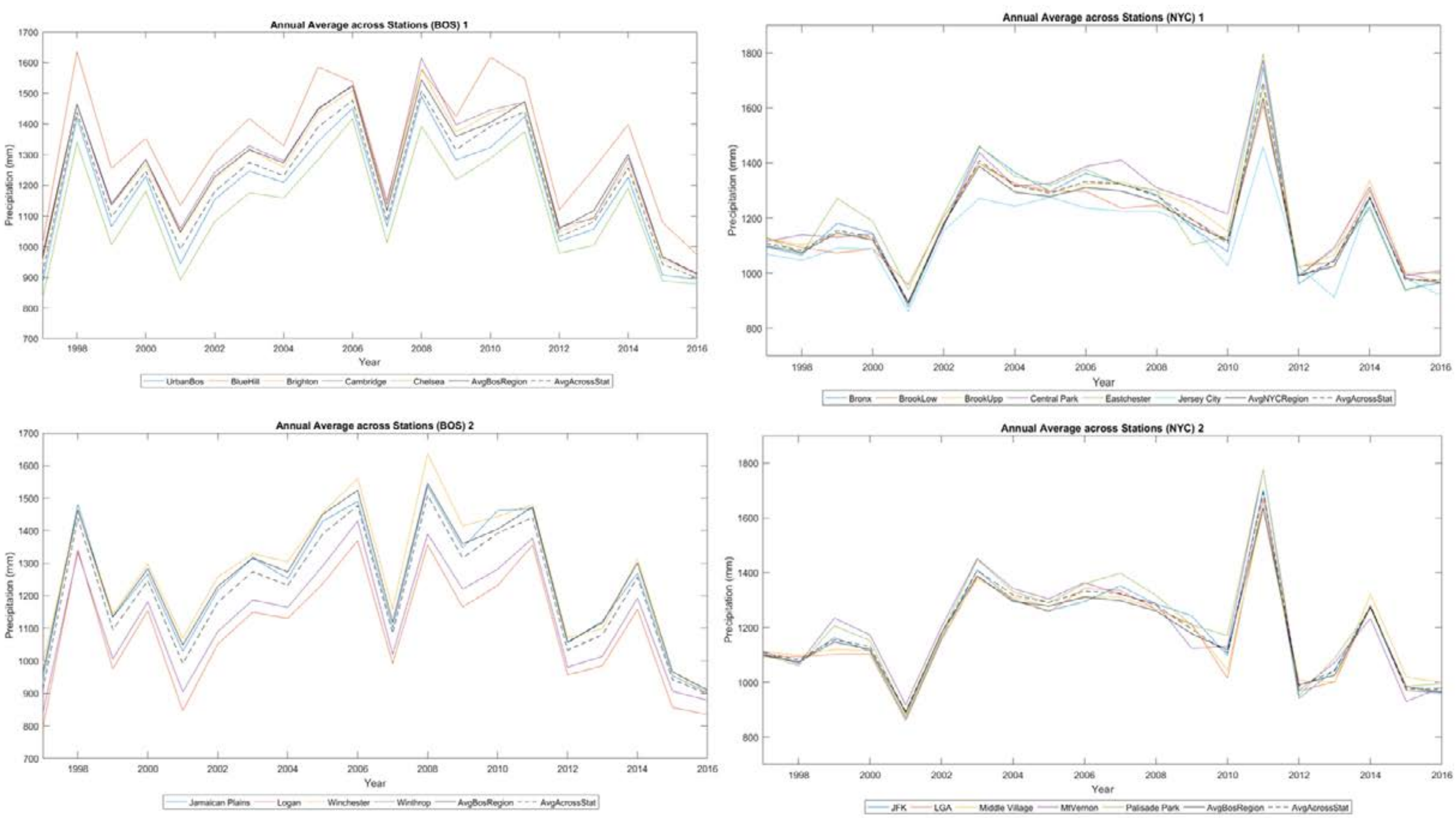
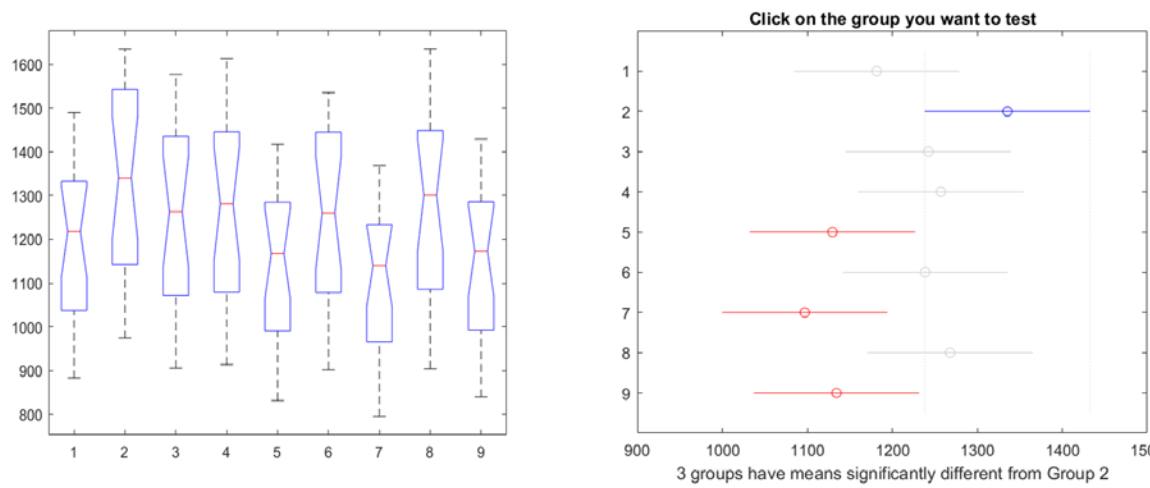


Figure 3. ANOVA testing results for comparing individual stations in the Boston area. Significant differences were found at both the 1% and 5% significance levels between the annual mean at the Blue Hill station and those at the Chelsea, Logan and Winthrop stations (highlighted in green), but not between the stations and the citywide average.



ANOVA Table

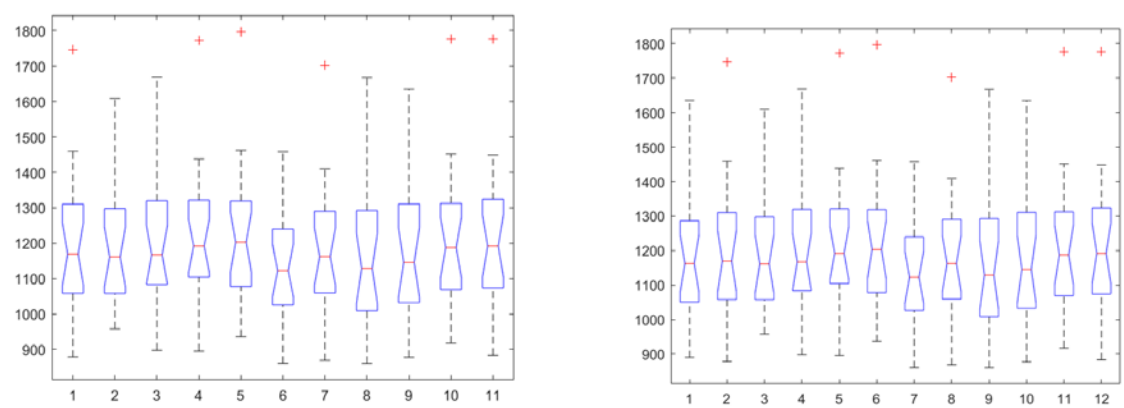
Source	SS	df	MS	F	Prob>F
Columns	979332.3	8	122416.5	3.12	0.0026
Error	6718881.9	171	39291.7		
Total	7698214.3	179			

1	2	3	4	5	6	7	8	9
UrbBos	BlueHill	Brighton	Cambridge	Chelsea	Jamaican Plains	Logan	Winchester	Winthrop

Figure 4. Same as Figure 3, but for the NYC area. ANOVA testing between stations (left) and between stations and the area-averaged mean for NYC (right). No significant differences were found.

1	2	3	4	5	6
Bronx	BrooklynLower	BrooklynUpper	Central Park	Eastchester	Jersey City
7	8	9	10	11	
JFK	LGA	Middle Village	Mt Vernon	Palisades	

1	2	3	4	5	6
NYC Region	Bronx	BrooklynLower	BrooklynUpper	Central Park	Eastchester
7	8	9	10	11	12
Jersey City	JFK	LGA	Middle Village	Mt Vernon	Palisades



ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	130983.3	10	13098.3	0.37	0.9596
Error	7461349.8	209	35700.2		
Total	7592333.1	219			

ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	133790.8	11	12162.8	0.35	0.9744
Error	8035840.9	228	35244.9		
Total	8169631.8	239			

Figure 5. Average Annual Precipitation amounts across all four study domains (1991-2005).

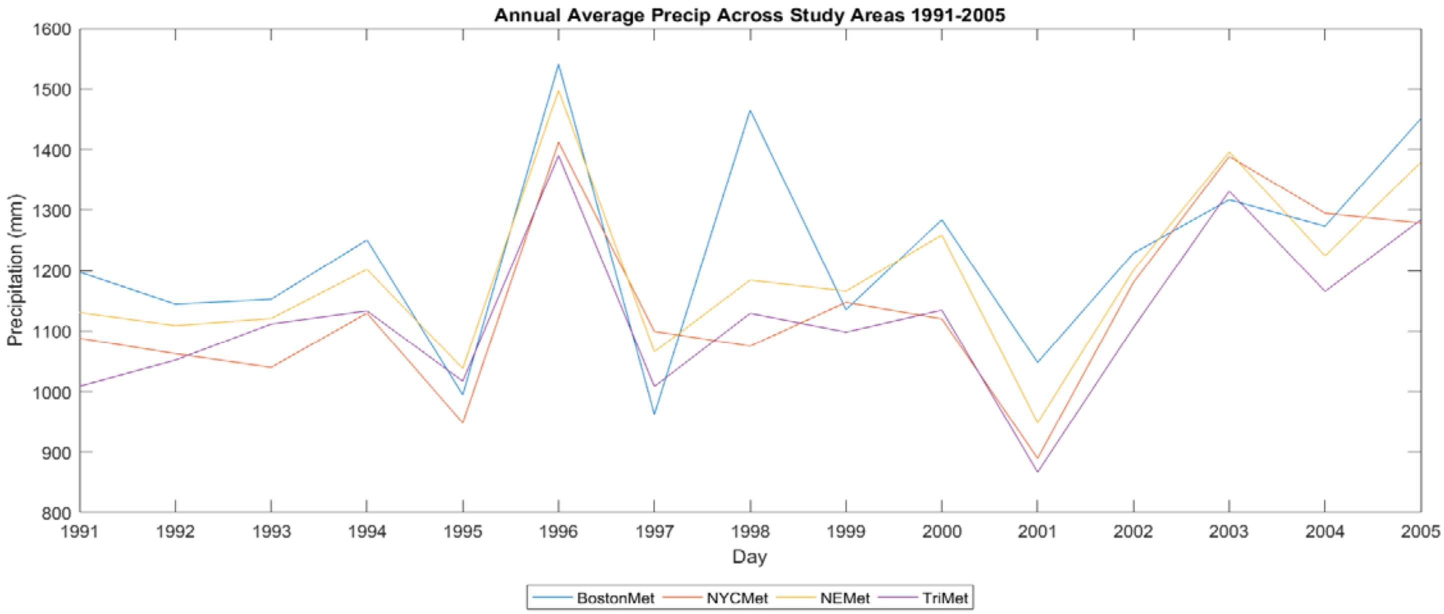
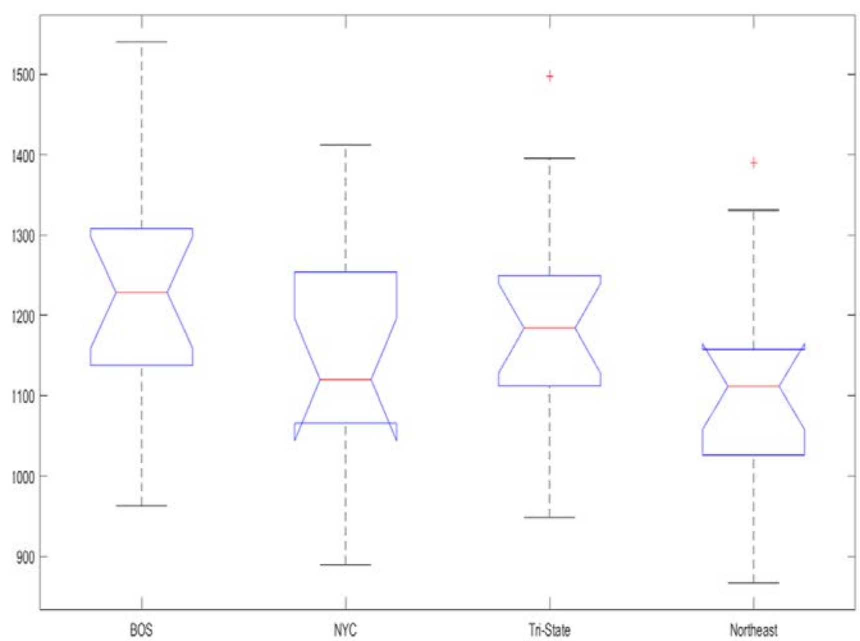


Figure 6. ANOVA testing results for comparison between the four study domains for average annual precipitation. No significant differences were found among any of the means.



ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	106322.8	3	35440.9	1.6	0.2003
Error	1242508.2	56	22187.6		
Total	1348831	59			

Figure 7. Annual GCM output compared with the Boston areal average (top), NYC area average (2nd row), Tri-State domain average (3rd row) and Northeast Regional average (bottom) from 1991-2005.

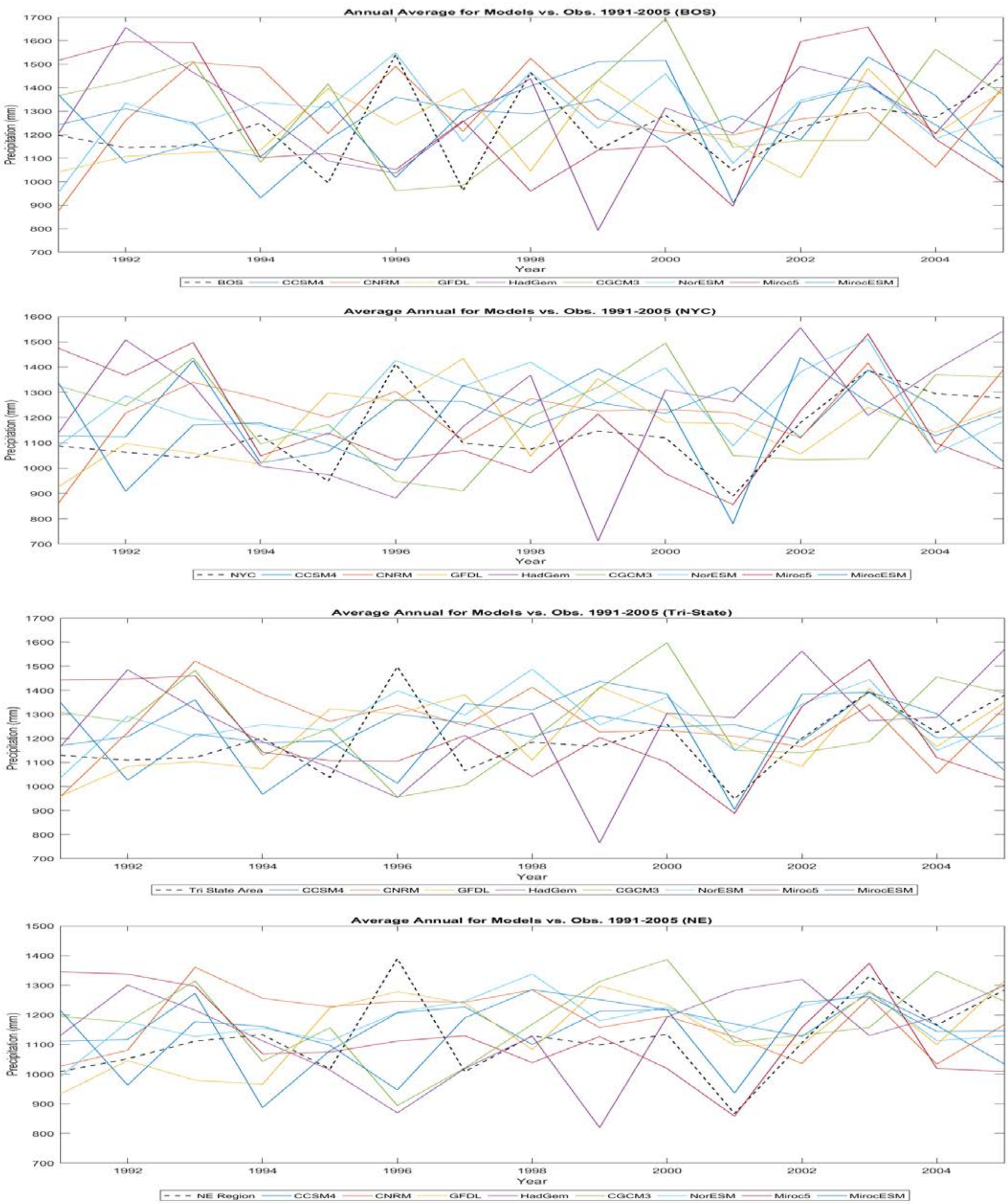


Figure 8. Bias corrected GCM output of average annual precipitation from CCSM4 for a historical run (1991-2005) and a future RCP 8.5 run (2041-2055) for the Boston area (top), NYC area (2nd row), Tri-state area (3rd row) and the Northeast region (bottom).

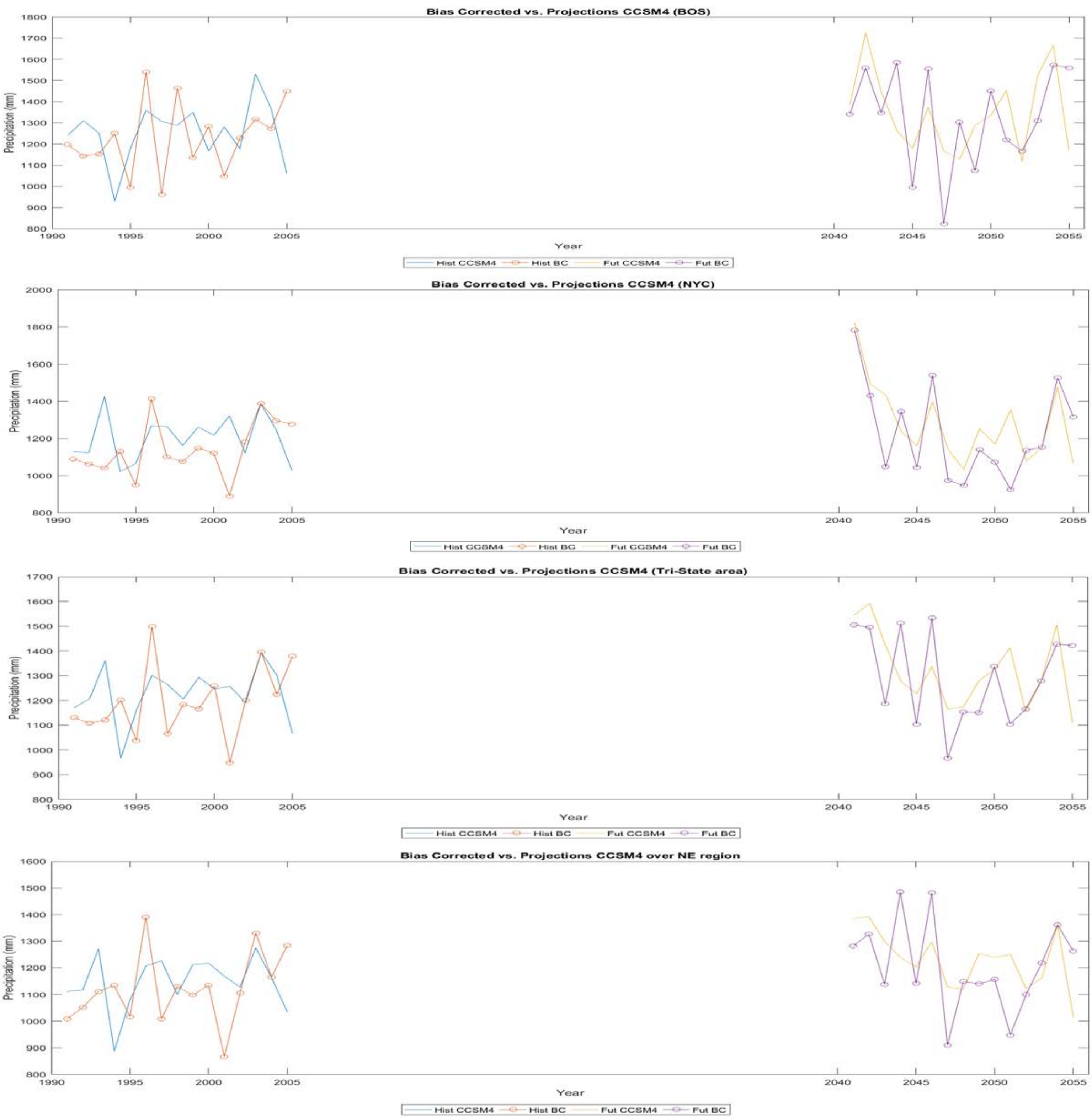


Figure 9. Spatial plots of precipitation from October 7-9, 2005. The top row shows the observed MetData precipitation, and the bottom row shows the GCM output from the NorESM1M model.

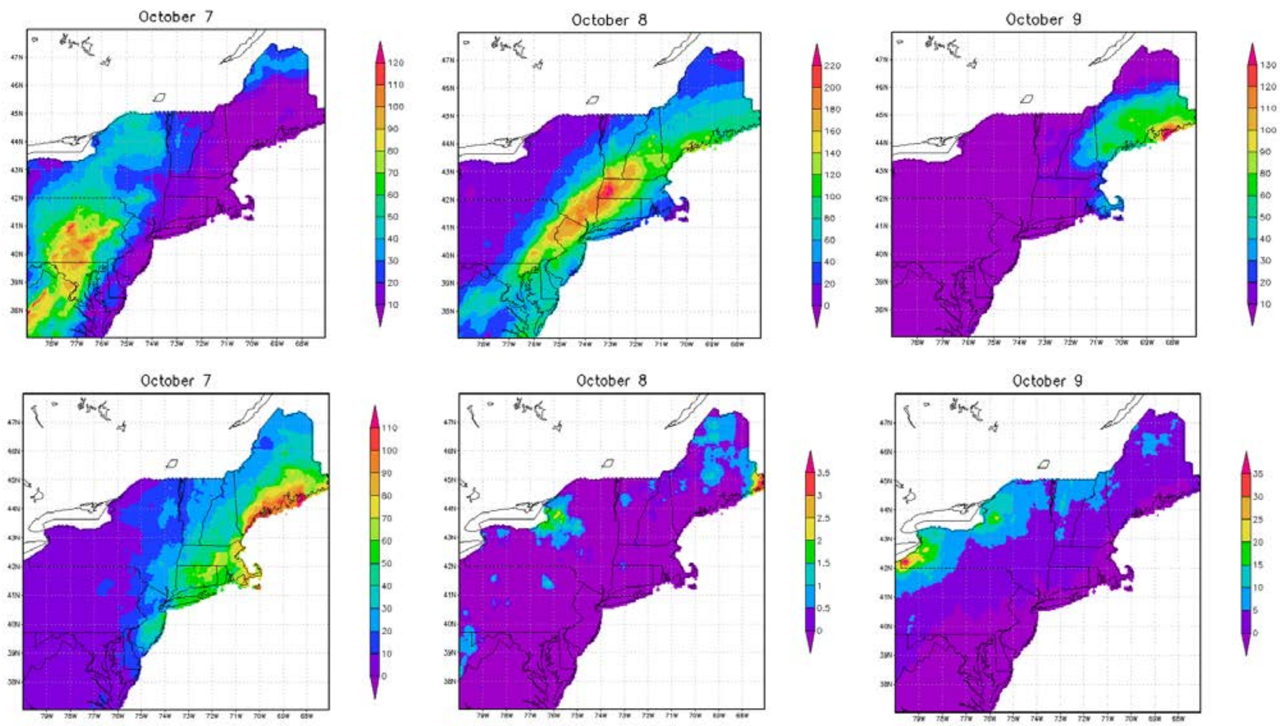


Figure 10. Daily MetData observed precipitation compared with selected CMIP5 GCM model output for the October 2005 event (October 7-12) for the Northeast region.

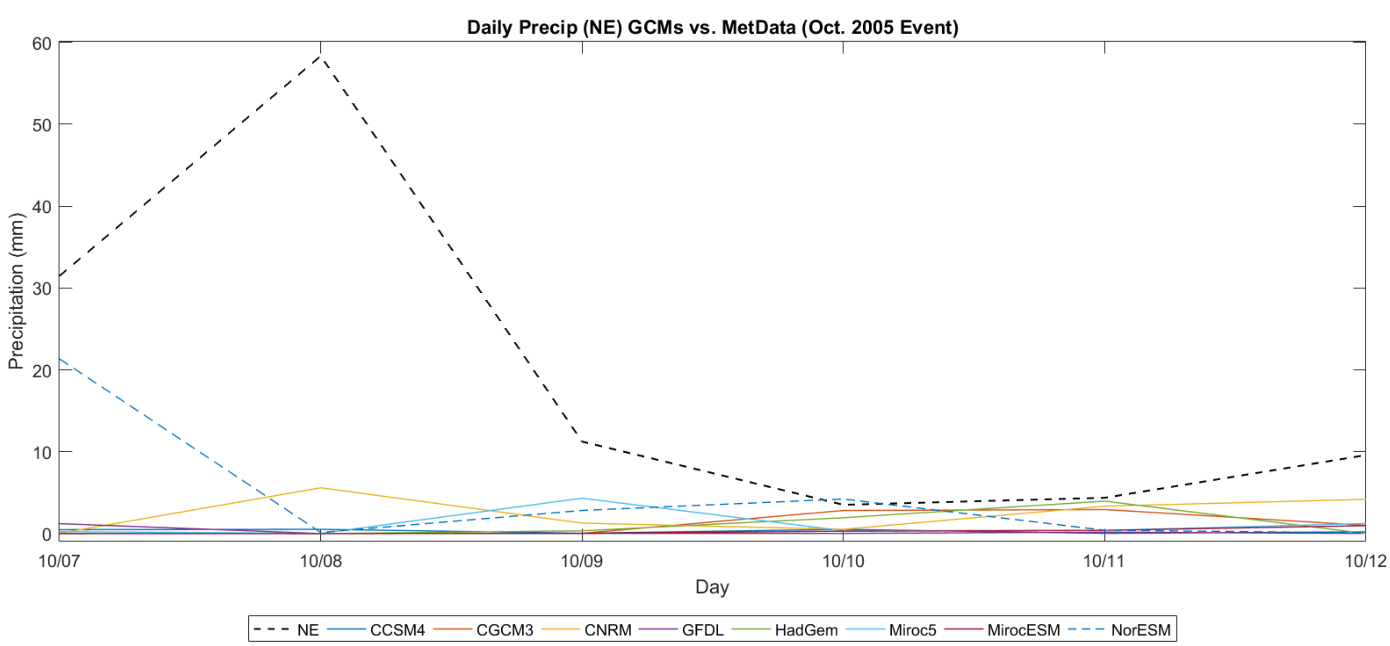


Figure 11. Precipitation plot for October 8th (peak observed precipitation) from MetData dataset (1st row) compared with spatial precipitation patterns and movements shown in the RCM3 run from 2005100800 (2nd row, 1st panel) to 2005100818 (2nd row, 4th panel).

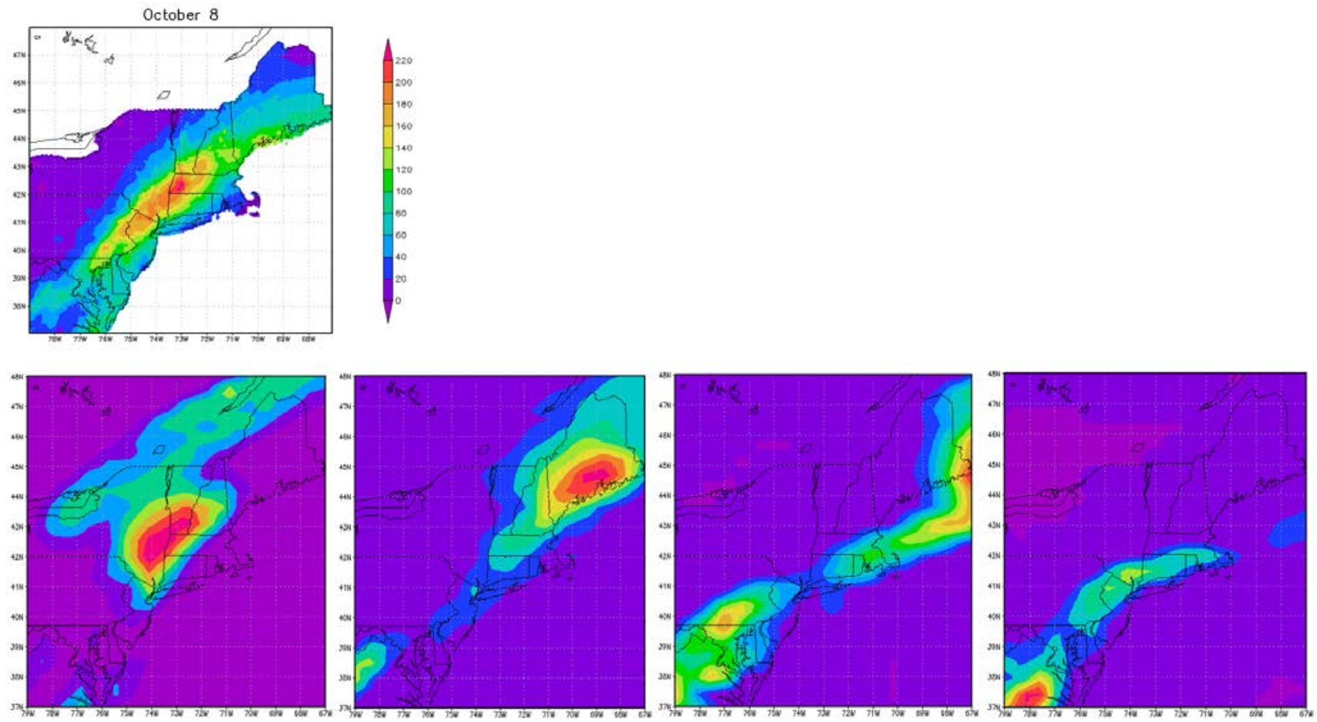
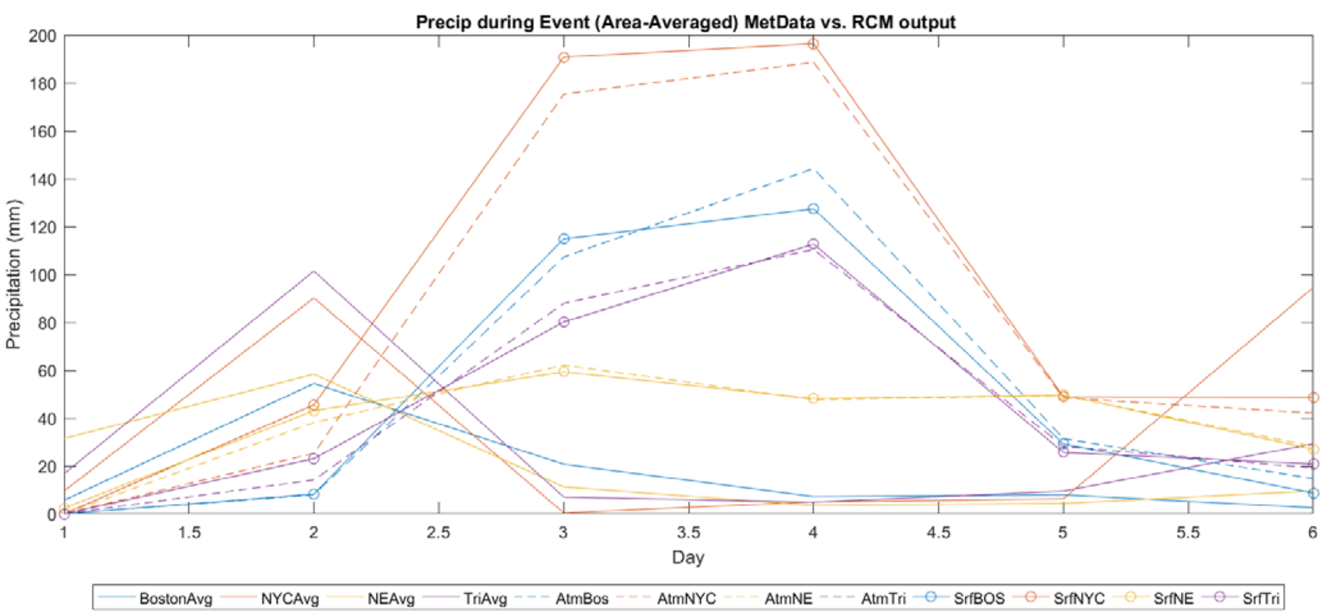


Figure 12. Area-averaged precipitation for October 2005 event shown by the atmospheric and land surface model components of the RCM, compared with the averaged MetData, for all four study domains.



The results of the ANOVA testing of individual stations shows us that there were significant differences at the 1% and 5% significance levels between one station (Blue Hill) and three others (Chelsea, Logan, and Winthrop) in Boston. This could be due to the “urban heat island/urban rain island” effect, with increased surface heating, landuse changes, and buildings impacting the development and movement of storms from city centers to just outside the boundaries. This same relationship was not found in NYC. The differences across Boston may also be attributed to local circulation patterns and oceanic forcings. An attribution study would need to be done in the future to take a closer look at this.

The results of the ANOVA testing for average annual precipitation over the four different domains showed that for the study period (1991-2005), there were no significant differences between any of the four stations at the 1% or 5% significance levels.

The comparison of MetData and historical GCM runs for the chosen CMIP5 model runs show a wide range of model variation in representing inter-annual precipitation variability over the study period (1991-2005). Each model has its own individual biases and uncertainty associated with it, with some overestimating and some underestimating the total precipitation amount, as well as capturing different trends or peaks at different years. These results should be taken lightly, however, due to the short time period, the natural variability of precipitation, and the different forcings in the GCMs as compared to the observational data.

The bias-correction for the CCSM4 model, applied for both a historical (1991-2005) and future (2041-2055) RCP 8.5 run shows some differentiation between the predicted annual precipitation and what the actual precipitation may look like based on past bias-correction. It is hard to tell by this indicator alone, and further study would need to be done to accurately represent the model biases in the past and future.

The RCM3 run looked at a shorter time period, only 5 days, in order to evaluate one event at the daily scale. Precipitation patterns in the MetData observational dataset show greater precipitation amounts over the whole Northeast regions from October 7th (before the peak), to October 8th (peak observed precipitation) to October 9th (after the peak) as compared to GCM output. Although the NorESM was able to capture more precipitation than the other GCMs in this study (on October 7th), it still severely underestimated the precipitation for the study period. The RCM results were also compared spatially at the 6-hourly scale with the MetData results over the course of the day on October 8th and showed a greater ability to capture the development, movement and precipitation over that day. The precipitation amounts from the RCM run are similar to the observed precipitation amounts on that day. The RCM peak precipitation, however, showed up about one or two days later. This could be due to a lack of spinup during the RCM run, and initial conditions and boundary conditions (ICBCs). The RCM results should be taken lightly due to model setup and natural biases and uncertainties associated with the model itself.

Summary and Conclusions

One key result of this study is the significant difference between a few individual stations within the Boston domain. Further research into this city, and other cities with similar characteristics (i.e. coastal, growing population and increased development) may lead to new information about the role of cities and urbanization on precipitations patterns, magnitude and frequency. The spread of model runs regarding precipitation at the annual time scale also shows the difficulty in predicting precipitation especially with natural and inter-annual variability. The RCM run being able to capture processes a little better at finer resolution could be helpful in future studies over smaller domains.

Some challenges I had throughout this project included utilization of data, quality control of the observational data, model comparisons and uncertainties, and choosing appropriate statistical and empirical measures to try and quantify results.

The land-based (GHCN) data for the cities (BOS and NYC) had very few consistent and reliable stations with continuous data for a long time period. There were many stations that only had data starting from the mid-2000s, which may be helpful in the next 20-30 years for comparison, but is not the most helpful right now as there isn't more historical data to compare the newer data with. It was also difficult to maintain a consistent time period for the study. Initially I wanted to look at 1997-2016, to have a 20-year period to compare with, but all of the GCM runs were from 1990-2005. I needed to take additional steps in extracting the data for all of the MetData and GCM output for 1991-2005 in order to get a 15-year period. Comparing the GCMs with the observations was also difficult for me to understand the most appropriate way to do so. If I had more time, I would have done more reading and research into the best methods of comparison and how to quantify the differences. The bias-correction done was also new to me, and trying to understand how to apply the bias for future runs, while not being sure about the initial comparison between the models and observational data did not make me feel confident about the model results. Lastly, the RCM run was challenging due to the limited years with data available for both the SST and Reanalysis datasets. I also ran into many errors when I tried to make modifications to the grid size, etc. which limited the applicability of the results. The model run was also too short, and didn't have an appropriate spinup included for model calibration.

Overall, I learned a lot through this project and course, and I hope to use the knowledge and skills in my future research, and do more careful analysis with a great deal of time spent on reading and reviewing statistical methods and how to properly compare model results with each other and with observations.

References

Collins, M. J. (2009). Evidence for changing flood risk in New England since the late 20th century. *JAWRA Journal of the American Water Resources Association*, *45*(2), 279-290.

Dore, M. H. (2005). Climate change and changes in global precipitation patterns: what do we know?. *Environment international*, *31*(8), 1167-1181.

Guilbert, J., Betts, A. K., Rizzo, D. M., Beckage, B., & Bomblies, A. (2015). Characterization of increased persistence and intensity of precipitation in the northeastern United States. *Geophysical Research Letters*, *42*(6), 1888-1893.

Horton, R., Yohe, G., Easterling, W., Kates, R., Ruth, M., Sussman, E., ... & Lipschultz, F. (2014). Ch. 16: Northeast. *Climate Change Impacts in the United States: The Third National Climate Assessment*, 371-395.

Liang, P., & Ding, Y. (2017). The long-term variation of extreme heavy precipitation and its link to urbanization effects in Shanghai during 1916–2014. *Advances in Atmospheric Sciences*, *34*(3), 321-334.

Tryhorn, L. and DeGaetano, A. (2011), A comparison of techniques for downscaling extreme precipitation over the Northeastern United States. Int. J. Climatol., 31: 1975–1989. doi:10.1002/joc.2208

Wang, G., Wang, D., Trenberth, K. E., Erfanian, A., Yu, M., Bosilovich, M. G., & Parr, D. T. (2017). The peak structure and future changes of the relationships between extreme precipitation and temperature. *Nature Climate Change*.

Washington, W. M., & Parkinson, C. L. (2005). *An introduction to three-dimensional climate modeling*. University science books.