Stony Brook University

An Application of Regression for Storm Surge Prediction along the New York/New Jersey Coast in Climate Models

A Thesis Presented
By

Keith J. Roberts

To

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Master of Science

In

Marine and Atmospheric Science

Stony Brook University

May 2015

Stony Brook University
The Graduate School

Keith J. Roberts

We, the thesis committee for the above candidate for the

Master of Science degree, hereby recommend

acceptance of this thesis.

Dr. Brian A. Colle, Thesis Advisor
Professor
School of Marine and Atmospheric Sciences

Dr. Malcolm J. Bowman, Thesis Reader
Professor
School of Marine and Atmospheric Sciences

Dr. Sultan Hameed, Thesis Reader
Professor

School of Marine and Atmospheric Sciences

This thesis is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Thesis

An Application of Regression for Storm Surge Prediction along the New York/New Jersey Coast in Climate Models

by Keith J. Roberts

Master of Science

in

Marine and Atmospheric Science

Stony Brook University

2015

Storm surge from extratropical cyclones can lead to significant coastal flooding, particularly along coastlines adjacent to wide and shallow continental shelves. In warming global climate simulations, studies have shown systematic regional changes to cyclones, which generate surges. However, there have been no formal studies that link regional cyclone changes to changes in the occurrence and intensity of surges along the densely populated New York/New Jersey coastline.

A multi-linear regression (MLR) approach is developed to predict 3-hourly surge during the cool season months (Oct. 1-March 31) using climate model data. At the three stations the approach is applied to (The Battery in New York, Atlantic City, New Jersey, and Montauk Point, New York), the MLR explains > 60.0 % of the observed surge variance in 3-h surge data. Predictions of storm maximum surge that meet or exceed the 95th percentile of storm maximum surges have a mean absolute error between 0.30 – 0.40 m and a mean error around zero. Using the same forecasted surface winds and pressures from the North American Mesoscale (NAM) model between October-March 2010 to 2014, surge predictions at The Battery are compared to raw output from a numerical hydrodynamic model's (SIT-NYHOPS) predictions. The accuracy of surge predictions between the SIT-NYHOPS and the MLR at The Battery are similar for predictions that meet or exceed the 95th percentile of storm maximum surge heights at 0-24 h leads.

The MLR is applied to an ensemble of six global climate models part of the 5th Generation of the Coupled Model Intercomparison Project (CMIP5) to study future surge changes at The Battery, New York. The predicted surges using climate models have realistic amplitudes, but the seasonal frequency of impactful surge events (≥ 0.61 m) is underpredicted by 2-5 events. While there are large interseasonal variations in storm surge, the majority of the models demonstrate no trends or intensity changes greater than natural or historically modeled surge variability. Only one GCM (CCSM4) predicts a noticeable shift toward more intense surge events in 2054-2079 compared to 1979-2004, which is explained with a northward shift in storm track. However, the effects of a regional sea level rise (SLR) scenario are much larger than any future modeled surge changes. Approximately 10 times more moderate coastal floods (≥ 2.44 m above MLLW) are predicted in 2069-2079 compared to 2009-2019 with the addition of a regional sea level rise scenario.

The intensity, spatial, and frequency distributions of classified tracks are related to surge events at The Battery to help explain the mechanisms that affect surge climatology. Surface cyclone tracks were automatically tracked in mean sea level pressure fields using the Hodges cyclone-tracking algorithm and were matched in time to modeled and observed surge time series. Tracks were automatically classified as either Miller Type A or Miller Type B following a set of rules. Both Miller Type A and Miller Type B cyclone tracks are shown to generate the majority (60/75) of observed impactful surge events at The Battery between 1979-2004 Nov.-March. The interseasonal frequency of Miller Type A tracks is shown to correlate positively with the interseasonal variations in the frequency of impactful surges in both observed (1979-2004) and modeled data (1979-2079). Between 2054-2079 and 1979-2004, there are no modeled changes that are consistent between models in the spatial, intensity, and frequency distribution of surge-generating cyclone tracks.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Financial support for this work was provided by New York Sea Grant.

This work is dedicated to my brother Brandon S. Roberts.

A tremendous amount of thanks and gratitude goes to Dr. Brian Colle for enabling me to explore my passion in science and writing. Without him, this project would never have been surmountable.

I sincerely thank my parents Debbie and Paul Roberts for which my inspiration to do everything comes from. Their continued efforts have been superhuman.

A special thanks goes to the SOMAS faculty, particularly Mark Lang, and some colleagues, particularly Mike Erickson and Nathan Korfe, who have made time spent here enjoyable and memorable.

Lastly, this work would not have been possible without my friends Alan Suchecki and Patrick Lanigan who have provided endless support, advice, and great friendship.

Chapter I: Introduction

a. Background

For low-lying coastal regions, the passage of storms force water on-shore resulting in the development of coastal floods, which can lead to significant damage. The portion of the total water level driven by meteorological factors is referred to as the storm surge, while the total water level determines the depth of flooding. Both the occurrence of coastal floods and storm surges characterize the frequency and intensity of nearby synoptic period (3-7 day) weather (Zhang et al. 2000; de Oliveria 2009); therefore, beside the obvious geophysical and human impacts these events incur, these data are also helpful to understand climate variations.

The inwardly curved and densely populated coastline that extends from near Atlantic City in New Jersey to Montauk Point in New York is termed the New York/Jersey (NY/NJ) Bight/coastline, and is vulnerable to storm surges (Fig. 1.1). The nearby bathymetry is relatively shallow, generally less than 100 m below mean sea level, and the curvature of the Bight enhances surges at its apex, which happens to be where New York City, New York is located (Bowman et al. 2005; Fig. 1.1). A typical storm surge scenario for NY/NJ Bight involves an extratropical cyclone propagating from south to north along the eastern seaboard (Colle et al. 2010).

During the cool-season months (Oct.-March), extratropical cyclones occur frequently nearby the New York/New Jersey coastline due to large land-sea temperature gradients between the mid-Atlantic and the northeast U.S and large amounts of oceanic moisture. For the prediction of the sensible impacts associated with these storms, such as storm surge and winter weather, it is helpful for forecasters to understand the track and evolution of these cyclones. Extratropical cyclones that occur over the Atlantic coastal region in the United States are classified into two

types of cyclones that were proposed by Miller (1946): Miller Type A and Miller Type B. An example of a Miller Type A and Miller Type B cyclone track are both illustrated in figure 1.2.

Miller Type A events are singular, well-defined, and usually intense surface lows that develop in or around the Gulf of Mexico and travel northeastward along the U.S. east coast as they intensify. Miller Type B events are more complex involving multiple surface lows and cyclone redevelopment induced by topographical effects from the Appalachian Mountains (Miller 1946). For Miller Type B events, the primary low's track over the Great Lakes weakens as a secondary low develops sometime (< 36 h) after the decay of the primary. Both types of tracks are associated with storm surge; however, there has been no formal investigation of storm surge generation from a classified cyclone track perspective.

Extratropical cyclones generated between 0 to 10 storm maximum surges that would at least warrant a coastal flood advisory if they had occurred during high tide between Nov.-March over 1979-2012 (Fig. 3b) at The Battery (Fig. 1.1). During the early Fall, hybrid storms or extratropical transitions can also generate surge in this region, such as Sandy in 2012 (Blake et al. 2013; Galarneau et al. 2013), Floyd in 1999 (Colle 2003; Atallah and Bosart 2003), and the "Perfect Storm" in 1991 (Drag 2000). For example, on 11-12 December 1992 a Nor'easter resulted in the fourth highest surge and third highest total water level at The Battery in the 20th century. During this event, a slow-moving surface cyclone tracked northeastward and nearly stalled as it made landfall onto the NY/NJ coastline. This resulted in 30-40 kt (15-20 m s-1) easterly winds over a 24 h period. The persistent easterly component of winds forced a storm maximum surge of 1.68 m and a storm maximum water level of 2.83 m above mean low-level water (MLLW; 1.0 m; 1983-2001) at The Battery (Fig. 1.4; Colle et al. 2008). The total water level over topped the wall in the Brooklyn-Battery tunnel to a height of 2 m, which short-

circuited the NYC subway system for up to 10 days while portions of the FDR drive were flooded with 1.5 m of sea water in lower Manhattan (National Weather Service Disaster Survey Report 1994).

Recently in late October of 2012, Hurricane Sandy underwent an extratropical transition before making landfall on the New Jersey shoreline on the 29th of October 2012 (Galarneau et al. 2013). Sandy's landfall produced a catastrophic storm surge stretching from New Jersey to Rhode Island that made it the sixth costliest U.S. tropical cyclone on record since 1900 (Blake et al 2013). In spite of the incredible damage Sandy incurred, two modeling studies have demonstrated that Sandy's storm surge was not the worst-case scenario for the NY/NJ coast (Forbes et al. 2014; Georgas et al. 2014). However, tropical-related surge events have occurred much less frequently than extratropical related events in the northeast U.S. Based on overwash sediments in marshes along the NY/NJ coastline, Scilleppi and Donnelley (2007) showed only five major hurricane (category 3 or greater) landfalls around a 700 y period from about the 12th to 20th century.

b. Storm surge physics

The temporal and spatial variations in storm surge and the total water level is a basin-specific response to a number of factors related to meteorology, the air-sea interface, and coastal and bathymetric attributes. Some of the most influential factors are: the magnitude, direction, and duration of surface wind stress, the wind fetch length, atmospheric surface pressure perturbations, bottom stresses, and wave breaking (Komar 1988). Storm surges are also influenced by coastal shelf geometry and local bathymetry. To illustrate the dependence of storm surge on the shelf geometry, one can assume uniform shelf geometry and storm surge then can

be represented by equation 1.1, where η is the storm surge height, τ_s is the wind stress at the air–sea interface, and g is the gravitational acceleration (Resio and Westerink 2008). The two geometric factors that represent the shelf geometry and bathymetry are h, the depth of the water, and W, the shelf width.

$$\eta \propto \frac{\tau_s}{gh} W \tag{1.1}$$

As one can see from equation 1.1, continental shelves that are wide and shallow, like the one in the NY/NJ Bight, are the most vulnerable to large surges. Admittedly, the bathymetry and shelf properties are realistically not uniform, and this can complicate the storm surge physics considerably.

Inertial forces become important for surge generation considering that storm surge events are the result of time-integrated surface wind stresses that occur over 12-36 h. The Coriolis force acts to transport the water at an angle to the right of the mean wind direction in the northern hemisphere, and this effect is referred to as Ekman transport. The action of the Coriolis force is represented by equation 1.2, where x is the across-shore distance, Ω is the rotation rate of the Earth (7.29×10^{-5} radians/sec), ϕ is the latitude, and v is the along-shore surface wind (Komar 1988).

$$\frac{\partial \eta}{\partial x} \propto 2\,\Omega \sin\phi\; v \tag{1.2}$$

From equation 1.2, the across-shore gradient in storm surge induced by the Coriolis effect depends on both the latitude and the magnitude of the along-shore surface wind field. The integral of equation 1.2 with respect to time demonstrates that the longer the duration of surface wind stress, the larger the transport of water due to the Coriolis force.

In deeper waters away from the coast, the influence on storm surge due to wind stress is considerably less (e.g., Eq. 1.1), and becomes primarily driven by atmospheric pressure perturbations. The effect pressure perturbations have on the water is commonly referred to as the inverted barometer effect and represents the rise in the ocean surface to restore hydrostatic balance with the atmosphere in the vertical column of water. The inverted barometer effect is quantified in equation 1.3 with $\overline{P}$ representing the average mean sea level pressure (MSLP) over some period (e.g., seasonal).

$$\eta = 0.01(P - \overline{P}) \tag{1.3}$$

For every one hPa decrease in MSLP, the ocean surface rises approximately 0.01 m (Harris and Angelo 1963).

Breaking waves can also add a positive and negative influence on the near-shore surge because of a momentum balance in the across-shore direction (Komar 1988). This is described by equation 1.4 where S_{xx} is the radiation stress, ρ is the density of water, φ is the difference between still water level and the water-level in the presence of waves (set-up/set-down), and h is the depth of the water column.

$$\frac{\partial S_{xx}}{\partial x} + \rho g(\bar{\varphi} + h)\frac{\partial h}{\partial x} \tag{1.4}$$

As the wave height increases, S_{xx} increases since it proportional to the wave height, therefore $\frac{\partial S_{xx}}{\partial x}$ increases and φ must decrease to maintain a momentum balance (Komar 1988). The decrease of water levels due to waves breaking on-shore is referred to as wave set-down. It follows from equation 1.4 that if waves are breaking near-shore, then $\frac{\partial S_{xx}}{\partial x} < 0$ and φ must rise to compensate. Thus, wave breaking near the shore adds a positive influence to surge height and wave-breaking on-shore induces a wave set down.

c. Storm Surge Modeling Approaches

Numerical hydrodynamic models (NHMs) are used primarily for the hindcasting and short-term forecasting of water levels and storm surge. Surface winds and pressure data from atmospheric models are input into NHMs to generate water level forecasts (Shen et al. 2005; DiLiberto et al. 2011; Orton et al. 2012; Georgas et al. 2014). For example, Shen et al. (2005) used the Advanced Three-Dimensional Circulation Model for Coastal Ocean Hydrodynamics or ADCIRC (Westerink et al. 1993) to hindcast the surge for Hurricane Isabel (2003) in the Chesapeake Bay and predicted the surge (1.9–2.5 m) to within 0.3 m of the observed surge. Colle et al. (2008) used ADCIRC to realistically predict the surge associated with hurricane Floyd and the December 1992 Nor'easter around New York City. A more comprehensive analysis was performed using an eight-member ADCIRC ensemble based on multiple operational NHMs, the NOAA Extratropical surge model based on the SLOSH model (ET-SURGE; Burroughs and Shaffer 1997), and the NYHOPS model at Stevens Institute based on the sECOM model (SIT-NYHOPS; Georgas 2010) for 74 forecasts between 2007-2008 (DiLiberto et al. 2011). The bias corrected root mean square error for the ensemble at a 12-48 h lead-time ranged from 0.10 m to 0.20 m between ensemble members. The mean of all three modeling systems, after bias correcting the individual ensemble members had less error than any one bias corrected ensemble member or the ADCIRC ensemble, thus illustrating the benefit of a multi-model surge ensemble.

Water level predictions can also be made using statistical relationships between observed surge and atmospheric conditions (wind & pressure) that are significantly faster to compute than NHMs and can offer considerable skill. Statistical models for storm surge and water level prediction were frequently used in the 1960 to the 1970s (Harris and Angelo 1963; Pore

1964,1965; Gjevik and RøEd 1975; Amin 1978). These studies employed regression models that featured correlations in verification data that ranged from 70.0 % to 90.0% (Harris and Angelo 1963; Amin 1978; Gjevik and RøEd 1975). For example, Pore (1964) used hourly surface wind and pressure observations at varying lags for three stations along the East coast of the United States as predictors in a MLR to predict the storm surge at Atlantic City, NJ. The MLR was evaluated for 18 independent extratropical storms between 1956 to 1961 and predicted surge during these storms accurately with a correlation of 0.93 and average root mean square error of 0.15 m (Pore 1964).

More recently, regression has been used to predict water levels along the northern coast of Germany to enhance the skill of NHM water level forecasts (Müller-Navarra and Knüpffer 2010). Salmun et al. (2011) developed a regression model for storm surge predictions at The Battery in NYC using sea level pressure and observed wave height at an offshore buoy. They showed that the regression had similar error metrics to NOAA's deterministic storm surge model (ET-SURGE), and, in another study, used the same regression to bias correct surge forecasts from a NHM (Salmun and Molod 2015). Regression models have also been employed to surrogate NHM ensemble spread for predictions of storm surge in Venice (Mel and Lionello 2014).

d. Motivation

An outstanding question for the NY/NJ coastline concerns how the occurrence and intensity of future storm surges and coastal floods may change throughout the 21st century in a warming global climate. Extreme events like Hurricane Sandy (2012) illustrated the vulnerability of the New York/New Jersey (NY/NJ) Bight to catastrophic flooding from landfalling storms

(Forbes et al. 2014; Georgas et al. 2014) and brought to light concerns about how frequent events like this may be in the 21st century (ASCE Conference Proceedings et al. 2013). These concerns were the combined result of research on the rarity of Sandy-like events (Hall and Sobel 2013), historical sea-level rise (SLR) trends (Talke et al. 2014), and future extratropical cyclone track changes in a warming global climate (Mizuta et al. 2011; Colle et al. 2013; Chang et al. 2012, 2013; Marciano et al. 2015)

Since storm surge events are forced by extratropical cyclones, any variations in these cyclones could result in changes to the surge climatology. Studies on 21st century cyclone track changes for North America and the north Atlantic have demonstrated a poleward shift and a weakening of the extratropical cyclone track into the late 21st century (2079-2100). This is the result of a weaker lower tropospheric meridional temperature gradient in a warmer global climate (Lambert 1995; Colle et al. 2013; Chang et al. 2012 and 2013). In spite of decreasing surface cyclone frequency, Mizuta et al. (2011) and Colle et al. (2013) show evidence that there may be an increase in the frequency of relatively deep (< 980 hPa) cyclones during the Northern Hemispheric winter by the late 21st century. Marciano et al. (2015) further demonstrated that future Miller Type A cyclones, that originate in moisture-rich locations like the Gulf of Mexico, could become more intense in the late 21st century due to enhanced latent heating in a warmer global climate.

Any effects that global warming induced shifts in the north Atlantic cyclone track could have on coastal flooding have not been formally investigated along the NY/NJ coastline. Even if future trends in cyclones were negligible into the 21st century, the occurrence of coastal flooding would still be affected by sea level rise (SLR). In the International Panel on Climate Change's 5th Assessment report (IPCC 2013), the consensus of research on SLR demonstrated that it

would very likely (90-100%) increase the occurrence of extreme sea level heights in some regions around the globe by 2100. Recently, a study for the NY/NJ coastline projected a regional sea level rise of 0.61 m to 0.87 m by 2100 above a reference level defined over 1984-2005 (Zhang et al. 2014; Fig. 1.5).

Both impacts associated with SLR and future storm changes on coastal flooding need to be quantified for geo-engineering purposes and for environmental reasons. For instance, the intense coastal flooding associated with Hurricane Sandy created breaches in the south shore beaches of Long Island, which led to significant changes to temperature, salinity, and tidal variations (American Littoral Society

2012). The rapid changes to the coastline associated with Sandy altered the marine ecosystem in many ways; for instance the breeding and spawning of Alewife Fish in the Great South Bay exhibited unusual behaviors in the years following Sandy (American Littoral Society 2012). If similar magnitude events occur like this more frequently into the 21st century, then certain actions may need to be taken.

Unfortunately, the use of NHMs with global climate model data for multi-decadal predictions of water levels is computationally prohibitive. To mitigate computational expense, a statistical method involving multi-linear regression between atmospheric data and observed storm surge was developed in Roberts et al. (2015) to predict the 3-h storm surge at The Battery. This tool enables long-term (~100 y) projections of storm surge and water levels from 6-h global climate model data at a point at virtually no computational expense. The statistical surge and water level predictions enable one to determine how the future cyclone changes associated with global warming may alter the occurrence and intensity of surges along the NY/NJ coast into the 21st century with all other things being equal.

The primary goals of this study are to determine how the occurrence of storm surge and coastal flooding may change into the 21st century and to quantify the relationship between the nearby cyclone track and the surge on seasonal. The previously mentioned studies on cyclone track changes in the north Atlantic showed an overall reduction in the frequency of cyclones but an increase in the occurrence of relatively deep ones (e.g., Colle et al. 2013; Chang et al. 2013). Based on these studies, it is hypothesized that a similar trend that demonstrates an overall decrease in the frequency but an intensification of surges would be exhibited in modeled storm surge during the 21st century. However, given the magnitude of predicted sea level rise (Zhang et al. 2014), it is hypothesized that the frequency of coastal floods (tide + surge + SLR) will predominately be affected by regional sea level rise scenarios and not surge changes.

This research will answer the following questions:

1. How well can a multiple linear regression (MLR) predict historical storm surges at Atlantic City, The Battery, and Montauk Point between 1 October to 31 March, using reanalysis data for wind and pressure forcing?
2. How sensitive are the MLR predictions to the size of the domain and cross-validation period used for the wind and pressure forcing?
3. How do the 3-h MLR surge predictions at these stations, before and after a bias correction, compare to the raw results from a hydrodynamic model (SIT-NYHOPS) currently used for operational predictions?
4. Can the historical storm surge climatology between 1979-2004 Nov.-March at The Battery be accurately captured by a multi-linear regression forced with atmospheric data from an ensemble of six GCM's?

5. Using this GCM ensemble, how do modeled storm surges that occur between Nov.-March vary during a future period (2006-2079) in warming global climate simulations and how does sea level rise impact modeled water levels?
6. How are cyclone tracks related to storm surge events at The Battery and how do variations in the frequency, intensity, and spatial distributions of cyclones relate with the variations in frequency and intensity of storm surge events at The Battery?

Chapter 2 describes the data and methods used to develop the regression model and to analyze climate model projections of surge. Chapter 3 concerns the verification of the regression model and discusses its results. In Chapter 4, the regression is applied to climate model data and predictions are analyzed from a time series and cyclone track perspective. Conclusions are presented in Chapter 5.

Chapter II: Data and Methods

a. Data used for Regression Development

The data, the treatment of the data, the methods used to develop the regression model, and the methods for its application to climate model data are described in this chapter. Raw sea level data were obtained from NOAA Tides and Currents (http://tidesandcurrents.noaa.gov/) in hourly increments from 1 January 1979 to 31 December 2012 for Atlantic City and Montauk Point and from 1 January 1948 to 31 December 2012 for The Battery (Fig. 2.1). Data from these three stations were selected based on the percent availability of verified hourly sea level data back to 1979 (Table 2.1) and their respective location along the NY/NJ coastline. University of Hawaii Sea Level Center (UHSLC) hourly water level data (http://uhslc.soest.hawaii.edu/data/)

was used to fill in missing NOAA Tides and Currents water level data for Atlantic City, NJ between 1979-1987 and 1993-1995.

One can define storm surge by equation 2.1 as the difference between the detrended (subtract line of best fit from data) observed water level from the astronomical tide. The trends removed from each station's data are also shown in table 2.1

$$\text{surge} = (\text{water level} - \text{sea level rise}) - \text{tide} - \text{surge interactions} \quad (2.1)$$

The definition of storm surge as the linear superposition of tide and water level assumes that tide-surge interactions are small compared to the surge height. One common way to test for the presence of tide-surge interaction is to compare the distributions of the tidal residuals (water level-tide) at different stages of the tide using Quantile-Quantile (qq) plots in which data are ranked based on probability (e.g., Bernier and Thompson 2007). In the absence of tide-surge interactions, a qq plot of tidal residuals recorded at different tidal stages should fall close to the one-to-one line. In order to test for this, the tidal residuals were binned based on three stages of the tide: high tide (HT or MHW-MLLW$\pm 0.4\ m$), mean tide (MT or Mean Tidal Range-MLLW$\pm 0.4\ m$), and low tide (LT or MLW-MLLW$\pm 0.4\ m$). The binned tidal residuals were then examined on qq plots with respect to the nine permutations of the three tidal stages. From figure 2.2 at the three stations, the tide-surge interaction appears to be negligible based on the conformity of the scatter to the one to one line. In other words, the surge does not appear to preferentially occur at a particular tidal stage.

The North American Regional Reanalysis (NARR) (Mesinger et al. 2004) at 32-km approximate horizontal grid spacing and 3-h was utilized for the development of the regression model. The Climate Forecast System Reanalysis (CFSR) (Saha et al. 2010) available at approximately 38-km horizontal grid spacing every 6-h was used to test the sensitivity of the

regression model to a different reanalysis dataset. CFSR data were linearly interpolated from 6-h to a 3-h time interval. The CFSR and NARR reanalysis datasets were chosen for this study because they featured similar high-resolution data on latitude longitude grids, similar data assimilation inputs, and are frequently cited in atmospheric science literature. The gridded 3-h zonal (x) and meridional (y) winds at 10-m above sea level and mean-sea level pressure (MSLP) were used from 1979 to 2012 and until 2010 for CFSR during the cool season months of 1 October to 31 March. Only the cool season was used for the regression development because this is when storm surge-generating cyclones occur most frequently along the NY/NJ coastline (Zhang et al. 2000; Colle et al. 2010), and our approach is designed for extratropical storms, not hurricanes given their small sample size over the study period.

All 10-m wind data were converted to wind stress components using a drag formulation documented in Garrett (1977). The formula for the conversion from wind speed to wind stress is shown in equations 2.2-2.3:

$$\tau_{x/\rho_w} = C_D * \rho_a * V_x * U \quad (2.2)$$

$$\tau_{y/\rho_w} = C_D * \rho_a * V_y * U \quad (2.3)$$

where U is the total wind speed, V_X and V_y are the east-west and north-south wind components, 999.97 kg/m^3 (ρ_w) is the density of water, τ_x and τ_y are the zonal and meridional wind stress, respectively, 1.23 kg/m^3 (ρ_a) is the air density, and C_D is the drag coefficient (Eq. 2.4).

$$C_D = 0.001 * (0.75 + 0.067 * U) \quad (2.4)$$

If C_D is > 0.003, then C_D is set to 0.003. This is a slip boundary condition that represents the shearing of wave tops when the wind speed exceeds a critical value (Garratt 1977). The drag formulation represented by equations 2.2-2.4 is used in some operational models, such as

ADCIRC (Westerink et al. 1993). A different drag coefficient developed by Large and Pond (1981) was also used to calculate wind stress, but large differences in results were not found.

Atmospheric reanalysis wind stresses and pressure used for the regression's development was collected within rectangular boxed regions extending to the south and east of each station, which are illustrated in figure 2.3 and with dimensions specified in table 2.2. The box orientation with the station on the top left corner captures the wind and pressure forcing that precedes the storm maximum surge since most significant cyclones that generate storm surges along the NY/NJ coast travel northward along the U.S. east coast (Colle et al. 2010). The geometry of the box was orientated in the east-west direction and rectangular to emphasize a fetch of winds that push water towards the NY/NJ coastline since a typical surge event is forced by a low-level easterly jet on the northern edge of the cyclone in the NY/NJ region (Pore et al. 1974; Colle et al. 2010). A square boxed region was also attempted but it produced models with less accurate predictions.

The size and station of the box is specified with a two-digit alphanumerical string (e.g., A1). The station the box is used for is indicated by a the letters A, B, or C, and represents Atlantic City, The Battery, and Montauk Point stations, respectively. The number 1, 2, or 3 that follows the letter specifies the box size in figure 2.3 for each station, with number 1 denoting the smallest box, number 2 the medium sized box, and number 3 for the largest box. A smaller box than number 1 was not attempted because this would increase interpolation errors for applications using global climate model data. For instance box 1 (smallest nested box) was 338 x 111 km (see Table 2.2) and contained approximately 30 NARR grid points (25 CFSR grid points), whereas box 1 could only contain approximately 4 GCM grid points for a typical GCM

resolution of 1°x1°. A box larger than box 3 was not used because the correlation between the atmospheric forcing and storm surge weakened considerably.

b. Development of Regression

The methodology used to develop multi-linear regressions to predict 3-hourly storm surge during the cool season months (Oct. 1-March 31) at three stations along the New York/New Jersey coast is described in this section. Multi-linear regression (MLR) assumes the data generating process is a linear combination of predictors that have zero-mean and unit variance based on the data they were developed with. Each predictor is related to the Y predictand by the coefficients β. The coefficients are determined by minimizing the sum of the squared error between predictions and the given data; this minimization criteria is called least squares.

1. Predictor Specification

The dynamics of storm surge generation are represented statistically in the regression with three predictors: two predictors represent the two components (x and y) of prolonged wind stress (τ_x and τ_y), and one predictor represents the surface atmospheric pressure forcing (P). The relationship between atmospheric conditions and storm surge is approximated to be linear and thus monotonic.

The prolonged action of wind stress on the sea-state is represented as a moving sum where a given predictor X at time t is the sum of L prior data intervals (Eq. 2.5) of an average wind stress component (x or y) over a particular boxed region (Fig. 2.3). These predictors are

termed accumulated (*A*) wind stress $A\tau_U(t)$. In equation 2.5, *L* denotes the lag in data intervals, the subscript denotes the component (*x* or *y*), *t* is the time interval, and Δt represents the data interval in hours (i.e., 3-h).

$$A\tau_U(t) = \sum_{i=1}^{L} \overline{\tau_U(t - L * \Delta t)} \tag{2.5}$$

The $A\tau_U(t)$ predictors represent the dynamical behavior of the water motion as a response to a prolonged wind stress component over a local spatial region. The calculation of the moving sum in this manner implies a constant weighting (i.e., box-car weight function) over the *L* time steps of data. A Gaussian weighted moving sum filter did not result in significantly different predictive accuracy.

The third predictor is the spatial minimum of the pressure at mean sea level (*P*) in a particular box. The *P* predictor is specified to take into account the inverted barometer effect, which represents the rise in the ocean surface to restore hydrostatic balance in the vertical column of water due to an atmospheric pressure perturbation. The *P* predictor is not lagged because storm surge-modeling studies demonstrate that a region of positive surge follows the pressure minimum of the cyclone (Angelo and Harris 1963; Weisberg 2008; Weaver 2004). MLRs trained with a lagged pressure predictor did not outperform models trained with a zero-lagged predictor in verification data. Different configurations of the *P* predictor were attempted, such as using the zero-lagged box-average MSLP as a predictor; however, the correlation between station storm surge and box-average or minimum MSLP was statistically similar ($\propto = 0.05$). Considering the insignificant differences in mean and minimum MSLP correlation with station storm surge, the minimum was used as a predictor.

In summary, the model has the generalized form of equation 2.6.

$$Predicted\ Surge(t) = \beta_0 + \beta_1 A\tau_u(t) + \beta_2 A\tau_v(t) + \beta_3 P(t) \quad (2.6)$$

Predictions of storm surge made by equation 2.6 are referred to as MLR and begin when *t* is greater than or equal to the larger of $(Lx/Ly)\Delta t$. The selection of the moving sum lengths *Lx* and *Ly* is explained in the section b2.

In order to develop the regression model, 3-h time series of surge data were regressed against the predictors described above. Atmospheric reanalysis and water level data were time-matched with each other and then partitioned into training and testing periods using an alternate-cool-seasons approach for cross-validation, as depicted in figure 2.1. Cool seasons that began on an odd year were used for training, and cool seasons that began on an even year were used for verification; this cross-validation configuration is referred to as CV1. All results were repeated with the training and testing periods switched (i.e., even years training and odd years verification), and this data configuration is referred to as CV2. When the data configuration used to compute a particular statistic is relevant, it is listed with the reanalysis data's name (either CFSR or NARR) first followed by the cross-validation period (either CV1 or CV2).

Analysis uncertainty is quantified as the range (min and max) of the statistic using the same cross-validation periods while varying the reanalysis dataset. Sampling uncertainty is likewise quantified as the range of the statistic calculated with different cross-validation periods using the same reanalysis data. The total uncertainty is the maximum and minimum of the statistic using all four combinations of data by considering both cross-validation and reanalysis uncertainty together and taking the range. All statistics are reported as the average value of the total uncertainty with square brackets indicating the range of the total uncertainty unless otherwise noted.

2. Candidate model selection

A range of possible combinations of $A\tau_U(t)$ and *P* are considered by allowing *Lx*, *Ly* to vary independently of each other between 0 to 27-h in 3-h increments while simultaneously varying the box size for all three predictors. The cross-correlation between station storm surge and the wind stress predictors were used to select an upper bound on the maximum length-moving sum to be considered. Moving sum lengths > 27 h were not considered as the cross-correlation between the moving sum wind stress predictors and station storm surge became statistically insignificant ($\propto = 0.05$) for these longer periods. This created a set of 2,700 candidate models for each station given that there were 30 configurations of $A\tau_x(t)$, 30 configurations of $A\tau_y(t)$, and 3 configurations of *P* for each candidate model. The Akaike Information Criterion (*AIC*) (Eq. 2.7) was calculated with the linear model function in MATLAB statistics toolbox for each candidate model. The notation *k* denotes the number of parameters in the model and *l* is the maximized likelihood function of the model (Akaike 1974).

$$AIC = 2k - 2\,ln(l) \tag{2.7}$$

The *AIC* is a commonly used model selection criterion that represents the goodness of fit of a model by rewarding a larger maximum likelihood while penalizing for overfitting (Burnham and Anderson 1998); therefore, a candidate model with a lower *AIC* is preferred. Since the absolute value of *AIC* is a likelihood (Burnham and Anderson 1998), the difference is calculated between each model's *AIC* from the model with the minimum *AIC*, and this is termed ΔAIC_i.

3. Bias Correction Approach

The MLR used in this application does not consider all of the dynamical effects and interactions that generate and affect storm surge, such as the Coriolis effect and wave set-up/set-

down. Model approximations such as linearity, the predictor's box geometry, and fixed moving sum lengths affect surge predictions as well. The omitted dynamical effects and model approximations may contribute to systematic errors in storm surge predictions by the MLR. In general, regression often has difficultly replicating the occurrence of extremes as frequently as observed (Klein et al. 1959; Pore et al. 1974; Wilks 2006).

It was apparent from various fitting approaches that the MLR exhibited a systematic bias using both training and verification data for large (>95th percentile) storm maximum surges. A MLR model also developed for predictions of storm surge at The Battery (Pore et al. 1974), which used point-based atmospheric predictors, exhibited a similar systematic underprediction bias for storm maximum surge predictions. In Pore et al. (1974), MLR predictions of storm surge were multiplied by the reciprocal of the multiple correlation coefficient or 1/R to improve large storm maximum surge predictions. An identical approach to bias correct, as shown in Pore et al. (1974), is used in this work. Since the predicted data is multiplied by 1/R, the overall variance explained by the MLR decreases since the fitted data no longer corresponds with the least-squares estimates. While the overall variance for the bias corrected (BC) predictions decrease because of the multiplicative factor (1/R), it is later shown that the mean bias is reduced for large surge events. Therefore, this BC approach sacrifices total variance explained for a smaller mean bias in the tails of the predicted surge distribution. When MLR predictions are multiplied by the reciprocal of the multiple correlation coefficient obtained from a training fit, predictions are referred to as a bias-corrected MLR or BC-MLR.

c. Regression Verification Approach

1. Verification using Reanalysis Data

The regression's predictions of storm maximum surge were calculated using reanalysis data and then were compared to observed storm maximum surges over 1979-2012 Oct.-March. In order to evaluate the performance of the regression model, time series of storm surge were separated into peaks, which represent the storm maximum surge. Only peaks in the 3-h surge time series > 0.10 m, and that had a minimum spacing of 24-h before and after the peak surge were considered for the calculation of thresholds described below. If two peaks in the time series occurred within 24-h of each other, only the larger was retained. Results were not sensitive to a larger minimum spacing interval between peaks until a spacing of 42-h was used.

The 95th and 99th percentiles of all surge peaks at each station were used as thresholds to bin data. The 95th and 99th percentiles correspond to a surge of 0.59 m at Atlantic City, 0.61 m The Battery, 0.57 m at Montauk Point, and 0.90 m at Atlantic City, 0.91 m The Battery, 0.87 m at Montauk Point, respectively. These thresholds are referred to as minor and moderate storm maximum events since they represent the maximum surge height for each storm. The minor event criteria corresponds to a surge height that occurs ≤ 5.0 % of all surge events but >1.0 % of all surge events, and the moderate event criteria occur ≤ 1.0 % of all surge events based on the 1979-2012 cool-season-month data at each station. The notation *all* denotes both minor and moderate surge events were considered together. Similarly, when the storm maximum total water level (tide + surge + SLR) exceeded 2.44 m above MLLW, it was considered a moderate coastal flood. Both the impactful surge event and the moderate coastal flood criterion are similar to the

flooding definitions used in Colle et al. (2010), which are based on flooding potential at The Battery during high tide and are used by the National Weather Service.

A useful summary statistic for a regression model is the fraction of variance explained by the fitted predictand or R^2 (Eq. 2.8). R^2 is referred to as the coefficient of determination, and its square root is equivalent to the multiple correlation coefficient (R) between the model and observations (Wilks 2006).

$$R^2 = 1 - \frac{\sum u^2}{\sum (Y - \bar{Y})^2} \tag{2.8}$$

Other useful metrics for summarizing the predictability of the MLR are: the binned mean error (ME; Eq. 2.9), when either predictions or observations meet minor and/or moderate storm maximum event thresholds, and the binned mean absolute error (MAE, Eq. 2.10). Note n denotes the binned sample size, not the total sample size.

$$ME = \frac{1}{n} \sum_{i=1}^{n} \hat{Y}_i - Y_i \tag{2.9}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{Y}_i - Y_i \right| \tag{2.10}$$

The scatter index is also calculated (Eq. 2.11) for when either observations or predictions of storm maximum surge meet the all event criteria. The scatter index represents the percentage of error compared to the mean observation (Howard et al. 2009) and is a measure of how consistent the error is.

$$scatter\ index = \frac{std(\hat{y} - y)}{\bar{y}} * 100 \tag{2.11}$$

The scatter index is a standard metric used for wave-model intercomparisons and assessing surface wind forecasts (Clancy et al. 1986).

2. Comparison with raw SIT-NYHOPS Predictions

SIT-NYHOPS *(*Stevens Institute of Technology-New York Harbor Observing and Predicting System*)* is an estuarine and coastal ocean 3-D numerical hydrodynamical model (NHM) that runs operationally for the waters of NY, NJ and CT; its model grid is indicated in light blue in figure 2.4. The model presently produces 72-h forecasts of water levels, waves, currents, and other quantities four cycles per day. The forecasts use meteorological information from the National Weather Service's atmospheric forecast model and USGS river flow gauges (Georgas 2010; Georgas et al. 2014). SIT-NYHOPS has been operational since 2006 and forecasts are available online at http://www.stevens.edu/maritimeforecast (Georgas et al. 2007; Blumberg and Georgas 2008; Georgas et al. 2014).

SIT-NYHOPS 0-24 h lead time forecasts from the last operational forecast cycle run each day are forced with 3-h North American Mesoscale (NAM) 10-m wind and MSLP pressure data. SIT-NYHOPS 10-minute water level predictions were provided for comparisons to the MLR's predictions forced using the same NAM meteorological data described above over the cool seasons between 2010-2014. Only MLR predictions at The Battery were compared with SIT-NYHOPS's predictions since less than 50% of the archived SIT-NYHOPS grid (Fig. 2.4) overlapped with the two boxes used for the MLR predictions at Atlantic City and Montauk Point, whereas about 75% of the grid overlapped for The Battery's box.

In order to compare SIT-NYHOPS predictions with predictions from the MLR, correlation statistics (Eq. 2.8), a set of basic error metrics (Eq. 2.9 and Eq. 2.10), and the scatter

index (Eq. 2.11) were calculated. These statistics were calculated for the raw (MLR) and bias-corrected MLR predictions (BC-MLR). Predictions were not separated into events for this comparison due to the limited number of events that occurred between 2010-2014. Instead, MLR predictions were linearly interpolated from 3-h to a 10-minute time-step to correspond with the SIT-NYHOPS prediction's time step. Interpolated 10-minute surge predictions are referred to as storm surge intervals. The sample size is reported as the number of 10-minute surge interval predictions or observations that meet the specified bin and lead time criteria.

d. Coupled Model Intercomparison Project Global Climate Model Data

Data from six global climate models (GCM) that are part of Phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012) are used to predict 3-h storm surge from 1979-2079 at The Battery. CMIP5 provides a repository of global climate model (GCM) simulations from various modeling institutions around the world, which all follow a consistent framework; thus, enabling direct comparisons of different GCM's data. The considered GCMs are listed in table 2.4 together with their resolutions, abbreviations, and relevant attributes. The CCSM4, CNRM-CM5, BCC-CSM1-1, GFDL-ESM2G, and CanESM2 are used since, with the exception of the CanESM2, were also studied over a similar region in Colle et al. (2013) and increase confidence in the cyclone track results presented below. CMIP5 GCM's follow plausible emission, land-use, and socio-economic population scenarios that extend to the year 2100 and beyond, which are otherwise known as Representative Control Pathways (RCPs). The data used here correspond with the highest emissions scenario in all CMIP5 experiment and it is termed RCP8.5. The RCP8.5 assumes little coordinated effort between nations to reduce carbon emissions (Taylor et al. 2012).

The data employed for surge predictions are gridded lowest model level (LML) u and v wind components (U) and surface mean sea level pressure (MSLP) data. All GCM data are separated into a historical period and a future period for the cool season months of November-March. Note that the Nov.-March period used here differs from the Oct.-March period used to develop the regression because Nov.-March is the period in which the cyclone track data described below was verified in Colle et al. (2013). The 25 cool season historical period began on 0000 UTC 1 January 1979 and terminated on 1800 UTC 31 March 2004. The future period began on 0000 UTC 1 November 2006 and terminated on 1800 UTC 31 March 2079. In this study, the 25 cool seasons historical period is compared to the last 25 cool seasons during our future period (2054-2079), and this period is termed the middle-late 21st century.

In addition to using atmospheric data, surface cyclone tracks were automatically tracked in the mean sea level pressure (MSLP) field of the six GCMs and Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) following the methods documented in Colle et al. (2013). The Hodges (1994, 1995) cyclone-tracking algorithm was used to obtain the cyclone track using 6-h spectral bandpassed (Wavenumbers equal or less than 5 and larger than 70 are removed.) MSLP data. The automated cyclone track used here was manually verified in Colle et al. (2013) for 11 Januaries every other year from 1980 to 2000 from 20.0 to 60.8 N and 40.0 to 90.8 W and carried approximately 5.0%-10.0% uncertainty in identifying surface cyclones.

e. Global Climate Model Surge Predictions

Common GCM evaluation methods generally aggregate all GCM to one standard grid as it enables the direct comparison of the results of the various models. Here we describe the steps used to accomplish this.

All GCM gridded data was linearly interpolated from 6-h to a 3-h time step in order to work with the MLR. After this, GCM data was spatially interpolated using a cubic spline interpolation to a uniform horizontal grid with a grid spacing of $0.5^{\circ}$by 0.5° (Fig. 2.5). The u and v wind speed data was then reduced to 10-m from it's estimated lowest model level height via a power law transform (Eq. 2.12), which is a common method to reduce wind speeds to various levels (Archer and Jacobson 2005; Pryor et al. 2005; Devis et al. 2014).

$$U(10) = U(HLML)(\frac{10}{z_{HLML}})^{\propto} \tag{2.12}$$

The term $\propto$ was which was assumed to be constant with a value of 0.11 and corresponds with a neutral stability stratification (Hsu et al. 1993). Neutral stability is assumed since there is no temperature data to characterize the GCM's atmospheric stability below the LML. The height of the LML was determined using the hypsometric equation with the difference in pressure at the LML sigma level and the surface pressure (Devis et al. 2014) (Table 2.4). The interpolated 3-h wind components were converted to wind stress components using the Garrett (1977) drag formulation.

Systematic biases in the wind and pressure predictions in the GCMs motivated the need to bias correct (BC). The mean and standard deviation of the MLR's predictors were conditionally BC (denoted by an *) according to CFSR data based on the sign of the average zonal wind along the NY/NJ coastline using the method of Xu and Wang (2014). In order to BC, GCM and CFSR fields were partitioned into training (denoted by a subscript T) and verifying datasets (denoted by a subscript V) and then further into means (denoted by an over bar) and anomalies (denoted by a single quote). In verification period data, the mean bias during the

training period was removed from the GCM, and the GCM anomaly was inflated by the ratio of standard deviations (*S*) (Eq. 2.13; Xu and Wang 2014).

$$GCM_V^* = \overline{CFSR}_T + (\overline{GCM}_V - \overline{GCM}_T) + GCM_V' \frac{S_{CFSR|T}}{S_{GCM|T}} \quad (2.13)$$

The BC approach was calibrated in the historical period with the CV1 configuration and then repeated with the CV2 configuration. The verification period data from each setup was concatenated to form a continuous time series of historical period bias corrected data. The future period data was BC using all historical period data. Figure 2.6 shows that the BC method improved the average of ≥ 0.6 m and ≥ 1.0 m 3-h surge intervals and decreased the ensemble spread around the corresponding observed values. Further, the standard deviation of ≥ 0.6 m and ≥ 1.0 m three hour surge intervals (Table 2.3) for the ensemble mean of GCMs during the historical period was shifted more toward the observed.

Modeled time series of storm surge were calculated by following the data treatment methods described above, converting the data to regression predictors, and then applying these data to the MLR. For the prediction of 3-h water level time series, the tidal uncertainty was taken into account. To quantify tidal uncertainty in water level predictions, a 3-h tidal time series from 1979-2012 relative to MLLW was added to the 3-h surge predictions 1,000 times. The tidal time series were shifted one 3-h time step forward iteratively 1,000 times to quantify the flooding potential at different tidal stages. The mean and 95th percentiles of this sampling distribution of water levels were used to represent tidal uncertainty.

Chapter III: Regression Results and Discussions

a. Regression Results

The multi-linear regression (MLR) has the generalized form of equation 3.1 with β representing the model's coefficients. The coefficients for each MLR are shown in table 3.1.

$$Predicted\ Surge(t) = \beta_0 + \beta_1 A\tau_u(t) + \beta_2 A\tau_v(t) + \beta_3 P(t) \qquad (3.1)$$

The selection of moving sum lengths for $A\tau_u(t)$ and the box used for *P* were determined by minimizing the ΔAIC_i calculated for each candidate MLR. This procedure was described in Chapter 2.

Figure 3.1 shows an example of the candidate model selection; the ΔAIC_i calculated for 2,700 candidate models represents all possible permutations of predictors at The Battery. The variation in ΔAIC_i demonstrates the sensitivity of the MLR to the choice of *Lx*, *Ly*, and box size. The selection procedure is most sensitive to the choice of box size followed by *Lx* and then *Ly*. The variation in *Lx* creates the 10 (0 to 27-h in 3-h increments) clusters in ΔAIC_i from left to right in figure 3.1, and the variation in *Ly* creates the variation in ΔAIC_i within the 10 clusters. The differences between solid colored lines, or MLR's ΔAIC_i calculated with predictors from the same box are significantly (i.e., they do not overlap) different from all other solid lines for all variations in *Lx* and *Ly*. As *Lx* increases to 15-h, the difference in $\Delta AIC_{\square}$ between MLRs created with the same box reached a maximum. A horizontal reference line in figure 3.1 demonstrates with its intersection with the ΔAIC_i curve that some different candidate models have equal likelihoods. For example, a MLR with box 2 that used an *Lx* of 6 h had the same ΔAIC_{i} as a MLR created with box 2 with an *Lx* of 21 h. The final selections of *Lx*, *Ly*, and box size for each station are listed in table 3.2 and used the smallest boxes or A1/B1/C1.

b. Verification of Regression using Reanalysis Data

A way to visualize the performance of a model is by scattering predictions against their corresponding observed values; predictions that fall on a 45-degree reference line (1:1 line) represent perfect predictions. The MLR verification (using the non-training years) 3-h surge predictions scattered against 3-h observed surges are highly organized around the 1:1 lines (Fig. 3.2). The MLR model explains around 60.0-70.0% of variance in verification storm surge data at the three stations and has a fairly stable R^2 that varies by ~10.0% among different cross-validation periods and between the NARR and CFSR (Table 3.2). Analysis uncertainty in R^2 is the largest at Montauk Point (8.0%) while at The Battery and Atlantic City is smaller, between 1.0% and 3.0%. Sampling uncertainty in R^2 is the largest at Atlantic City and is associated with two storms that occurred on 12 December 1992 and 14 March 1993. The multiplicative factor (1/R) used to bias-correct inflates predictions from the values that minimize the sum of the squared error (least-squares). Therefore, predictions made by BC-MLR explain less 3-h surge variance than the raw surge predictions. The decline in variance explained using BC-MLR is the most apparent at Atlantic City in the NARR CV1 configuration because of the two mentioned extreme storms.

The scatter at all three stations has an average slope of less than one compared to the 1:1 reference line (Fig. 3.2). This indicates that the MLR has an underprediction bias or negative mean error (ME). The expected MLR prediction of all storm maximum surge events has a negative ME from -0.05 m to -0.30 m and a mean absolute error (MAE) from 0.30 m to 0.50 m (Fig. 3.3a; Fig. 3.4a). Compared to MLR, BC-MLR surge predictions for the all event criteria have a ME that is more centered around zero at The Battery of 0 m [-0.03 m to 0.01 m] and at

Montauk Point of 0.02 m [-0.03 m to 0.05 m] (Fig. 3.3b). However, for the NARR CV1 model at Atlantic City, BC-MLR predictions for the all event criteria had a larger MAE (Fig. 3.4b) compared to MLR predictions, which resulted in a positive ME of 0.11 m [0.02 m to 0.23 m]. However, BC-MLR improved minor and moderate storm maximum surge predictions in the NARR CV2 and CFSR CV2 cross-validation configurations of data at Atlantic City. In summary, minor and moderate storm maximum surge predictions made by BC-MLR explain less variance overall (Table 3.2); however, minor and moderate storm maximum surge predictions are less biased (Fig. 3.3b), more accurate (Fig. 3.4b), and have a lower average scatter index compared to predictions made by MLR (Table 3.3).

c. Comparison of Regression with Raw SIT-NYHOPS Predictions

Each 1800 UTC surge forecast is "hot-started" with the previous 1800 UTC forecast to populate the moving sum predictors used in the MLR. The raw SIT-NYHOPS predictions are compared to the MLR and BC-MLR surge predictions. Predictions of storm surge intervals are separated into 5 bins from > 0 m & ≤0.20 m, etc. to ≥ 0.80 m in increments of 0.20 m and two lead-time categories (3-12 h and 15-24 h). All statistics are averaged for the four cool seasons between 2010-2014. The bars in figure 3.5 and figure 3.6 indicate the 95th percentile range of the binned error metric, which were obtained by bootstrapping 1,000 times. Confidence intervals could not be utilized because the small sample size in some bins made some of the binned sampling distributions non-Gaussian.

Raw SIT-NYHOPS and BC-MLR predictions of storm surge intervals ≥0.6 m for 3-12 h (n =1,315, 10-min intervals) and 15-24 h (n= 511, 10-min intervals) lead times at The Battery have similar expected MAE, whereas MLR predictions are less accurate and more negatively

biased than raw SIT-NYHOPS predictions (Figs. 3.5 and 3.6). Raw SIT-NYHOPS predictions are slightly more negatively biased for surge intervals ≥0.6 m at 15-24 h lead-time than BC-MLR. For example, SIT-NYHOPS has an expected ME of -0.09 m [-0.12 m to -0.04 m] compared to BC-MLR's ME of -0.05 m [-0.17 m to 0.02 m] (Fig. 3.5a). For 3-12 h and ≥0.6 m surge intervals, raw SIT-NYHOPS has a more slightly negative expected ME of -0.08 m [-0.03 to 0.14 m] compared BC-MLR's expected ME of -0.06 m [-0.20 m to 0.07 m]. Due to the limited sample of moderate storm surge intervals, moderate interval statistics are not elaborated here, but it is noted that binned moderate storm surge interval MAE have similar values (i.e., overlapping 95th percentiles).

For storm surge intervals ≥0.6 m, the expected MAE for forecasts made by BC-MLR at 3-12 h leads is 0.19 m [0.14 m to 0.27 m, 95th percentile range] while the raw SIT-NYHOPS expected MAE is similarly 0.19 m [0.17 m to 0.24 m] (Fig. 3.5b). For storm surge intervals ≥0.6 m made at 15-24 h lead times, BC-MLR has a slightly lower expected MAE of 0.17 m [0.13 m to 0.23 m] compared to raw SIT-NYHOPS expected MAE of 0.21 m [0.18 m to 0.22 m]. However, the 95th percentiles for all MAE bins ≥0.6 m for both sets of lead times overlap, thus these differences are not considered significant. Overall, the raw SIT-NYHOPS 95th percentile range for both sets of lead times is consistently smaller than MLR and BC-MLR, which indicates that raw SIT-NYHOPS predictions feature errors that are more consistent from interval to interval.

The surge from Sandy at The Battery falls into the comparison window and predictions for this event by all modeling systems are analyzed briefly. Three forecasts made by raw SIT-NYHOPS, MLR, and BC-MLR all valid for 24 hours after 1800 UTC 28 Oct 2012, 1800 UTC 29 Oct 2012, and 1800 UTC 30 Oct 2012 demonstrate that the prediction of Sandy's storm maximum surge was consistently and similarly under-predicted at 0-24 h leads for the three

consecutive 1800 UTC forecasts (Fig. 3.7). BC-MLR predicted a slightly more accurate storm maximum surge (2.33 m) than raw SIT-NYHOPS (1.98 m) or MLR (1.85 m) for the same lead time, but underpredicted the observed storm maximum surge of 2.83 m occurring at 0115 UTC 30 Oct 2012. The large underprediction of storm maximum surge by all modeling systems shown here was a result of an inaccurate atmospheric model forcing (Georgas et al. 2014), and not a symptom of a MLR deficiency. In fact, using 3-h NARR forcing to predict Sandy's storm maximum surge at The Battery using the MLR is within 0.12 m of the observed storm maximum (Fig. 3.8).

An extratropical storm surge event that fell within the comparison time and had a storm maximum storm surge at The Battery of 0.92 m at 0300 UTC 8 Nov 2012. Three forecasts made by raw SIT-NYHOPS, MLR, and BC-MLR valid for 24 h after 1800 UTC 6 Nov 2012, 1800 UTC 7 Nov 2012, and 1800 UTC 8 Nov 2012 demonstrate that the prediction of the storm maximum surge was similarly under-predicted by all modeling systems for the three forecasts spanning the event (Fig. 3.9). The 1800 UTC 8 Nov 2012 raw SIT-NYHOPS forecast of 0.50 m at 0500 UTC on 8 Nov 2012 underpredicted the observed storm maximum storm surge of 0.92 m at 0300 UTC 8 Nov, while MLR and BC-MLR captured the storm maximum surge more accurately albeit too early (0.62 m and 0.79 m) for the same forecast times.

d. Discussion and Conclusion

A methodology for the development of multi-linear regressions was trained and verified at Atlantic City, The Battery, and Montauk Point to predict 3-h storm surge. Although there have been statistical surge models developed in the past for the NY/NJ coastal region, there has been no formal attempts to train and verify a statistical model with multiple atmospheric reanalysis

datasets to ascertain the robustness of the regression. Salmun et al. (2011) demonstrated the forecast skill of their regression surge predictions at The Battery against an operational hydrodynamical model's using observed significant wave heights as a predictor. In our work, regression surge forecasts at The Battery were compared with an operational hydrodynamical model but using forecast data from atmospheric weather models as predictors. As a result, our work expands and improves on prior studies that used regression to predict storm surge in the mid-Atlantic region (Pore et al. 1974; Salmun et al. 2011; Salmun and Molod 2015) and develops a methodology that can be applied to other coastal stations for real time predictions.

The relationship between local surface atmospheric data and the magnitude of the storm surge is reflected in the regression can be conceptualized in terms of basic physical processes. The MLR approximates the station's storm surge as a linear response to prolonged wind stress components averaged over a adjacent boxed region and the instantaneous minimum in sea level pressure in this same region. The results from our MLR methodology demonstrate the physical sensitivities and time scales storm surge occurs on. Both the choice of the boxed region to collect atmospheric data and the degrees of freedom associated with each of the three predictors were determined based on maximizing a penalized likelihood function known as the Akaike Information Criterion (AIC; Akaike 1974). The results of maximizing penalized likelihood demonstrate that surge is a station-specific response to a local along-shore surface wind stress on 9-27 h time scales, which agrees with the well-studied physics of storm surge generation (e.g., Komar 1988). By assuming zero covariance between predictors, the zonal wind stress term ($A\tau_x$) has the largest independent effect on storm surge predictions by the MLR, followed by the pressure term (P), and then the meridional term ($A\tau_y$) at all three stations. The signs and magnitudes of regression coefficients are similar between different training datasets and show

the importance of a persistent easterly wind component for storm surge generation. The importance of the prolonged 10-m zonal wind in generating NY/NJ coast storm surges is similarly highlighted in Colle et al. (2010) and Pore et al. (1964, 1974). The commonalities in results between MLR coefficients and predictors demonstrate a high similarity in the physical processes that generate storm surge between all three stations along the New York/New Jersey coast.

Storm maximum surge events that exceeded the 95th (99th) percentile of all storm maximum surges were termed minor (moderate) events. Raw storm surge predictions by MLR exhibited negative mean error (ME) or an underprediction bias at the three stations for minor and moderate events. It is hypothesized that the underprediction bias could be attributed to the omitted dynamic effects and interactions from the wave set-up, as suggested in DiLiberto et al. (2011). However, the physical explanation for the MLR's bias cannot be established within the context of regression analysis alone and suggests future sensitivity work needs to be done with a numerical hydrodynamic model (NHM) to help explain forcing contributions, such as in Orton et al. (2012). For storms that occur over longer time scales (i.e., > 18-h), the Coriolis force becomes more important for surge generation when there is an along-shore wind flow by driving a water flux in the across-shore direction (Komar 1988). The Coriolis effect could be responsible for the alternating selection of either $A\tau_x$=0 h and $A\tau_x$=27 h in some configurations of data for Atlantic City due to different duration training storms. A more thorough examination of Ekman dynamics in storm surge generation in the NY/NJ Bight is necessary to understand the budgeted role of this forcing mechanism.

In order to remedy the underprediction bias in storm maximum surge predictions, the predictions were multiplied by the reciprocal of the MLR's training period correlation (1/R). This

multiplicative factor was applied in Pore et al. (1974) for storm maximum surge prediction and in Klein et al. (1959) for 2-m temperature prediction. In both applications, the technique was used successfully. In this study, the use of the multiplicative factor acted as a bias correction and reduced the ME for minor and moderate surge event predictions to approximately zero at the three stations in verification data. Additionally, the bias correction reduced the mean absolute error (MAE) by 10 to 15 % for verification period minor and moderate storm maximum surge events. The accuracy of the MLR is considered robust given that the regression coefficients remain consistent between different cross-validation periods and two reanalysis datasets.

The MLR was also compared to the raw (not bias corrected) NHM (SIT-NYHOPS) surge predictions at 0-24 h lead time for the 723 forecasts made daily at 1800 UTC for Oct.-March over 2010 to 2014. The bias-corrected predictions of minor and moderate storm surges by the MLR (BC-MLR) had similar accuracy to the SIT-NYHOPS forecasts. The MLR surge predictions were largely negatively biased compared to the raw SIT-NYHOPS forecasts at 3-24 h lead times, whereas the bias corrected MLR (BC-MLR) predictions had a near zero expected ME. However, the raw SIT-NYHOPS featured more consistent error metrics or smaller 95th percentile ranges compared to either the raw or bias corrected MLR.

e. Real-time Applications and Future Work

The similar accuracy between MLR predictions and the SIT-NYHOPs at 0-24 h leads motivated running the MLR for surge and water level prediction in real-time as a forecasting tool. Forecasts of surge and water level are made at 03, 09, 15,and 21 UTC forced using the 21-member Short Range Ensemble Forecasting (SREF) system and are available online at: http://wx.somassbu.org/products/surge/surgeregression.php. A screenshot of the webpage is

shown in figure 3.10a, which illustrates the layout of the website. Surge and water level time series are predicted using the four run cycles of the SREF at Atlantic City, New Jersey, The Battery in New York, and Montauk Point, New York for 24-84 h lead times. The previous two forecasts at each station are archived each day and are also available for viewing; additionally, water level time series are annotated with flooding thresholds provided by the local National Weather Service office (Fig. 3.10b).

Future work could concern generating MLR forecasts to enhance other pre-existing numerical storm surge ensembles at tide gage locations. The greatest limitation for the application of this regression approach is sample size. Due to the infrequent nature of surge events (0-10 per season), at least 20 y of continuous storm-surge time series is preferred for robust results. Considering that the verification-period sample size for the SIT-NYHOPs and MLR comparison is somewhat limited (only four years), future work is also necessary to better quantify the forecast predictability of the MLR. This work would involve a longer verification window with different deterministic/ensemble atmospheric modeling systems, such as SREF forecasts.

There is also potential to further improve on the regression methodology by incorporating more complex statistical methods. Since a linear model can explain a large fraction of variance (> 60.0%) in verification period observed 3-h surges at the three stations studied here, the possibility to add non-linear storm surge dynamics using a more sophisticated statistical technique remains. Some examples of this for storm surge prediction include the use neural networks (e.g., de Oliveira et al. 2009) or a fully non-linear model (e.g., Siek M. and Solomatine 2010). More simply, the pre-existing MLR methodology designed could be improved by using a more complex predictor selection technique that incorporates some aspect of predictor penalties

(e.g., LASSO regression) or classification (e.g., regression trees). For instance, storms could be classified into meteorological relevant categories and then relationships could be trained within each class. These more complex statistical techniques could potentially improve predictions of extreme storm surges, but the issue of a small sample size would remain. Generalized linear models could also be used with the prior assumptions of non-normal probability distributions to better fit the heavy tails of the surge distribution. Essentially, the use of the multiplicative factor modifies the regression model's Gaussian probability distribution to one that is more like a Cauchy distribution with heavier tails, so generalized linear modeling in this application could be promising.

Overall, these results demonstrate that even with the large improvements to numerical prediction of storm surge, statistical methods can still offer competitive forecast skill. In general, statistical methods applied to surge prediction are highly useful due to the highly linear and highly localized nature of surge generation. Forecasters could benefit from the application of the regression in real-time since numerical predictions of surge often fail to take into account local relationships that regression can represent (Müller-Navarra and Knüpffer 2010).

Chapter IV: Surge Climate Predictions

a. Surge Predictions using CMIP5 Global Climate Model Data

Climatological (25-y to 100-y) predictions of storm surge are made using global climate model (GCM) data with a multi-linear regression (MLR) to determine how future changes in the wind and pressure climatology may affect storm surge along the NY/NJ coast. The purpose of

this section is to compare the frequency and intensity of surges on interseasonal and climatological time scales between a historical (1979-2004) and a future (2006-2079) period.

Modeled time series of storm surge using GCM data were calculated by following the methods described in Chapter 2. Both modeled and observed surges at The Battery had similar amplitudes compared to observed surges (-1.0 m to +2.0 m). Approximately 99.0 % of both modeled and observed 3-h surge intervals were between -0.60 m to +0.60 m. Considering 99.0 % of the raw (not bias corrected) 3-h surge were between -0.40 m and 0.43 m, the use of bias-correction applied to atmospheric GCM data drastically improved the representation of modeled impactful surges.

In this section, the definition of a count is a storm maximum surge that is greater than 0.61 m; thus, a count represents an impactful surge event. The definition of a moderate coastal flood is when the storm maximum water level (tide + surge + SLR) exceeds 2.44 m above MLLW. All surge predictions were separated into a historical period and a future period for the months of November-March. The 25 cool season historical period began on 0000 UTC 1 January 1979 and terminated on 1800 UTC 31 March 2004. The future period began on 0000 UTC 1 November 2006 and terminated on 1800 UTC 31 March 2079. The 25 cool season historical period is compared to the last 25 cool seasons during our future period (2054-2079), and this period is termed the middle-late 21st century.

Three metrics are used to detect changes and compare storm surge activity between historical and future time periods.

- Counts: the number of impactful surge events during a certain period. Counts per season represent interseasonal surge activity.

- Skill score: the percent of overlap between modeled and observed surge histograms. A score of 100 % represent a pair of histograms that are identical in probability structure (Perkins et al. 2007).

$$\text{skill score} = \sum_{1}^{n} \min\left(z_1, z_2\right) \tag{4.2}$$

- Probability differences: the binned differences between two histograms. Statistical significance ($\propto = 0.05$) between binned probabilities is determined using a chi-square test when the bin sample size for both histograms under comparison exceeds five.

All statistical comparison are made at $\propto = 0.05$ and. In equation 4.2, n represents the number of bins used to calculate the histogram, and z_1 and z_2 are a pair of histograms of surge data that are under comparison. Histograms of surge are compared over a range of impactful storm surge levels defined from 0.6 m to 2.0 m in 0.2 m increments (7 bins). Surge levels less than 0.6 m are not considered since any coastal flooding is relatively minor.

1. Historical Period Surge Climatology

Predictions of 3-h surge and storm maximum surge are compared to observed data at The Battery to determine their skill on interseasonal time scales throughout the historical period (1979-2004 Nov.-March). To assess skill, we measure the model-to-model range (max/min), average counts per historical period and per season, and the probability of 3-h surge heights. The metrics calculated for observed data at The Battery from 1948-2012 serve as a representation of natural variability (Table 4.1).

Between 1979-2004 Nov.-March, The Battery observed 93 counts, whereas the ensemble mean predicted 62.2 with a model-to-model range from 45 to 95 (Fig. 4.1a). Without bias

correction of atmospheric data, the ensemble mean predicted 35 counts with an model-to-model range from 20-55; therefore, all quantities herein are reported with the use of bias corrected GCM data. Five GCMs (BCC, NORESM, CANESM, CNRM, and CCSM4) under predicted by 15-30 counts, and one GCM (GFDL-ESM2G) over predicted by 2 counts. Observed interseasonal counts ranged from 0 to 10 counts/season with an average of 3.98 counts/season, whereas the average model-to-model range was 0 to 7.33 counts/season with an ensemble mean of 2.47 counts/season (Fig. 4.1b). Because of relatively large interannual variations, there are large differences in count frequency between consecutive seasons; additionally, modeled interseasonal counts are insignificantly correlated with observed interseasonal counts during the historical period (Fig. 4.1b).

The probability of 3-h surge heights is realistically simulated in modeled surges (Fig. 4.2). Both modeled and observed impactful surge (≥ 0.61 m) heights occur infrequently (less than 1.0 % of all intervals). The ensemble mean has a skill score 92.0 % with a model-to-model range between 89.0 – 93 % in reproducing the probability of 3-h surge. There is a tendency for surge heights ≥ 0.80 m to be overpredicted by 0.01 % to 2.4 % compared to observed. However, these differences are considered small.

In order to characterize and later compare the intensity of storm maximum surge, histograms of storm maximum surges were normalized so that the probability of all bins summed to one. These distributions are referred to as intensity distributions and show how frequently storm maximum surge heights occur. Thus, the calculation of skill scores for intensity distributions demonstrate how similar the occurrence of modeled storm maximum surge heights are to observed data. For storm maximum surge, the ensemble mean skill score was 80.0% with a model-to-model range between 72.0 % to 86.0 % (Fig. 4.3).

There are similar probability errors between modeled storm maximum surge heights and the probability of 3-h surge; modeled surges have more >0.80 m surge peaks, but less overall counts (>0.61 m). The histogram in figure 4.3 demonstrates that the modeled 0.60 m-0.80 m surge peak bin has consistently lower probabilities (15.0 % - 25.0 %) compared to observed data. In contrast, the modeled 0.80-1.0 m surge bins tend to have larger probabilities (5.0 % - 20.0 %) compared to observed data. A similar probability error is reflected in the daily probability of 3-h surge heights (Fig. 4.2). However, due to a cancellation of binned mean probability error between GCMs, all surge bins of the ensemble mean intensity distribution (Fig. 4.2g) were statistically similar to observed data that had a sample size greater than five.

2. Future Period Surge Climatology

With the historical differences between GCM modeled and observed surges considered, the modeled future period (2006-2079 Nov.-March) storm surge climatologies are compared to the historical period. While future surge changes are studied only at The Battery, the results can be generalized for the nearby tide gages at Atlantic City and Montauk Point given the large decorrelation space of storm surge. For example, the 3-h surge at The Battery has greater than an r = 0.85 zero-lagged cross-correlation with either Montauk Point or Atlantic City 3-h surges. Overall, throughout the future period at The Battery, interseasonal counts in each of the six GCMs do not exceed natural or modeled data over the historical period.

Modeled future period storm surge exhibit relatively large interseasonal and interdecadal variations but exhibit no long-term trends (25 y to 100 y) in frequency, which is similar to the results found using observed surge data throughout the 20th century at The Battery in Zhang et al. (2000). A centered sliding window of length 10-y applied to count time series provides a

measure of decadal frequency. Modeled counts/decade ranged between 25-49. For comparison, the decadal count frequency in observed data between 1948-2012 ranged between 25 to 52 counts. The interseasonal counts during the future period are also similar to both the observed and modeled historical period data. For instance, the model-to-model range of interseasonal counts was 0 to 8.0 counts/season, which is close to the model-to-model range during the historical period (0 to 7.33 counts/season). For comparison, the observed interseasonal counts ranged from 0 to 10 counts/season between 1948-2012.

Linear regressions were fit to centered sliding windows of length 10-y, 20-y, and 30-y of seasonal count data throughout the future period to identify trends (Fig. 4.4b-d). Each point in the sliding window corresponds to the trend over the preceding and following n (2 yr)$^{-1}$, where n is the interval in question. Statistical significance of these trends are determined by comparing the slope to the level that would reject the null hypothesis that the slope differs from zero of a t-test at $\propto = 0.05$.[1] As the sliding window period used to calculate the linear trend increases from 10-y to 30-y, the magnitude of trend consistently diminishes, which demonstrates that count variability is greater on shorter time scales. For example, 10-year trends fit to seasonal counts vary between -0.55 to +0.57 counts/season, while 20-y trends vary between -0.26 to +0.17 counts/season, and 30-y trends vary between -0.11 to +0.11 counts/season (Fig. 4.4). For reference, using observed data between 1948-2012 at The Battery, 10-y trends varied between -0.49 counts/season to 0.73 counts/season and had similar values to that of modeled data for 20-y and 30-y trends.

[1] A Monte Carlo approach was used to determine the linear trend that would reject the null hypothesis of a t-test given that the sample of the sliding window was less than 30.

To determine if trends in modeled counts are similar or unrelated between other GCMs, the correlation matrix of the each modeled 10-y, 20-y, and 30-y trends was calculated (Fig. 4.5). The average correlation between some GCM's trend increase from 20.0 % to 40.0 % to 60.0 % to 80.0 % as the sliding window period increases from 10-y to 30-y. This suggests the presence of similar patterns in multi-decadal surge activity. For instance, using 20-y and 30-y centered sliding windows, the trends calculated for the CCSM4, GFDL, and CNRM are significantly positively correlated (r = 0.40 to 0.80) to each other, whereas the remaining three models do not have a significant correlation. These three GCMs demonstrate a similar insignificant increasing trend of +0.10 to +0.15 counts/season from 2040-2054 and then an insignificant decreasing trend (-0.05 to -0.10 counts/season) from 2054-2060 (Fig. 4.6). However, using sliding windows of 40-y or larger (not shown), the mid-century 2040-2054 activity and middle-late inactivity (2054-2060) in surge counts shown in these three GCMs was not detectable and was on the order of 0.001 counts/season.

To test the hypothesis that surges may become more intense into the 21st century due to an increased frequency of intense cyclones in the western north Atlantic (Colle et al. 2013), the annual maximum surge (AMS) height was calculated for all six GCMs and observed. In a scenario where more intense cyclones occurred more frequently over the U.S. east coast region, one would expect to see an increase in the average annual maximum surge height and generally larger variation about the mean AMS. However, throughout the future period (2006-2079 Nov.-March), there was no significant variation in the mean (Fig. 4.8a) or standard deviation of the AMS (Fig. 4.8b). The average modeled AMS throughout the future period was 1.01 m ($\bar{\sigma} = 0.30$ m) with a model-to-model range of 0.90 m to 1.05 m. As a reference, the average observed AMS

at The Battery between 1948-2012 was 0.96 m ($\sigma = 0.27$ m) so there is little change in this statistic over the 21st century (Fig. 4.8a-c).

While the AMS does not appear to change throughout the 21st century, a more thorough way to examine intensity shifts in surge is to compare storm maximum intensity distributions between equal length periods. For example, while the AMS may not exhibit an increase, the intensity shift could be present in another range of the surge distribution, and therefore would show up in a histogram comparison. However, statistical significance in the ≥ 1.0 m surge bin differences could not be determined because of the small binned sample size. Overall, there are small shifts (<1.0 m) in the expected surge intensity between the mid.-late 21st and the historical period, with means that are insignificantly different (Fig. 4.7). The CCSM4 has the most well defined increase (0.12 m) in storm maximum surge intensity between periods, but the other GCMs show changes between 0.01-0.05 m in expected storm maximum surge height.

Intensity shifts in storm maximum surge in the future period are also searched for by calculating the 50th, 75th, and 95th percentile of modeled impactful surge heights in 30-y sliding windows (Fig. 4.9). Overall, the ensemble mean's 50th and 75th percentiles of storm maximum surge heights exhibit little variation throughout the future period. The ensemble mean's 95th percentile of storm maximum surge increases from 1.36 m to 1.42 m but this represents only a 4.0 % change. The largest increase is associated with the CCSM4's 95th percentile, which increased from 1.29 m to 1.63 m by 2063 with represents a 25.0% change. However, the changes to the 95th percentile have to be taken with caution given that they are highly sensitive to different samples; further, the sample size used to calculate these percentiles are generally less than 15. More confidence is placed in the 50th and 75th percentiles that have larger sample sizes

(>30); the changes over time associated with these percentiles agree with the changes described above for the AMS and expected intensity distributions shifts.

The trends in frequency and intensity distributions of moderate coastal flooding (WL+ tide + SLR) similarly follow the results for surge (i.e., no change beyond natural variability) since the number of surge events is directly proportional to the number of moderate coastal floods in any given period. In general, moderate coastal flooding events occur less frequently than impactful surge events because of the variations in water level introduced by the tidal stage. During the historical period, The Battery observed eight water levels that met moderate coastal flood criteria (2.44 m above MLLW), whereas the ensemble mean predicted seven (Fig. 4.10). In a similar matter to the results described above for impactful surge, over the seven decades that span 2006-2079, there were no significant trends in the frequency or intensity of moderate coastal flooding events.

3. Impacts from Sea Level Rise

Since the total water level is also influenced by sea level rise (SLR), the number of moderate coastal floods per season is projected to increase into the 21st century. When a regional SLR scenario consistent with the global warming expected with RCP8.5 was added to 3-h predicted water levels (Zhang et al. 2014), the expected number of moderate coastal floods per decade increased at an exponential rate (Fig. 4.10). By fitting the ensemble mean difference (counts with SLR – counts without SLR; Fig. 4.10c) in moderate coastal floods per decade to an exponential model, the expected number of moderate coastal floods attributed solely to SLR in a given future decade (beyond 2009-2019) can be calculated. The SLR impact quantifies the

difference in moderate coastal flooding events per decade attributed to SLR while ignoring any storm changes attributed to global warming.

$$SLR\ Impact\ (year) = 1.104\ e^{0.53 * \frac{[year - 2009]}{10}} \quad (4.3)$$

Based on this relationship (Eq. 4.3), which is the line of best fit through figure 4.10c, there is an expected 41 more moderate coastal flooding events attributed to SLR in 2069-2079 compared to 2009-2019.

The annual probability (Eq. 4.4) and return period (Eq. 4.5) or the expected frequency of the storm tide ($H = tide + surge$) set by Sandy ($H_{Sandy} = 4.02$ m above MLLW) has been of interest to calculate.

$$Annual\ Probability[H(t) \geq H_{Sandy}(y)](y) = \frac{m}{n} \quad (4.4)$$

$$Return\ Period[H(t) \geq H_{Sandy}(y)](y) = \frac{1}{Annual\ Probability(y)} \quad (4.5)$$

Where both the return period and the annual probability are functions of the year because of SLR, *n* is the number of years on record, and m is the number of occurrences of a storm tide that exceeds $H_{Sandy}(y)$. In the 100-y period between 1979-2079 without SLR, a storm tide greater than 4.02 m above MLLW did not occur for any GCM; therefore, the return period was intractable for these data. However, the storm tide necessary to exceed Sandy's water level (tide + surge + SLR) will become less by the level of SLR. For instance, by 2079 the predicted regional SLR is ~0.60 m; therefore, an equivalent depth coastal flood to that of Sandy's at The Battery in the year 2079 would require a peak storm tide of 3.46 m instead of 4.02 m (Fig. 4.11a).

Given the realism of the modeled surge and water level time series on interseasonal time scales, we take these data and calculate the return period of an equivalent coastal flood to that of Sandy at The Battery given sea level rise. In our calculations, each iteration of the tidal bootstrap was considered a separate year; thus n = 1,000. Due to the infrequent nature of extreme events and tidal uncertainty, the calculated annual probability can vary from season to season. Therefore, the annual probability (Eq. 4.4) was fit to an exponential model using least squares to ensure a smooth time function and then the reciprocal of the fitted model was calculated to calculate the return period (Eq. 4.5). An exponential model was deemed appropriate given the acceleration of predicted regional SLR (cf., Fig. 1.5). The calculation of return period for the coastal flood depth set by Sandy in this manner takes into account both storm changes attributed to global warming, SLR, and tidal uncertainty. There is a consensus between four of the six GCMS that Sandy's return period in 2012 was between 100-y to 250-y. Two GCMs (CANESM and CNRM) had very few exceedances prior to 2050 and this highly affected their return period calculation. In general, return period calculations are highly sensitive for extreme (infrequent) events because of limited sample. By 2079, the model-to-model range predicts that a Sandy-like coastal flood, which would require a 3.46 m storm tide, would have an expected return period between 10-85 y.

b. Track and Surge Approach

Future surge changes are the response to changes in the movement, intensity, and frequency of cyclones since cyclones generate surges; therefore, it is useful to study surge from a cyclone track perspective (Zhang et al. 2000; Colle et al. 2010). In the previous section, we used six GCM's from the fifth generation of the Coupled Model Intercomparison Project (CMIP5) to

predict surge with a multi-linear regression (MLR). During the historical period, the MLR forced with atmospheric GCM data predicted less impactful surge events (≥0.61 peak surge) than observed even with a bias correction technique applied to atmospheric data. Since the verification of the MLR demonstrated unbiased (near zero mean error) impactful surge predictions using CFSR (Chapter 2), the underprediction of impactful surges must originate from GCM data itself. Identifying biases in GCM data can allow us to pose questions for future studies and acknowledge pre-existing challenges with data that might preclude certain results. For instance, an underprediction in the occurrence of impactful surge could be motivated by an under prediction of nearby cyclones.

Observed impactful surge events (≥0.61 m storm maximum surge) were associated with ~8.0 % or 155 tracks of CFSR tracks in our study region during the historical period. Therefore, the skill in predicting the climatological cyclone field does not imply accurately modeled surges. By classifying cyclone tracks into Miller Type A and Miller Type B classes, we reduce the number of tracks that need to be analyzed at a time, and we can determine if a particular cyclone class is more or less related to the development of impactful surge events at The Battery. Tracks are also considered with respect to the concurrent surge height at The Battery; for example, impactful surge-generating cyclone tracks have at least one track point at a time when a >0.61 m surge event occurs at The Battery.

All cyclone tracks considered for further analysis had a minimum duration of 24-h within the black-dotted boxed region in figure 4.12 and collectively represent the Nov.-March cyclone track climatology. These cyclone tracks, composed of greater than or equal to four cyclone features in time and space were matched in time to 3-h surge data at The Battery. Cyclone tracks were then classified as either Miller Type A or Miller Type B tracks since these classes of tracks

fit into a conceptual model of storm tracks over the Atlantic coastal region proposed by Miller (1946). Cyclones that did not meet either class were termed "other" so that the total climatological track field is the sum of all three classes.

In order to classify a cyclone as a Miller Type A, the cyclone had to form near or within the Gulf of Mexico, which is indicated as the red-hatched region in figure 4.12, or pass through this boxed region for at least 24-h. The cyclone additionally had to move within the red polygon region along the U.S. east coast for a specific fraction of its total track time. This rule was defined by a quantity called the lifetime ratio (Eq. 4.6)

$$\text{lifetime ratio} \geq \frac{\text{time within polygon}}{\text{total track lifetime}} \tag{4.6}$$

By choosing a lifetime > 0.5, the classification of Miller Type As were more restrictive since the cyclone had to persist for a longer time within the red polygon along the U.S. east coast. Likewise, by selecting a lifetime ratio < 0.5, the classification were more permissive as tracks could persist for less time within the polygon but still be considered Miller Type A tracks. An optimal lifetime ratio was determined by perturbing the parameter and each time manually identifying cyclone tracks that did not resemble Miller Type As (i.e., track along or within 400 km of the U.S. east coast). The lifetime ratio that produced the least amount of falsely classified Miller Type As was selected for the final classification rules. The value that minimized falsely classified tracks was found to be >0.40. The classification of cyclones as Miller Type As with the rules described above is similar to the rules set forth in Marciano et al. (2015). Similarly, Marciano et al. (2015) imposed a genesis requirement of at least 400 km from the Gulf of Mexico coastline and a subsequent track requirement within 400 km from the United States eastern seaboard up to the latitude of around 39 N.

For a surface cyclone track to be classified as a Miller Type B, it had to track within the boxed region over the Great Lakes outlined in yellow for a minimum time of 24-h (Fig. 4.12). In addition, within 36-h of the last track point in the yellow box, another cyclone had to develop over the mid-Atlantic coastal region, which is indicated as the blue-hatched region in Fig. 4.12. A time difference of 36-h between the last track point within the Great Lakes region and the genesis of the mid-Atlantic cyclone was determined by maximizing the number of events that could be time matched with a surge event at The Battery while minimizing the number of falsely associated cyclone redevelopments.

All tracks that did not meet Miller Type A or Miller Type B classes but were considered for study were classified as "other" tracks. Thus, the sum of all classified tracks (i.e., Type A + Type B + other) sum to the climatological track field. The track density per cool season (Nov.-March) for each of the three cyclone classes in CFSR and GCMs during the historical period is illustrated in figure 4.13. During the historical period, there were a total of 1929 CFSR tracks considered for classification of which 388 tracks were classified as Miller Type A events, 589 tracks were associated with Miller Type B, and 952 tracks other.

1.Historical Evaluation of Classified Tracks

Overall, the GCMs underpredicted the frequency of Miller Type A, Miller Type B, and the "other" cyclone class by 100-200 tracks (6.0 % - 15 % compared to CFSR) during the historical period (Fig. 4.13). Related studies have shown a similar underprediction of the climatological Nov.-March cyclone frequency over North America (Chang et al. 2012) and over the western north Atlantic (Colle et al. 2013) in GCMs. In CFSR tracks, Miller Type As occurred ~15 per season and Miller Type Bs ~24 per season (Table 4.2). Modeled Miller Type As

occurred 9-15 per cool season (Nov.-March), whereas Miller Type Bs occurred more frequently, between 13-23 per cool season. The seasonal frequencies of observed Miller Type A and Miller Type B are similar but slightly larger by 2-5 per season to the results of Miller (1946); however, Miller (1946) was based off a limited sample and a data sparse observing network at the time of that study.

The track density correlation and absolute mean difference in grid point track density (TD) were calculated using CFSR tracks to quantify historical skill of classified cyclones (Table 4.4). Overall, GCMs were able to realistically simulate the three classified fields track density but underpredicted the track density maxima by 10.0-15.0% (Fig. 4.13). Track density correlations were similar and high between models, ranging from 0.89-0.90, and TDs ranged from 0.40-0.56 cyclones per season per grid point (Table 4.4). The track density correlations for Miller Type A feature comparatively the lowest correlations with CFSR tracks and ranged from 0.85 to 0.87. Miller Type B and other class tracks had similarly high correlations ~0.90 and TDs between 0.16-0.25 compared to that of Miller Type As and climatology. Most noticeably, the track density maxima over the Great Lakes for Miller Type B and over the Gulf of Mexico for Miller Type A storms was underpredicted (15-25 %) compared to the climatology track density (Fig. 4.13).

c. Track and Surge Relationships

The relationship between classified cyclone tracks and observed surge is qualitatively established by comparing the classified track frequency within the study region (100.0 W to 50.0 W and 20.0 N to 50.0 N) to the occurrence of impactful surge events (counts) at The Battery. During impactful surges at The Battery, the mean sea level pressure is typically weakly defined

and often involves a broad area of surface low pressure over the eastern U.S. (Colle et al. 2010). Considering this, differently classified tracks (i.e., Miller Type A or Miller Type B) can sometimes occur simultaneously during the same surge event. With this in mind, the number of tracks that occurred with the number of surges is reported without consideration to the double counting of surge events.

There were 155 CFSR tracks (~8.0 % of the climatological CFSR tracks) that were classified and occurred during 75 observed impactful surge events between 1979-2004 Nov.-March (historical period). The remaining 18 observed surge events during the historical period could not be matched in time with a CFSR track because of tracking difficulties. The seasonal surge-generating track density is associated with coastal tracks that originate near the Gulf of the Mexico and move northeastward along the U.S. east coast (Fig. 4.14). As these tracks move northward along the U.S. east coast they converge near the mid-Atlantic coastline; this is shown as a track density maximum of 1-2 tracks per 50,000 km^2 per season located off-shore the Delmarva peninsula (Fig. 4.14). A similar coastal track and track density maximum off the mid-Atlantic coast during >0.60 m surge events at The Battery was illustrated in Colle et al. (2010).

Miller Type A and Miller Type B tracks occurred during 61 of the 75 observed impactful surge events during the historical period, whereas the remaining 14 occurred during "other" class tracks. During the historical period, 39 Miller Type A tracks occurred during 38 observed impactful surges at The Battery, while there were 56 Miller Type B tracks that occurred during 42 observed impactful surge events. Some surge events were associated with multiple cyclone tracks in the study region since there were 75 observed surge events at The Battery period that occurred in time with 155 cyclone tracks. However, many of the surge events associated with

multi-cyclone tracks had one or more upstream or downstream cyclone that was reasonably unrelated (> 500 km distance from The Battery) to the surge height at The Battery.

The classified surge-generating track frequency was also realistically simulated in GCMs but underpredicted by 6.0 %-16.0 % compared to CFSR classified surge-generating tracks (Table. 4.3). Classified Miller Type A surge-generating tracks appear similar in location and trajectory to observed events (Fig. 4.14). In both the CFSR and GCMs, Miller Type A tracks that generate impactful surges generally form either in the western Gulf of Mexico or east of Florida. For the coarser GCMs (BCC, CANESM, NORESM), tracks appeared to be more zonally orientated than in CFSR data, whereas, in higher resolution GCMs (CCSM4 and CNRM), were more meridionally orientated.

There are large track frequency differences between CFSR Miller Type B and the GCM's Miller Type B surge-generating tracks. Generally, the frequency of GCM tracks that originate in the western Gulf of Mexico are largely (20.0 %-75.0 %) underpredicted compared to CFSR, whereas tracks associated with the coastal redevelopment off the Mid-Atlantic are comparatively less under predicted compared to CFSR (Fig. 4.14). The underprediction of cyclone tracks over the western Gulf of Mexico is also evident in the climatological TR for GCMs in figure 4.14. For the climatological field (i.e., not surge-generating), GCMs appear to have difficulty resolving cyclogenesis events that are downwind of the Rocky Mountains located over the central and southern U.S. Midwest; therefore, this underprediction also appears for Miller Type B surge-generating tracks (Fig. 4.14). Additionally, the GCM's Miller Type B track underprediction becomes larger in coarser horizontal resolution models (Fig. 4.13c-o).

As demonstrated in the beginning of Chapter 3, the largest variations in the frequency of surge events occur on the interseasonal (1-9 years) time scale; these variations are created by

variations in the cyclone fields. Considering surge-generating tracks compromise a small fraction (8.0 %) of the climatological cyclone track in the study region and the variations in cyclone frequency can be large (50-100 tracks per season), the relationship between all tracks and the frequency of surge events is difficult to establish. In fact, the interseasonal frequency of CFSR climatological cyclone tracks in our study region was insignificantly correlated ($r = -0.08$, $p = 0.70$) to interseasonal counts (Fig. 4.17a). Likewise, the interseasonal frequency of CFSR Miller Type B tracks were also insignificantly correlated ($r = -0.003$, $p = 0.98$) to interseasonal counts (Fig. 4.17b). Only the frequency of Miller Type A tracks demonstrate a positive correlation ($r = 0.28$) to interseasonal counts, but the correlation is still insignificant ($p = 0.16$).

Modeled impactful surges and cyclone tracks in the six GCMs exhibit similar correlative relationships to what is described above for CFSR (Fig.4.18a-c). Both the interseasonal frequency of Miller Type B and climatological tracks have near zero correlations with interseasonal counts (Fig. 4.18a,c). However, interseasonal frequency of Miller Type As have a similar magnitude positive correlation ($r=0.27$, $p=0.21$) to the relationship described above using CFSR tracks with observed surges (Fig. 4.18b). Considering that both Miller Type A and Miller Type B tracks contribute to the majority of observed surges, the variation in interseasonal counts appears to be influenced primarily by the seasonal frequency of Miller Type A cyclones. This agrees with the hypothesis put forth by Colle et al. (2010): that moderate surge events at The Battery were forced by coastal cyclones resembling Miller Type A tracks, and the interseasonal variations of counts were related to the Southern Oscillation (ENSO), which influenced the frequency of Miller Type A events. Colle et al. (2010) found that the years with the most counts at The Battery occurred during ENSO events. During years with positive ENSO, the frequency

of coastal cyclones like Miller Type A events has been shown to be more prevalent than compared to either La Nina and neutral Southern Oscillation years (Hirsch et al. 2001).

Since the seasonal frequency of Miller Type A tracks are positively correlated to counts, systematic changes to the intensity of Miller Type A tracks could influence the expected storm maximum surge at The Battery. Marciano et al. (2015) demonstrated increases in cyclone intensity of Miller Type A storms in the form of a decrease of minimum sea level pressure and increase of the 10-m wind field in a warming global climate. The changes in future storm intensity of Miller Type A storms were dynamically attributed to increases in diabatic potential vorticity; thus, supporting the hypothesis that enhanced latent heating was responsible for the regional increase in intense (<980 hPa) storms posed by Colle et al. (2013). However, here it is shown that the changes to the mean average track MSLP and maximum intensity (minimum track MSLP) of Miller Type A storms are not consistent between GCMs and do not relate to the weak shifts in the modeled storm maximum surge intensity distributions throughout the future period.

Here we find that, while some GCMs project an intensification of the average along-track Miller Type A MSLP, some GCMs also project a weakening; thus, the changes are inconsistent among GCMs between 1979-2004 and 2054-2079. The average along-track Miller Type A in CFSR data was 987 hPa and GCMs underpredicted this during the historical period by 3-15 hPa (Fig. 4.19). An overall underprediction of cyclones in GCMs has also been noted in Colle et al. (2013) and Chang et al. (2012). The differences in shape of the average track MSLP distribution are too peaked in the GFDL and NORESM (Fig 4.19) compared to CFSR and suggest these models may be inaccurately simulating some dynamical aspects of cyclogenesis. The large differences in average along-track MSLP is expected considering the steps that had to be taken to

remove the bias from the GCM data. GCM-relative changes per period (historical vs. mid.-late 21st) show an increase (weakening) to the expected along-track Miller Type A MSLP for the BCC, GFDL, CNRM, CANESM and decreases (intensification) for the CCSM4 and NORESM (Fig. 4.19).

However, surge is a local response to surface wind stress and pressure fields; therefore, the along-track pressure may not represent local wind and pressure changes that affect surge generation. When the average along-track pressure for the segment of Miller Type A tracks within the boxed mid-Atlantic region (blue-hatched region in figure 4.12) is calculated (Fig. 4.20), the GCM-relative changes per period are weaker (-3.0 to -3.0 hPa) compared to the average total-track changes (-7.0 to 7.0 hPa). The expected track MSLP changes along the segment within the mid-Atlantic region show an increase (weakening) for the CCSM4, NORESM and decrease (intensification) for the BCC, CNRM, GFDL, and CANESM. The MSLP changes do not agree between the average along-track and average track segment within the Mid-Atlantic boxed region; GCMs that predict a deeper expected along-track MSLP, predict a weaker expected along-track MSLP along the mid-Atlantic region. Further, the period shifts in expected track MSLP and expected track segment do not correspond with the increases/decreases in the storm maximum surge (e.g., Fig. 4.7, Fig. 4.9) since one would expect deeper more intense storms locally to drive larger surges. Additionally, the minimum along-track pressure or maximum intensity of the cyclone for the segment of Miller Type A track within the boxed mid-Atlantic region (Fig. 4.21) features similar changes to the average along-track segment within the mid-Atlantic box MSLP. Given the overall disagreement between the cyclone intensity distribution shifts locally and along-track, it cannot be concluded that there is a general intensification of cyclones in the middle-late 21st century compared to the historical period.

Overall, cyclone pressure changes are weak and in disagreement; therefore, they do not demonstrate future trends beyond that of natural variations.

Potentially, a shift in cyclone track could alter the intensity distribution of surges at The Battery during 2054-2079 compared to the historical period. Studies have demonstrated a poleward expansion of the northern hemispheric surface cyclone track attributed to a reduced meridional surface temperature gradient produced by polar amplification in a warming global climate (Lambert et al. 1995; Chang et al. 2012; Colle et al. 2013). A northward shift in the expected cyclone track could alter the duration, intensity, and fetch of surface wind forcing that generates surges. Thus, the shift in cyclone track would be apparent in the track density of surge-generating track.

However, the differenced impactful surge-generating track density between the mid.-late future and the historical period demonstrate no consistent northward shift in surge-generating cyclones between periods (Fig. 4.22). There is a consensus of a weak reduction in the seasonal surge-generating track density by -0.2 to -0.4 cyclones per 50,000 km^2 over the western north Atlantic region in most of the GCMs. Both the NORESM and CCSM4 tracks, which had the largest increases to the 95th of storm maximums surge during 2054-2079, also have an increase of 0.2 to 0.4 cyclones per 50,000 km^2 to the seasonal surge-generating track density along the mid-Atlantic coast. Given that the NORESM and CCSM4 both show a track density enhancement over the mid-Atlantic coastline, it is speculated that the expected shift of cyclone tracks more onto the U.S. east coast could increase the expected surge height at The Battery. For the CCSM4, the differenced (2054-2079 minus 1979-2004) surge-generating seasonal track density of Miller Type A storms demonstrates a poleward shift (Fig. 4.23); however, the Miller Type A frequency does not demonstrate any large change in frequency (> 5 cyclones per period).

Therefore, considering the lack of frequency and intensity changes to Miller Type A storms, the CCSM'4 shift toward more intense surge events appears to be related to a poleward shift in the Miller Type A storm track.

In conclusion, given the lack of consistent trends in the frequency, intensity, and spatial distributions of cyclone tracks, future changes in the modeled future (2054-2079) surge climate appear to remain within the bounds of modeled historical variability and exhibit no pronounced change.

Chapter V: Conclusions

a. Regression Model

A multiple linear regression (MLR) approach was developed for Atlantic City, The Battery, and Montauk Point along the New York/New Jersey (NY/NJ) coast to predict storm surge every three hours. At all stations, the MLR approach was successful considering it explained > 60 % of the variance in verification period 3-h surge heights, had a mean absolute error between 0.30 – 0.40 m, and a mean error around zero for storm maximum surge predictions that met the 95th percentile of storm maximum surge heights. Further, MLR was trained and tested with various configurations of reanalysis data with similar accuracy; thus, the results are considered robust to different datasets. The results of the regression model development were published in Roberts et al. (2015).

The original motivation behind the development of the regression was to predict surge heights using 6-h climate model data for a few sites over several decades into the 21st century. It was realized that the accuracy between MLR predictions and the SIT-NYHOPs at 0-24 h leads

was similar and motivated the use of MLR for surge and water level prediction in real-time as a forecasting tool. Forecasts of surge and water level are made at 03, 09, 15, and 21 UTC forced using the 21-member Short Range Ensemble Forecasting (SREF) system and are available online at: http://wx.somassbu.org/products/surge/surgeregression.php

Overall, these regression results demonstrate that even with the large improvements to numerical prediction of storm surge, statistical methods can still offer competitive forecast skill. In general, statistical methods applied to surge prediction are highly useful due to the linear and highly localized nature of surge generation. Forecasters can benefit from the application of the regression in real-time since numerical predictions of surge often fail to take into account local relationships that regression can represent (Müller-Navarra and Knüpffer 2010).

b. Surge Predictions using Climate Models

Climatological (25-y to 100-y) predictions of storm surge were made using global climate model (GCM) data with a multi-linear regression (MLR) to determine how future changes in the wind and pressure climatology could affect storm surge along the NY/NJ coast. Identifying trends in storm surge into the 21st century naturally extends from research on the future extratropical cyclone track variations (e.g., Chang et al. 2012; Colle et al. 2013; Marciano et al. 2015) as storm surge is a regional phenomenon associated with extratropical storms. While surge changes were studied only at The Battery in this work, the results can be generalized for the nearby tide gages at Atlantic City and Montauk Point given the large decorrelation space of storm surge.

A relatively small ensemble of GCMs composed of six members was employed to determine future surge changes at The Battery solely attributed to future changes in the

atmosphere. Unfortunately, many CMIP5 GCMs that had 6-h data did not have winds on a sigma pressure vertical coordinate; thus, the methodology to reduce the winds to 10-m was not applicable for these GCM's data and they could not be used for surge predictions. Due to the limited member size of the ensemble and highly varied surge predictions, this work primarily highlights some of challenges faced using global climate model data to predict surge and some fundamental deficiencies in 6-h GCM data that need to be improved.

There was a general underprediction of cyclone frequency and intensity in climate model data compared to the tracks calculated using the Climate Forecast System Reanalysis, which has also be shown in other studies (Chang et al. 2012; 2013; Colle et al. 2013). As a result, atmospheric data from GCMs exhibited large biases in mean and standard deviation compared to the CFSR, thus motivating the use of a simple bias correction (Xu and Wang 2014). After applying the bias correction to atmospheric data from GCMs, the predicted surges had realistic amplitudes, but the seasonal frequency of impactful surge events (≥ 0.61 m) was still underpredicted by 2-5 events in all GCMs. The use of bias correction is controversial given that it implies the physical processes that drive storm surge are not correctly modeled.

Overall, throughout the future period at The Battery, the interseasonal variations in the number of impactful surges (counts) in each of the six GCMs did not exceed natural or modeled variations calculated over the historical period (1979-2004 Nov.-March). Sliding windows of length 10-y to 30-y were used to identify trends in count data; however, there was no coherent 21st century trend in count frequency in any of the six GCMS. Three different metrics were used to assess whether or not surges were becoming more intense into the 21st century at The Battery: the annual maximum surge (AMS), storm maximum intensity distributions, and percentiles

calculating using 30-y sliding windows. All three metrics demonstrated little change in the intensity of storm maximum surge heights between 2054-2079 and 1979-2004.

Regardless of any changes to storm surge intensity, sea level rise (SLR) was found to be the largest factor that affects the frequency of moderate coastal floods, which is what incurs flooding damage. SLR masks any of modeled future trends in surge occurrence or intensity and induced an exponential shift in the frequency of damaging moderate coastal floods by the late 21st century (2079 and onward). With the addition of regional SLR consistent with Zhang et al. (2014) to 3-h predicted surges, the ensemble mean predicted 41 more moderate coastal flooding events attributed to SLR in 2069-2079 compared to 2009-2019. More concerning, the storm tide necessary to exceed the historical Sandy flood (tide + surge + SLR) will become less by the level of SLR. For instance, by 2079 the predicted regional SLR is ~0.60 m; therefore, an equivalent depth coastal flood to that of Sandy's at The Battery in the year 2079 would require a peak storm tide of 3.46 m instead of 4.02 m. To determine the expected frequency of this even given SLR, the return period of an equivalent depth Sandy coastal flood at the Battery was calculated for all GCMs. In 2012, four out of the six GCMs estimated the magnitude of event to have a return period between 100-300 y, and, with SLR, all six GCMs predicted an inter-model range that a equivalent coastal flood would have a return period between 10-85 y by 2079.

In order to explain variations in surges and quantify relationships between the cyclone track and surge, surge data was matched in time to cyclone features part of nearby surface cyclone tracks. Surface cyclone tracks were automatically tracked in the mean sea level pressure (MSLP) field of the six GCMs and Climate Forecast System Reanalysis using the Hodges cyclone-tracking algorithm. Tracks were then classified as either Miller Type A, Miller Type B,

and "other" following a set of rules based off Miller (1946). The rules for Miller Type A were based off the methods in Marciano et al. (2015) and invented for Miller Type B cyclone tracks.

It was shown that both Miller Type A and Miller Type B tracks were realistically simulated, but consistently underpredicted by approximately 10.0%-15.0%. In CFSR tracks, on average Miller Type As occurred 15 per season and 24 per season for Miller Type Bs. In the study region (100.0 W to 50.0 W and 20.0 N to 50.0 N), modeled Miller Type As occurred 9-15 per cool season (Nov.-March), whereas Miller Type Bs occurred more frequently than Miller Type As between 13-23 per cool season. The track density maxima over the Great Lakes for Miller Type B and over the Gulf of Mexico for Miller Type A storms were more largely underpredicted (15-25 %) compared to the climatology or Miller Type A classes. Further, the simulation of Miller Type B tracks in GCMs appeared to be sensitive to horizontal model resolution. The regional dependence of cyclone skill is hypothesized to be influenced by geographical effects (i.e., mountain ranges, Gulf of Mexico), which has been shown to affect cyclone development dynamically. Future work could involve using the cyclone classification rules to determine the GCM's classified cyclone skill in more depth, with more ensemble members, and in idealized experiments that could attribute GCM classified cyclone skill to dynamics.

There were 155 CFSR tracks, which compromised ~8.0 % of the climatological CFSR tracks, that were classified as either Miller Type A, Miller Type B, or other and occurred during 75 observed impactful surge events between 1979-2004 Nov.-March. The impactful surge event track in both CFSR and GCMs involved two track clusters: one over the U.S. eastern seaboard and one over the Great Lakes, which represent tracks similar to Miller Type A and Miller Type B

events, respectively. Miller Type A and Miller Type B tracks occurred during 61 of 75 observed impactful surge events during the historical period, and the other class tracks occurred during the remaining 14. Some classified cyclone tracks occurred at the same time during one event; therefore, the total number of tracks is greater than the number of surge events. During the historical period, 39 Miller Type A tracks occurred during 38 observed impactful surges, whereas 56 Miller Type B tracks occurred during 42 observed impactful surge events. The remaining tracks were considered spurious and unrelated to surge generation at The Battery through manual inspection.

Classified cyclone frequencies were related to the frequency of impactful surge events to determine how variations in the seasonal frequency of cyclones are related to surges. In a previous study on the surge climatology at The Battery (Colle et al. 2010), it was hypothesized that the interseasonal frequency of Miller-Type-A-like-tracks explained the majority of variations in interseasonal surge frequency at The Battery, which is likely influenced by the Southern Oscillation (ENSO) (Hirsh et al. 2001. Here we find that, the interseasonal frequency of Miller Type A tracks are positively correlated (r=0.28) to the interseasonal surge frequency, thus supporting Colle et al. (2010)'s hypothesis. The seasonal frequency of climatological and Miller Type B cyclones showed no relationship between the seasonal surge variations. Similar relationships were exhibited in modeled impactful surges and cyclone tracks in the six GCMs between 1979-2079.

The impact of the poleward expansion of the cyclone track (e.g, Chang et al. 2012; Colle et al. 2013) was briefly investigated in its relation to surge-generating cyclone tracks at The Battery. An overall weak reduction -0.1- 0.2 cyclones per 50,000 km^2 was found in the number of seasonal surge-generating cyclone tracks over 2054-2079, but there was no coherent

northward shift between GCM's track density difference between periods. The one GCM (CCSM4) that demonstrated a defined increase in the expected surge intensity in 2054-2079, also demonstrated a relatively large track density increase (0.2 to 0.4 cyclones per 50,000 km^2) off the U.S. east coast. Further, a general intensification of Miller Type A events as posed by Marciano et al. (2015) and more frequent intense cyclones (<980 hPa) Colle et al. (2013) were only observed in the NORESM and CCSM4. However, over the mid-Atlantic coastline, the changes to the cyclone pressure distribution were weak (-3.0 to 3.0 hPa) and not consistent with the changes to the average along-track MSLP.

Thus, future work needs to investigate why some GCMs do not demonstrate an overall enhancement of U.S. east coast storms while others do. If the lack of intensification of Miller Type A events in some GCMs is primarily due to inaccurate model physics, then the CCSM4's more intense future surge scenario is plausible given its historical cyclone skill. However, to date, studies have only assessed the GCM's skill of all cyclones probabilistically. While these approaches are useful for climate studies, they combine all errors thus masking physical deficiencies of climate data on shorter time scales, which are more influential to regional scale processes. In order to gain confidence in GCM's realism and skill for predictions of surge, it would be helpful to simulate observed historical cyclone events and compare climate modeled and observed regional scale measurements.

Literature Cited

Akaike, H., 1974: A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723, doi:10.1109/TAC.1974.1400705.

American Littoral Society, 2012. Assessing the impacts of hurricane Sandy on coastal habitats. Available at: http://www.littoralsociety.org/images/PDFS/Policy/alssandyassessmentreport.pdf

Amin, M., 1978: A statistical analysis of storm surges in Torres Strait. *Mar. Freshwater Res.*, **29**, 479–496.

Atallah, E. H., and L. F. Bosart 2003: Extratropical transition and precipitation distribution of Hurricane Floyd '99. *Mon. Wea. Rev.*, **131**, 1063–1081.

ASCE Conference Proceedings, et al. Against the deluge: storm surge barriers to protect New York City. Reston, Virginia: American Society of Civil Engineers, 2013.

Bernier, N. B., and K. R. Thompson (2007), Tide-surge interaction off the east coast of Canada and northeastern United States, J. Geophys. Res., 112, C06008, doi:10.1029/2006JC003793.

Blake, E. S., T. B. Kimberlain, R. J. Berg, J. P. Cangialosi, and J. L. Beven III, 2013: Tropical cyclone report: Hurricane Sandy (AL182012), Technical Report, Natl. Hurricane Cent., Natl. Oceanic and Atmos. Admin., 157 pp. [Available at http://www.nhc.noaa.gov/data/tcr/AL182012_Sandy.pdf.]

Burnham, Kenneth P., and David R. Anderson, 2004: Multimodel inference understanding AIC and BIC in model selection: *Sociological Methods & Research*, **33**, 261–304, doi:10.1177/0049124104268644.

Burroughs, L. B., and W. A. Shaffer, 1997: East Coast extratropical storm surge and beach erosion guidance. NWS Tech. Pro- cedures Bull. 436, National Oceanic and Atmospheric Administration, 24 pp.

Bowman, M. J., B. A. Colle, R. Flood, D. Hill, R. E. Wilson, F. Buonaiuto, P. Cheng, and Y. Zheng, 2005: Hydrologic feasibility of storm surge barriers to protect the Metropolitan New York - New Jersey region. Final Rep., Marine Sciences Research Center Tech. Rep., Stony Brook University, 28 pp. [Available from School of Marine and Atmospheric Sciences, Stony Brook University, Endeavour Hall, Stony Brook, NY 11794-5000.]

Chang, E. K. M., Y. Guo, and X. Xia, 2012: CMIP5 multimodel ensemble projection of storm track change under global warming. *J. Geophys. Res.*, **117**, D23118, doi:10.1029/2012JD018578.

Chang, E. K. M., 2013: CMIP5 Projection of significant reduction in extratropical cyclone activity over North America. *J. Climate*, **26**, 9903–9922, doi:10.1175/JCLI-D-13-00209.1.

Colle, B. A., 2003: Numerical Simulations of the Extratropical Transition of Floyd (1999): Structural Evolution and Responsible Mechanisms for the Heavy Rainfall over the Northeast United States. *Mon. Wea. Rev.*, **131**, 2905–2926, doi:10.1175/1520-0493(2003)131<2905:NSOTET>2.0.CO;2.

_____,_____, F. Buonaiuto, M. J. Bowman, R.E. Wilson, R. Flood, R. Hunter, A. Mintz, and D. Hill, June 2008: Simulations of Past Cyclone Events to Explore New York City‟s Vulnerability to Coastal Flooding and Storm Surge Model Capabilities. Bull. Amer. Meteor. Soc., 89, 829-841.

_____, _____, K. Rojowsky, and F. Buonaito, 2010: New York City Storm Surges: Climatology and an Analysis of the Wind and Cyclone Evolution. *J. Appl. Meteor. Climatol.*, **49**, 85–100, doi:10.1175/2009JAMC2189.1.

Clancy, R., J. Kaitala, and L. Zambresky, 1986: The Fleet numerical oceanography center global spectral ocean wave model. Bull. Amer. Meteor. Soc., 67, 498– 512.

Colle, B. A., Z. Zhang, K. A. Lombardo, E. Chang, P. Liu, and M. Zhang, 2013: Historical evaluation and future prediction of eastern North America and western Atlantic extratropical cyclones in the CMIP5 models during the cool season. *Journal of Climate*, 130325101629005, doi:10.1175/JCLI-D-12-00498.1.

de Oliveira, M. M. F., N. F. F. Ebecken, J. L. F. de Oliveira, and I. de Azevedo Santos, 2009: Neural Network Model to Predict a Storm Surge. *Journal of Applied Meteorology and Climatology*, **48**, 143–155, doi:10.1175/2008JAMC1907.1.

Devis, A., N. P. M. van Lipzig, and M. Demuzere, 2014: A height dependent evaluation of wind and temperature over Europe in the CMIP5 Earth System Models. *Clim Res*, **61**, 41–56, doi:10.3354/cr01242.

Di Liberto, T., B. A. Colle, N. Georgas, A. F. Blumberg, and A. A. Taylor, 2011: Verification of a Multimodel Storm Surge Ensemble around New York City and Long Island for the Cool Season. *Weather and Forecasting*, **26**, 922–939, doi:10.1175/WAF-D-10-05055.1.

Drag, Walter (2000-07-14). "A comparative retrospective on the Perfect Storm". Boston National weather Service Office. Retrieved 2011-07-01.

Forbes, C., J. Rhome, C. Mattocks, and A. Taylor, 2014: Predicting the storm surge threat of Hurricane Sandy with the National Weather Service SLOSH model. *Journal of Marine Science and Engineering*, **2**, 437–476, doi:10.3390/jmse2020437.

Garratt, J. R., 1977: Review of drag coefficients over oceans and continents. *Monthly Weather Review*, **105**, 915–929, doi:10.1175/1520-0493(1977)105<0915:RODCOO>2.0.CO;2.

Georgas, Nickitas, 2010: Establishing confidence in marine forecast systems: The design of a high fidelity marine forecast model for the NY/NJ Harbor estuary and its adjoining coastal waters. PhD thesis, Stevens Institute of Technology, Hoboken, NJ, USA. http://web.stevens.edu/ses/documents/fileadmin/documents/pdf/PhD-dissertation_signed-o.pdf

_____, _____ P. Orton, A. Blumberg, L. Cohen, D. Zarrilli, and L. Yin, 2014: The impact of tidal phase on Hurricane Sandy's flooding around New York City and Long Island Sound. *J. of Extr. Even.*, **01**, 1450006, doi:10.1142/S2345737614500067.

Gjevik, B., and L. P. Røed, 1976: Storm surges along the Western coast of Norway. *Tellus*, **28**, 166–182, doi:10.1111/j.2153-3490.1976.tb00664.x.

Hall, T. M., and A. H. Sobel, 2013: On the impact angle of Hurricane Sandy's New Jersey landfall: HURRICANE SANDY IMPACT ANGLE. *Geophysical Research Letters*, **40**, 2312–2315, doi:10.1002/grl.50395.

Harris, D. Lee, and Aldo Angelo, 1963: A regression model for storm surge prediction. *Mon. Weather Rev.,* **91**, 710–26, doi:10.1175/1520 0493(1963)091<0710:ARMFSS>2.3.CO;2.

Hirsch, M. E., A. T. DeGaetano, and S. J. Colucci, 2001: An East Coast winter storm climatology. J. Climate, 14, 882–899.

Howard, K., M. Splitt, S. Lazarus, G. Zarillo, S. Chiao, P. Santos, and D. Sharp, 2009: The impact of atmospheric model resolution on a coupled wind/wave forecast system. Preprints, 16th Conference on Air-Sea Interaction, Phoenix, Arizona,

Hsu, S. A., E. A. Meindl, and D. B. Gilhousen, 1994: Determining the Power-Law Wind-Profile Exponent under Near-Neutral Stability Conditions at Sea. *Journal of Applied Meteorology*, **33**, 757–765, doi:10.1175/1520-0450(1994)033<0757:DTPLWP>2.0.CO;2.

Intergovernmental Panel on Climate Change, 2013: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp.

Klein, William, Lewis Billy M., and Enger Isadore, 1959: Objective Prediction of five- day mean temperatures during winter. *J. Meteor.*, **16**, 672–682. doi: http://dx.doi.org/10.1175/1520-0469(1959)016<0672:OPOFDM>2.0.CO;2

Komar, P. D., 1998: Beach Processes and Sedimentation. 2 nd ed. Prentice Hall. 544 pp

Lambert, S. J., 1995: The effect of enhanced greenhouse warming on winter cyclone frequencies and strengths. J. Climate, 8, 1447–1452.

Large, W. G., and S. Pond, 1981: Sensible and Latent Heat Flux Measurements over the Ocean. J. Phys. Oceanogr., 12, 464-482.

Mel, Ricardo and Lionello, Piero, 2014: Storm Surge Ensemble Prediction for the City of Venice. *Wea. Forecasting*, **29**, 1044–1057. doi: http://dx.doi.org/10.1175/WAF-D-13-00117.1

Mesinger, F., and Coauthors, 2006: North American regional reanalysis. *Bulletin of the American Meteorological Society*, **87**, 343–360, doi:10.1175/BAMS-87-3-343.

Miller, J. E., 1946: Cyclogenesis in the atlantic coastal region of the united states. *J. Meteor.*, **3**, 31–44, doi:10.1175/1520-0469(1946)003<0031:CITACR>2.0.CO;2.

Marciano, C. G., G. M. Lackmann, and W. A. Robinson, 2014: Changes in U.S. East Coast Cyclone Dynamics with Climate Change. *J. Climate*, **28**, 468–484, doi:10.1175/JCLI-D-14-00418.1.

Mizuta, R., M. Matsueda, H. Endo, and S. Yukimoto, 2011: Future change in extratropical cyclones associated with change in the upper troposphere. *J. Climate*, **24**, 6456–6470, doi:10.1175/2011JCLI3969.1.

Müller-Navarra, Sylvin H., and Knüpffer Klaus, 2010: Improvement of water level forecasts for tidal harbors by means of model output statistics (MOS) part 1 (skew surge forecast), 22 pp.

National Weather Service, Eastern Region, 1994: The great nor-easter of December 1992. Disaster Survey Rep., 59 pp. [Available from the National Weather Service, Eastern Region Headquarters, 630 Johnson Ave, Bohemia, NY 11716.]

Orton, P., N. Georgas, A. Blumberg, and J. Pullen, 2012: Detailed modeling of recent severe storm tides in estuaries of the New York City region. *J. Geophys. Res.*, **117**, C09030, doi:10.1029/2012JC008220.

Perkins, S. E., A. J. Pitman, N. J. Holbrook, and J. McAneney, 2007: Evaluation of the AR4 Climate Models' Simulated Daily Maximum Temperature, Minimum Temperature, and Precipitation over Australia Using Probability Density Functions. *J. Climate*, **20**, 4356–4376, doi:10.1175/JCLI4253.1.

Pore, N. A., 1964: The relation of wind and pressure to extratropical storm surges at Atlantic City. *Journal of Applied Meteorology*, **3**, 155–163, doi:10.1175/1520-0450(1964)003<0155:TROWAP>2.0.CO;2.

____, _____, 1965: Chesapeake Bay extratropical storm surges. *Chesapeake Science*, **6**, 172–182, doi:10.2307/1350850.

____, ____, 1965: Chesapeake Bay extratropical storm surges. *Chesapeake Science*, **6**, 172–182, doi:10.2307/1350850

____, ____, W. S. Richardson, and H. P. Perrotti, 1974: Forecasting extratropical storm surges for the northeast coast of the United States. Tech. Memo. 50, National Weather Service, NOAA, U.S. Department of Commerce, 70 pp.

Pryor SC, Schoof JT, Barthelmie RJ (2005a) Empirical down- scaling of wind speed probability distributions. J Geo- phys Res 110:D19109, doi:10.1029/2005JD005899

Resio, D.T., and J.J. Westerink, 2008, "Modeling the physics of storm surges", (PDF File) Physics Today, September 2008, pp. 33-38.

Roberts, Keith J.; Colle, Brian A.; Georgas, Nickitas; Munch, Stephen: "A Regression- based Approach for Cool-Season Storm Surge Predictions along the New York/New Jersey Coast"

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast system reanalysis. *Bulletin of the American Meteorological Society*, **91**, 1015–1057, doi:10.1175/2010BAMS3001.1.

Salmun, H., A. Molod, K. Wisniewska, and F. S. Buonaiuto, 2011: Statistical prediction of the storm surge associated with cool-weather storms at the Battery, New York. *J. Appl. Meteor. Climatol.*, **50**, 273–282, doi:10.1175/2010JAMC2459.1.

_____, ____. _____, ____2015: The use of a statistical model of storm surge as a bias correction for dynamical surge models and its applicability along the U.S. East Coast. *Journal of Marine Science and Engineering*, **3**, 73–86, doi:10.3390/jmse3010073.

Shen, J. Gong, W., Wang, H., 2005: Simulation of Hurricane Isabel using the Advanced Circulation Model (ADCIRC). Hurricane Isabel in perspective. Proceedings of a conference. Chesapeake Research Consortium. K. G. Sellner (ed.), 107-116.

Talke, S. A., P. Orton, and D. A. Jay, 2014: Increasing storm tides in New York Harbor, 1844–2013. *Geophys. Res. Lett.*, **41**, 2014GL059574, doi:10.1002/2014GL059574.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Scileppi, E., and J. P. Donnelly, 2007: Sedimentary evidence of hurricane strikes in western Long Island, New York. *Geochem. Geophys. Geosyst.*, **8**, Q06011, doi:10.1029/2006GC001463.

Shen, J. Gong, W., Wang, H., 2005: Simulation of Hurricane Isabel using the Advanced Circulation Model (ADCIRC). Hurricane Isabel in perspective. Proceedings of a conference. Chesapeake Research Consortium. K. G. Sellner (ed.), 107-116.

Siek M, D. P. S., 2010: Nonlinear chaotic model for predicting storm surges. *Nonlinear Processes in Geophysics*, doi:10.5194/npg-17-405-2010.

Weaver, R. J., 2004: Effect of wave forces on storm surge. M.S. thesis, Dept. of Civil and Coastal Engineering, University of Florida, 75 pp. http://www.essie.ufl.edu/~slinn/weaver.slinn.surge.pdf

Weisberg, R. H., and L. Zheng, 2008: Hurricane storm surge simulations comparing three-dimensional with two-dimensional formulations based on an Ivan-like storm over the Tampa Bay, Florida region. *J. Geophys. Res.*, **113**, C12001, doi:10.1029/2008JC005115.

Westerink, J. J., R. A. Luettich, and N. Scheffner, 1993: ADCIRC: An Advanced Three-Dimensional Circulation Model for Shelves, Coasts, and Estuaries. Report 3. Development of a Tidal Constituent DataBase for the Western North Atlantic and Gulf of Mexico.

Wilks, D. S., 2006: Statistical methods in the Atmospheric Sciences. 2nd ed. Academic Press, 627 pp.

Xu, Zhongfeng and Yang, Zong-Liang, 2012: An Improved dynamical downscaling method with GCM bias corrections and its validation with 30 years of climate imulations. *J. Climate*, **25**, 6271–6286. doi: http://dx.doi.org/10.1175/JCLI-D-12-00005.1

Zhang, K., B. C. Douglas, and S. P. Leatherman, 2000: Twentieth-Century Storm Activity along the U.S. East Coast. *J. Climate*, **13**, 1748–1761, doi:10.1175/1520-0442(2000)013<1748:TCSAAT>2.0.CO;2.

Zhang, Minghua, Bokuniewicz, Henry, Lin, Wuyin, Jang, Sung-Gheel, Liu Ping, 2014: NYRISE Climate Risk Report for Suffolk and Nassau TR-0-14-01. Available at: http://nysrise.org/docs/NYSRISE-SBU-ClimateRiskReportforNassauandSuffolk-August2014.

Appendix

Tables

Table 2.1: The year in which training and verification cool seasons begin and the percent of water level data from NOAA Tides and Currents that was available during those cool seasons. This cross-validation approach is referred to as CV1 and when the testing and training periods are reversed, it is referred to as CV2. The * denotes a half season from 0000 UTC Jan 1, 1979 to 0000 UTC March 31, 1980. The ** denotes half season from 0000 UTC October 12, 2012 to 0000 UTC December 31, 2012. Otherwise, all periods begin on 0000 UTC October 1 and terminate at 0000 UTC March 31.

Training Years (17):	**Testing Years (17):**
1979,1981,1983,1985, 1987,1989,1991,1993, 1995,1997,1999,2001 2003,2005,2007,2009,2011	1979*,1980,1982,1984 1986,1988,1990,1992, 1994,1996,1998,2000, 2002,2004,2006,2008, 2010,2012**

THE BATTERY		**ATLANTIC CITY**		**MONTAUK**	
Trend Removed	**3.0 mm/yr**	**4.14 mm/yr**		**5.1 mm/yr**	
Training	**Testing**	**Training**	**Testing**	**Training**	**Testing**
95.0 %	92.0 %	99.0 %	99.0 %	90.0 %	81.0 %

Table 2.2: The latitude and longitude boundaries, the horizontal and vertical dimension (km) of the different boxes. The verification R^2 values with total uncertainty in square brackets for a MLR model using the box specified to calculate predictors is also shown.

Station	**Bounds**		
	Box 1 (338 x 111 km)	**Box 2 (442 x 222 km)**	**Box 3 (507 x333 km)**
THE BATTERY (B)	39.5 N: 40.5 N -74.0 W: -70.0 W	38.5 N: 40.5 N -74.0 W: -69.0 W	37.5 N : 40.5 N -74.0 W: -68.0 W
R^2	**0.66 [0.64-0.67]**	**0.62 [0.60-0.64]**	**0.57 [0.56-0.59]**
ATLANTIC CITY (A)	38.3 N: - 39.3 N -74.42 W: -70.42 W	37.3 N:-39.3 N -74.42 W: -69.42 W	36.3 N: 39.3 N -74.42 W: -68.42 W
R^2	**0.57 [0.53-0.60]**	**0.52 [0.48-0.55]**	**0.52 [0.49-0.55]**
MONTAUK POINT (C)	40.0 N: 41.0 N -72.0 W: -68.0 W	39.0 N: 41.0 N -72.0 W: -67.0 W	38.0 N : 41.0 N -72.0 W: -66.0 W
R^2	**0.71 [0.69-0.73]**	**0.68 [0.66-0.69]**	**0.63 [0.61-0.65]**

Table 2.3: Storm surge heights (m) that correspond with 1, 2, and 3 standard deviations departure from the average > 0.61 m 3-h surge height for raw predictions, bias-corrected predictions (BC), and observed three-hourly storm surge data at The Battery between 1979-2012.

Model Name	$\mu_\eta + \sigma_\eta$		$\mu_\eta + 2\sigma_\eta$		$\mu_\eta + 3\sigma_\eta$	
	Raw	**BC**	**Raw**	**BC**	**Raw**	**BC**
CCSM4	1.09	1.02	1.29	1.21	1.50	1.40
BCC	0.98	1.09	1.21	1.34	1.44	1.58
NORESM	0.88	1.07	0.98	1.32	1.08	1.58
CANESM	0.90	1.03	1.02	1.20	1.13	1.38
GFDL	0.83	0.99	0.94	1.15	1.05	1.32
CNRM	1.11	1.02	1.35	1.20	1.59	1.39
Observed	**1.01**	----	**1.23**	----	**1.45**	----

Table 2.4: Attributes and abbreviations of the six GCMs from the highest resolution to the lowest resolution.

Model Name	**Horizontal Resolution**	**LML Height (m)**	**Reference**
Community Climate System Model, version 4 (CCSM4)	1.25°x 0.94°	63.0 [58.0-67.4]	Gent et al. (2011)
Centre National de Recherches Meteorologiqes Coupled Global Climate Model, version 5.1 (CNRM)	1.4°x 1.4°	32.0 [30.0-34.3]	Michou et al. (2011)
Norwegian Earth System Model, version 1 (intermediate resolution) (NORESM)	2.5°x 1.9°	60.0 [58.3-60.4]	Zhang et al. (2012)
Geophysical Fluid Dynamics Laboratory Earth System Model with Modular Ocean Model 4 (MOM4) ocean component (ESM2G) (GFDL)	2.5°x 2.0°	35.0 [33.0-37.0]	Donner et al. (2011)
Canadian Earth System Model Version 2 (CANESM)	2.81°x 2.79°	42.3 [40.3-44.3]	McFarlane et al. (1992)
Beijing Climate Center, Climate System Model, version 1.1 (BCC)	2.8°x 2.8°	60.0 [57.9-60.3]	Wu et al. (2010).

Table 3.1: MLR coefficients for all configurations of reanalysis data (CFSR and NARR) and sampling data (CV1 & CV2) that minimize ΔAIC_i. All coefficients are significant ($\propto = 0.05$) and correspond with z-score predictors based on their respective training data. Standard error on all coefficients was two orders of magnitude smaller ($1x10^{-5}$) than all the coefficient's magnitudes due to a large (n~24,000) sample size and is not reported.

		NARR				**CFSR**			
STATION	**CV**	**y-int.**	$\mathbf{A\tau_x}$	$\mathbf{A\tau_y}$	**P**	**y-int.**	$\mathbf{A\tau_x}$	$\mathbf{A\tau_y}$	**P**
Atlantic City	**CV1**	0.03	-0.12	-0.04	-0.10	0	-0.13	-0.05	-0.05
	CV2	0.02	-0.11	-0.04	-0.08	0.01	-0.12	-0.04	-0.03
The Battery	**CV1**	0.04	-0.14	-0.02	-0.08	0.03	-0.13	-0.02	-0.07
	CV2	0.04	-0.14	-0.02	-0.07	0.03	-0.13	-0.02	-0.06
Montauk Point	**CV1**	-0.04	-0.10	-0.02	-0.09	-0.04	-0.09	-0.02	-0.09
	CV2	-0.03	-0.10	-0.02	-0.09	-0.04	-0.09	-0.01	-0.09

Table 3.2: Predictor summing lengths in hours for the zonal direction (Lx), the meridional direction (Ly), R^2calculated using verification predictions made with MLR, and R^2 for predictions made with BC-MLR in square brackets for different cross-validations.

		NARR			CFSR		
STATION	**CV**	**Lx**	**Ly**	**R^2**	**Lx**	**Ly**	**R^2**
Atlantic City	**CV1**	15	27	0.53 [0.44]	12	0	0.56 [0.49]
	CV2	12	27	0.60 [0.57]	12	0	0.63 [0.61]
The Battery	**CV1**	12	27	0.64 [0.60]	15	18	0.63 [0.59]
	CV2	18	27	0.67 [0.63]	18	27	0.66 [0.61]
Montauk Point	**CV1**	9	27	0.69 [0.66]	12	27	0.61 [0.56]
	CV2	9	27	0.73 [0.71]	12	27	0.65 [0.61]

Table 3.3: The scatter index calculated for MLR and BC-MLR surge event predictions that meet the all criteria in verification period data for the various cross-validation configurations.

		NARR		**CFSR**	
STATION	**CV**	**Scatter Index (%)**		**Scatter Index (%)**	
		MLR	**BC-MLR**	**MLR**	**BC-MLR**
Atlantic City	**CV1**	14.10	12.16	15.13	14.07
	CV2	17.31	14.89	15.9	13.96
The Battery	**CV1**	10.23	9.35	11.14	10.10
	CV2	11.42	10.65	12.15	11.40
Montauk Point	**CV1**	7.36	6.81	9.41	8.20
	CV2	6.76	5.75	12.82	11.49

Table 4.1: The mean and range of observed (natural variability) surge counts from 1948-2012 and for the modeled surge climate during the historical period (1979-2004 Nov.-March).

Metric	Natural Variability	Modeled
Count/ Historical	**93**	**63 [45-90]**
Count/Season	**3.98 [0-10.0]**	**2.47 [0-7.33]**

Table 4.2: The average number of classified cyclone tracks per season (Nov.-March) during the historical period (1979-2004) for cyclones tracked with the six GCMs and the CFSR within the study region (100 W to 50 W and 20 N to 50 N). Note that Miller Type B includes tracks associated with Great Lakes and Mid-Atlantic storms.

Class	**CFSR**	**CCSM4**	**BCC**	**GFDL**	**CANESM**	**NORESM**	**CNRM**
Miller Type A	15.52	11.12	14.5	10.96	9.36	13.36	13.2
Miller Type B	23.56	21.76	13.7	23.12	20.6	23.56	22.2
Other	38.08	34.88	48.0	31.28	30.6	37.44	34.96
Climatology	77.16	67.76	76.2	65.36	60.56	74.36	70.36

Table 4.3: The number of classified cyclone tracks that generated impactful surge events (storm maximum surge ≥0.61 m) during the historical period. Observed surge data was used for CFSR and modeled surge was used for the GCMs. Note that Miller Type B include tracks associated with both Great Lakes and Mid-Atlantic cyclones.

Class	**CFSR**	**CCSM4**	**BCC**	**GFDL**	**CANESM**	**NORESM**	**CNRM**
Miller Type A	40	43	33	52	34	20	19
Miller Type B	55	49	27	76	52	42	51
Other	53	43	46	43	35	62	70
Climatology	155	136	107	171	121	124	140

Table 4.4: Classified track density (TR) correlations and the mean absolute difference in track density difference per grid point (TD) per season (Nov.-March) with CFSR cyclones during the historical period for the region between 100 W to 50 W and 20 N to 50 N. Note that Miller Type B includes tracks associated with Great Lakes and Mid-Atlantic storms.

Class	**CCSM4**		**BCC**		**GFDL**		**CANESM**		**NORESM**		**CNRM**	
	TR	**TD**	**TR**	**TD**	**TR**	**TD**	**TR**	**TD**	**TR**	**TD**	**TR**	**TD**
Miller Type A	**0.87**	*0.16*	**0.87**	*0.15*	**0.85**	*0.15*	**0.86**	*0.18*	**0.85**	*0.17*	**0.87**	*0.14*
Miller Type B	**0.88**	*0.16*	**0.83**	*0.25*	**0.93**	*0.14*	**0.91**	*0.15*	**0.90**	*0.15*	**0.91**	*0.15*
Other	**0.91**	*0.20*	**0.89**	*0.29*	**0.87**	*0.25*	**0.89**	*0.27*	**0.91**	*0.21*	**0.89**	*0.21*
Climatology	**0.90**	*0.40*	**0.89**	*0.41*	**0.91**	*0.40*	**0.90**	*0.50*	**0.91**	*0.41*	**0.92**	*0.36*

Figures

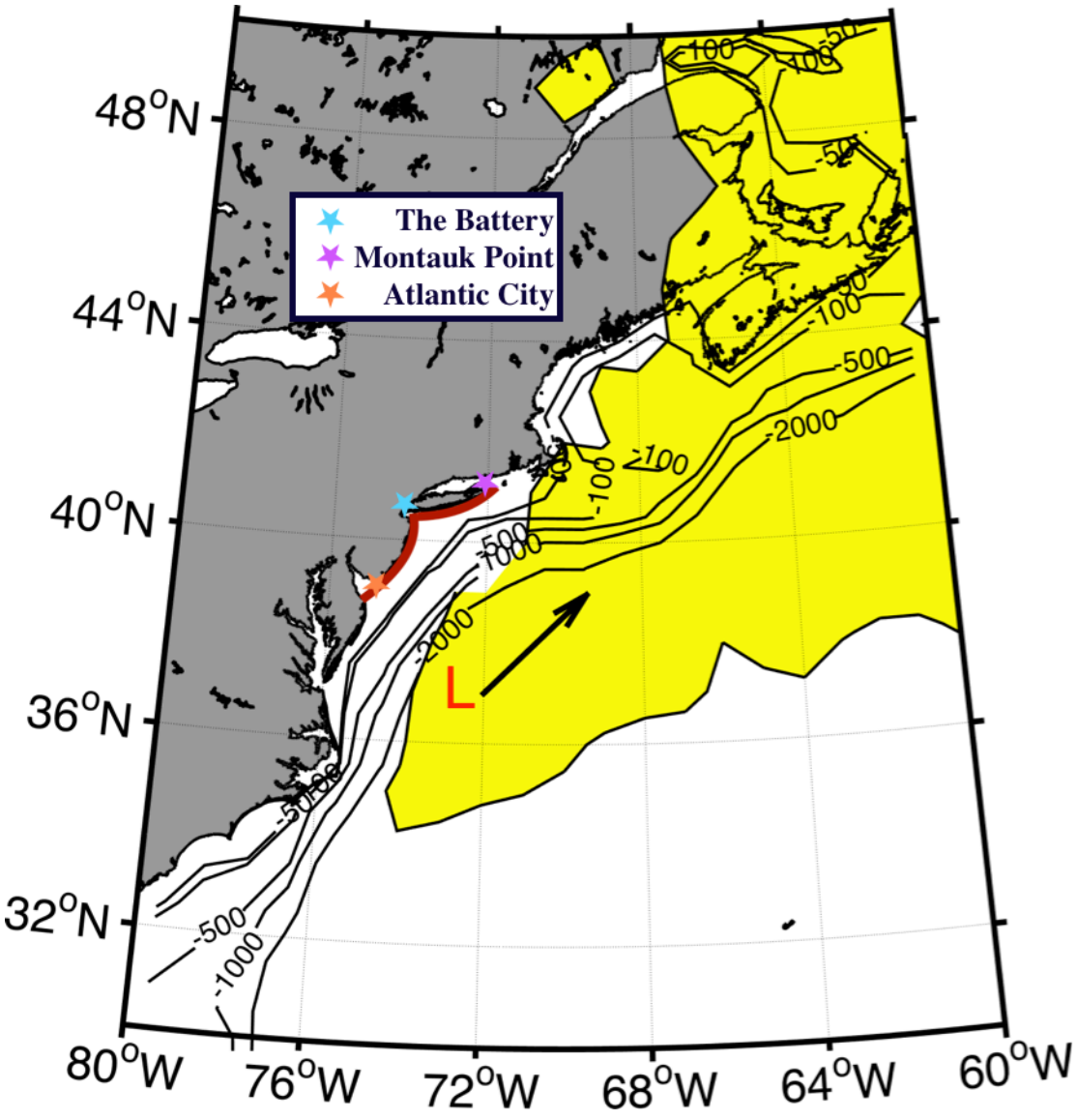


Figure 1.1: Depicts the study region along with bathymetric contours labeled in meters below mean sea level. The New York/New Jersey coastline/Bight is highlighted in dark magenta along with the three stations studied. The yellow contour depicts the north Atlantic cyclone Nov.-March track.

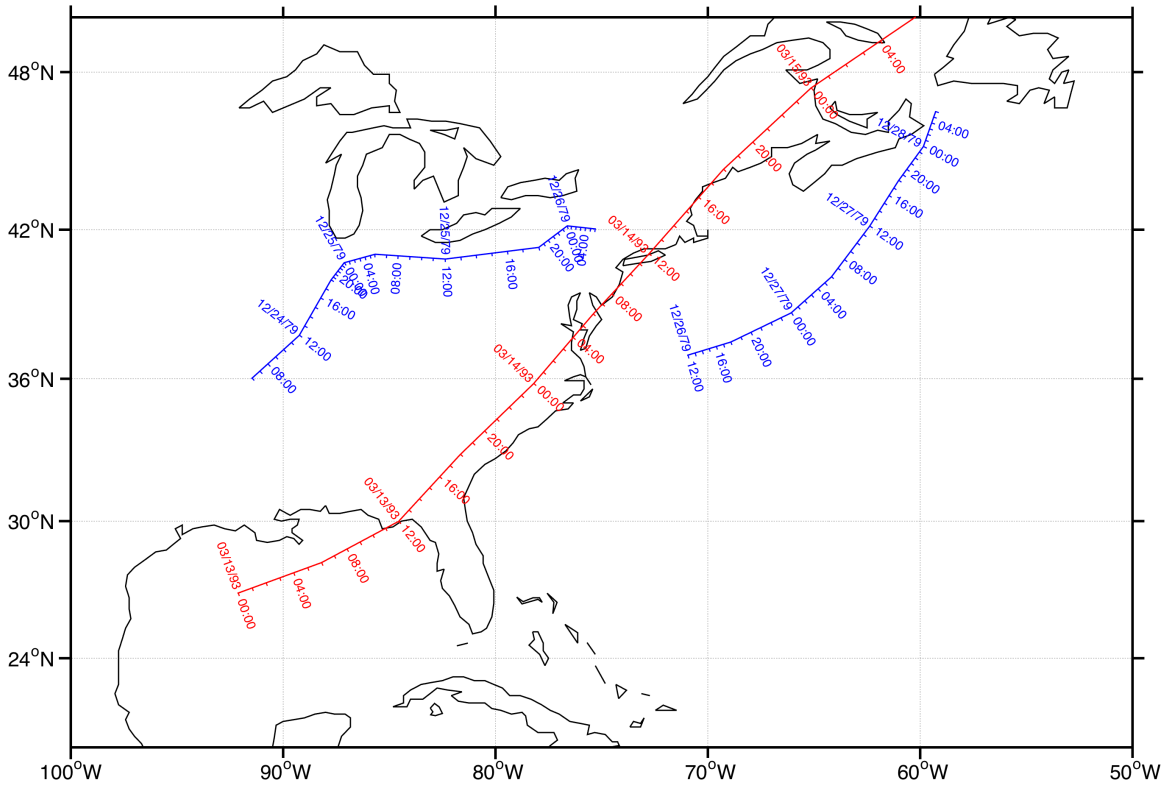


Figure 1.2: An example of a Miller Type A event in red occurring between the dates of 13 March 1993 and 15 March 1993, and a Miller Type B event in blue occurring between the dates of 23 December 1979 and 28 December 1979. Ticks are plotted every hour, dates are plotted every 12 h, and times are plotted every 4 hours.

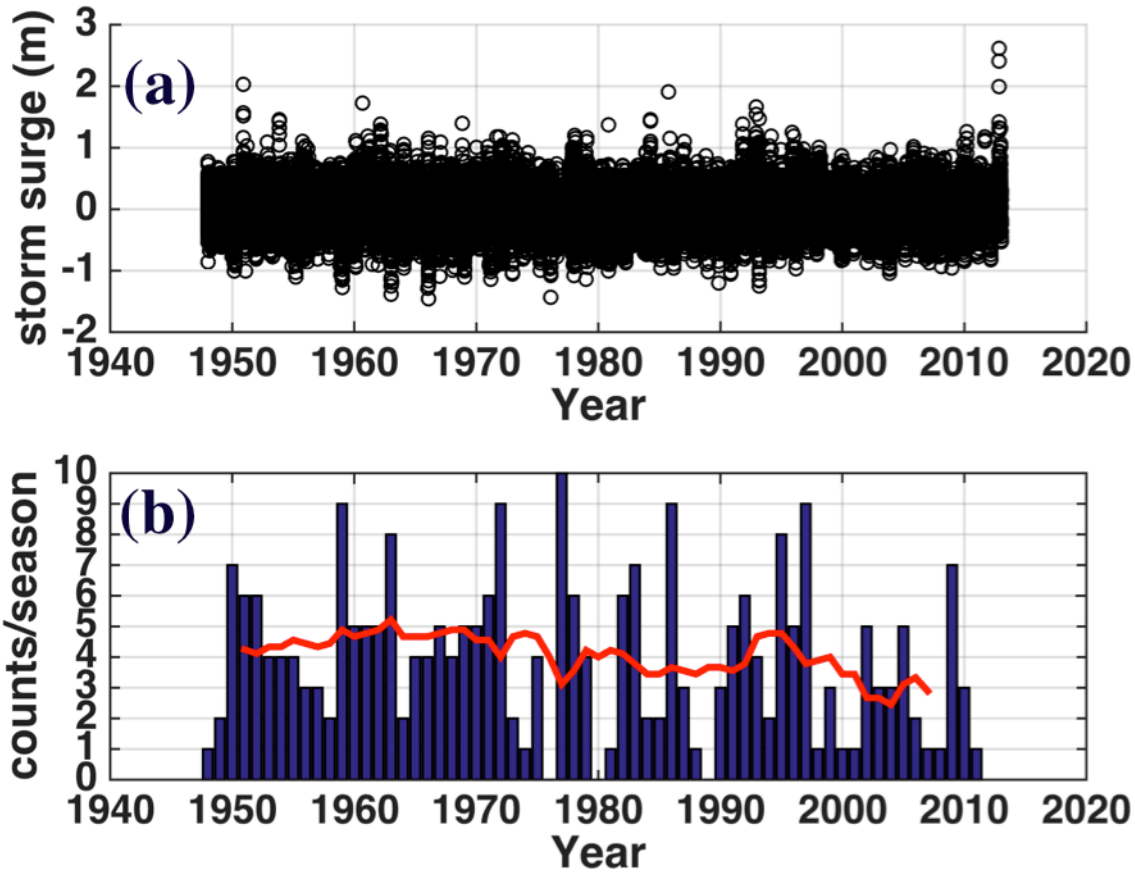


Figure 1.3: (a) Observed three-hourly detrended storm surge at The Battery from 1948-2012 and (b) the number of impactful surge events ($\geq 0.61\ m$) or counts during each season (Nov.-March) from 1948-2012. The 10-year moving average of counts per season is drawn as a red line in (b) where each point corresponds to the preceding and following 5-y of data.

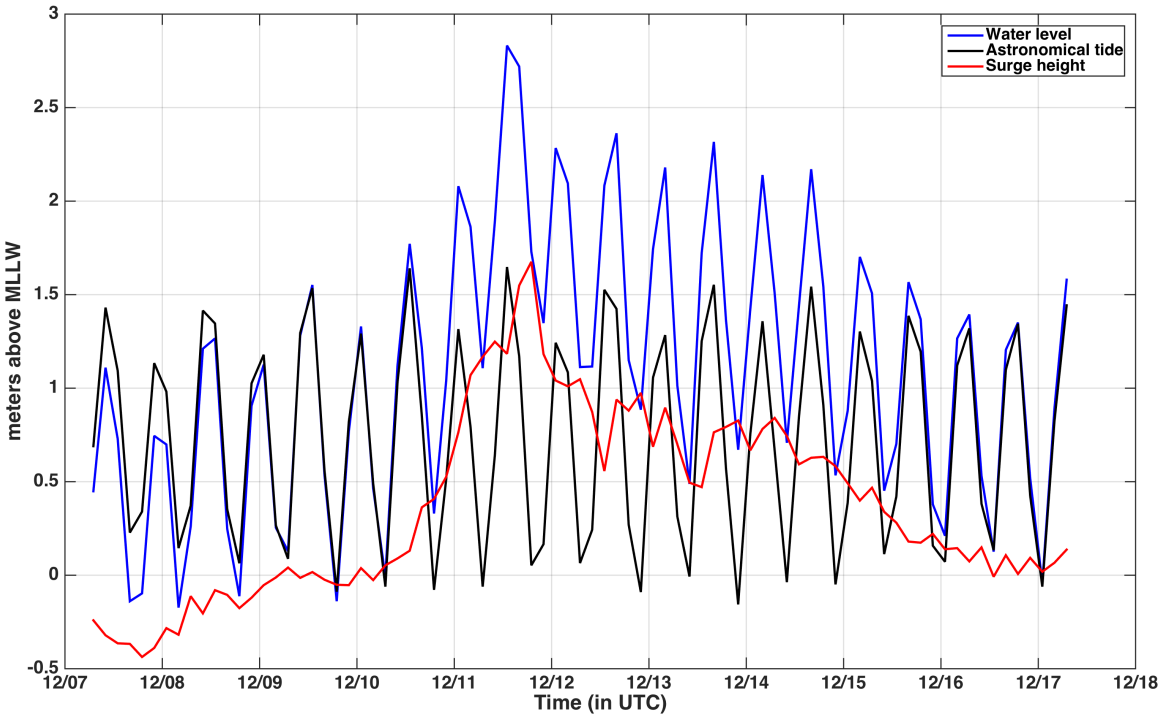


Figure 1.4: Time series of storm surge (m), total water level (m above MLLW), and astronomical tide (m above MLLW) from the 7 December 1992 to 18 December 1992.

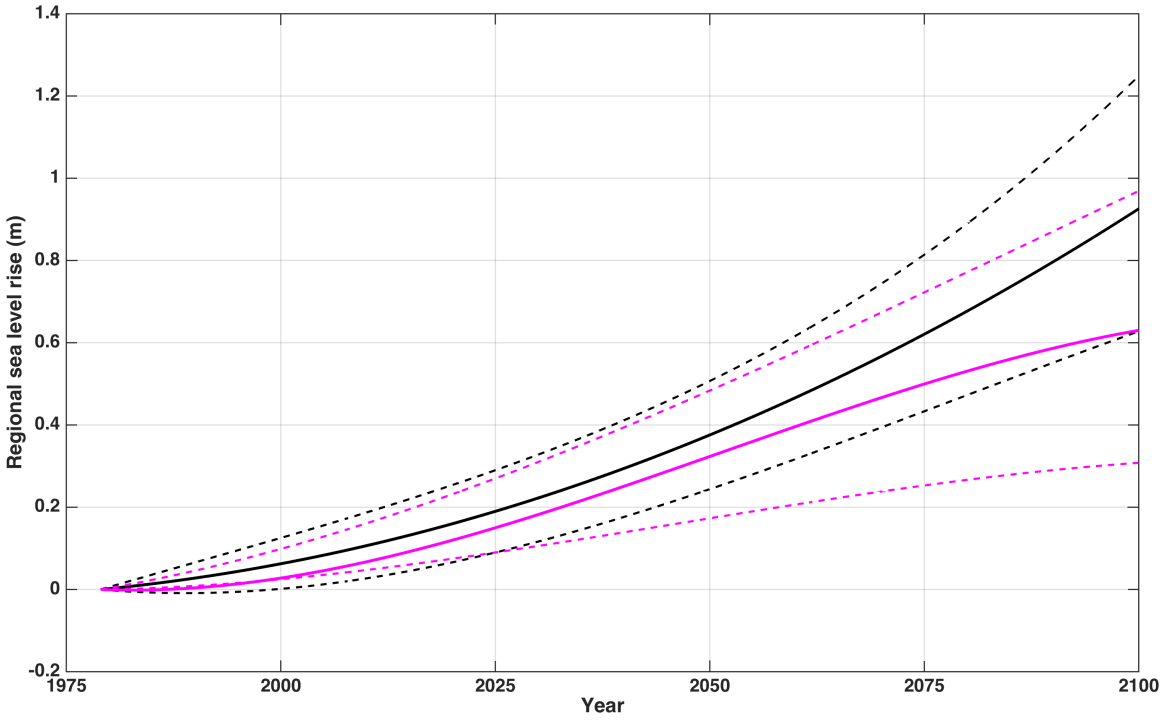


Figure 1.5: Sea level rise scenarios from 1979 to 2100 above a reference level defined over 1984-2004 for New York coastal waters. These scenarios are consistent with the global warming associated with the RCP4.5 (magenta lines) and RCP8.5 (black lines) IPCC (2013) emissions scenario (Zhang et al. 2014). Dotted lines represent 95th uncertainty ranges due to physical uncertainties in the Earth-ocean coupled system.

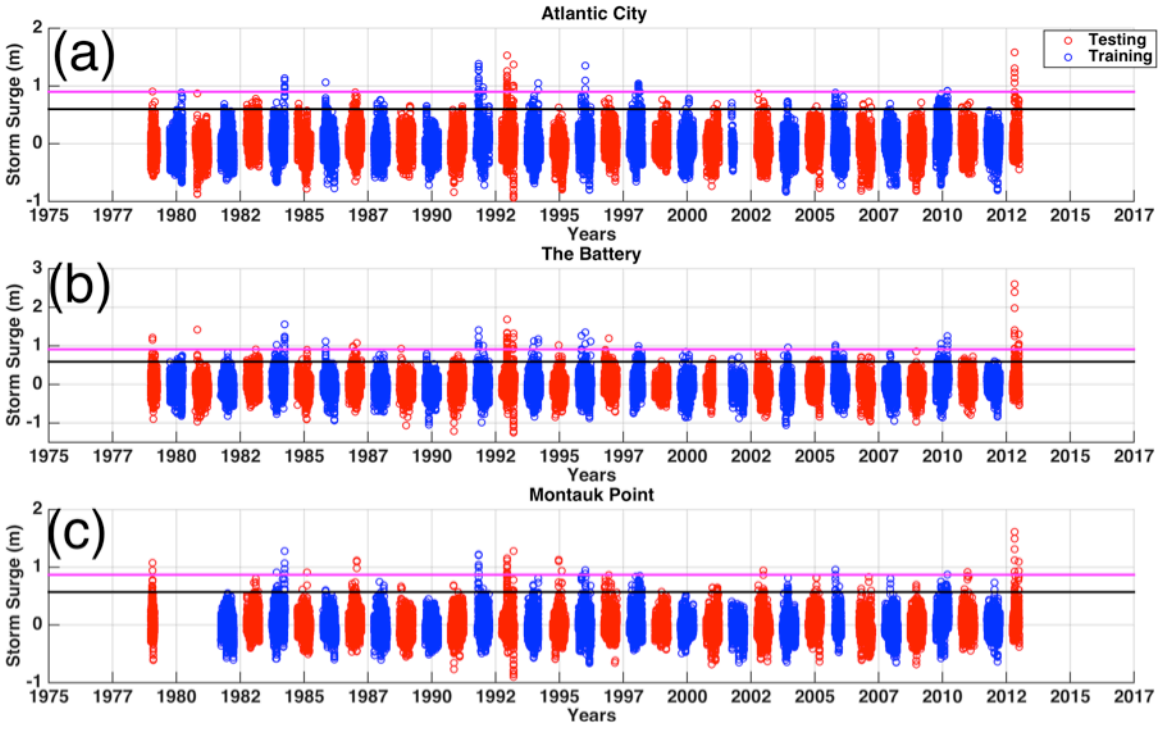


Figure 2.1: Time series of (a) Atlantic City, (b) The Battery, and (c) Montauk Point detrended storm surge with the training and testing cool seasons highlighted in blue and red corresponding with the cross-validation 1 configuration (CV1). The minor and moderate storm maximum surge thresholds are drawn as black and magenta lines for each station respectively.

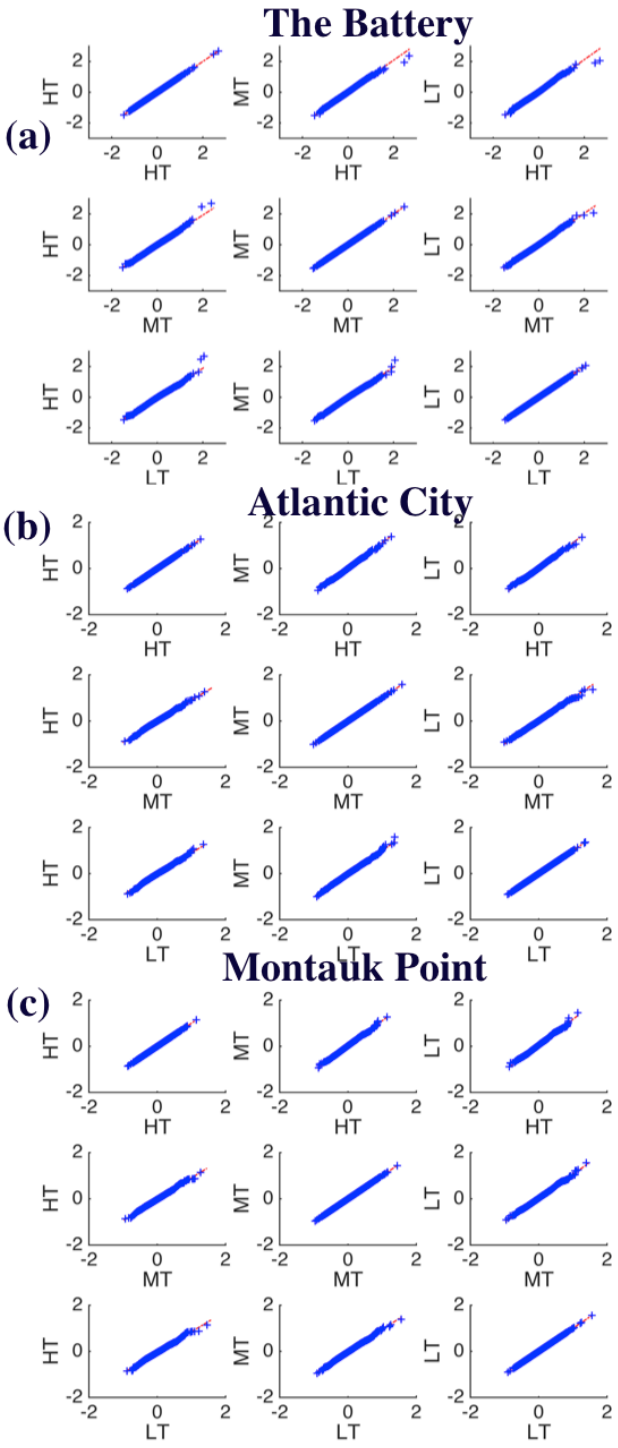


Figure 2.2: Quantile-Quantile (qq) plots of the tidal residual (surge-tid) during high tide (HT), low tide (LT), and mean tide (MT) for 1948-2012 at (a) The Battery, (b) Atlantic City, and (c) Montauk Point.

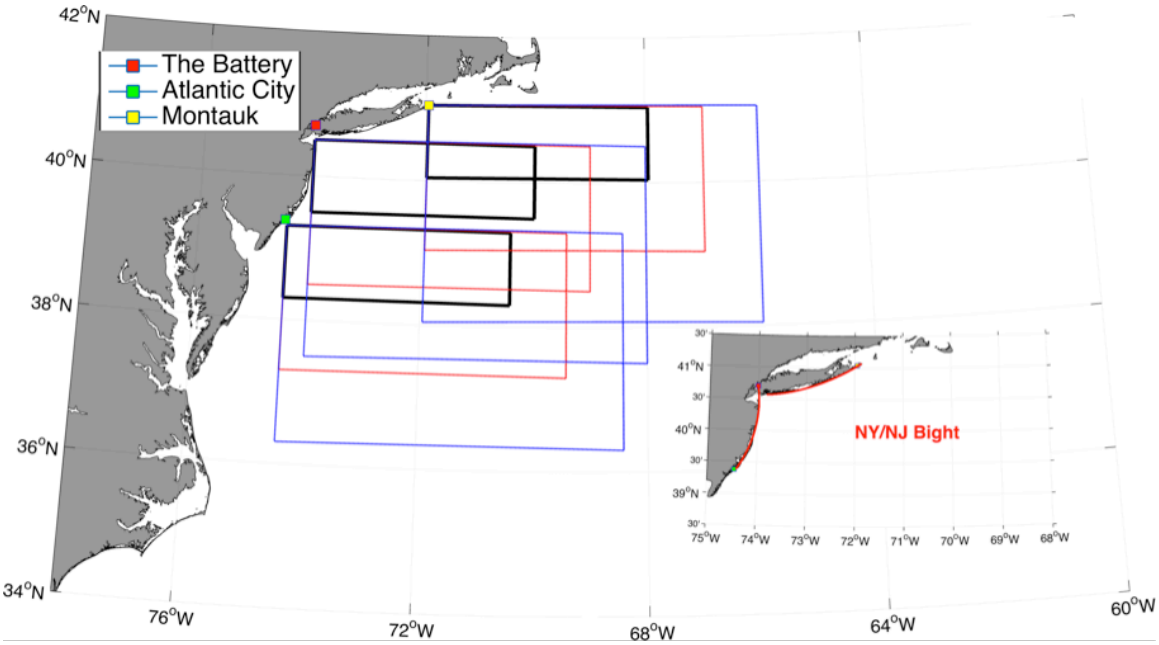


Figure 2.3: The boxes used for the collection of predictors for each station. The coastline between Montauk Point, New York and Atlantic City New York (otherwise known as the NY/NJ Bight) is highlighted in red in the inset.

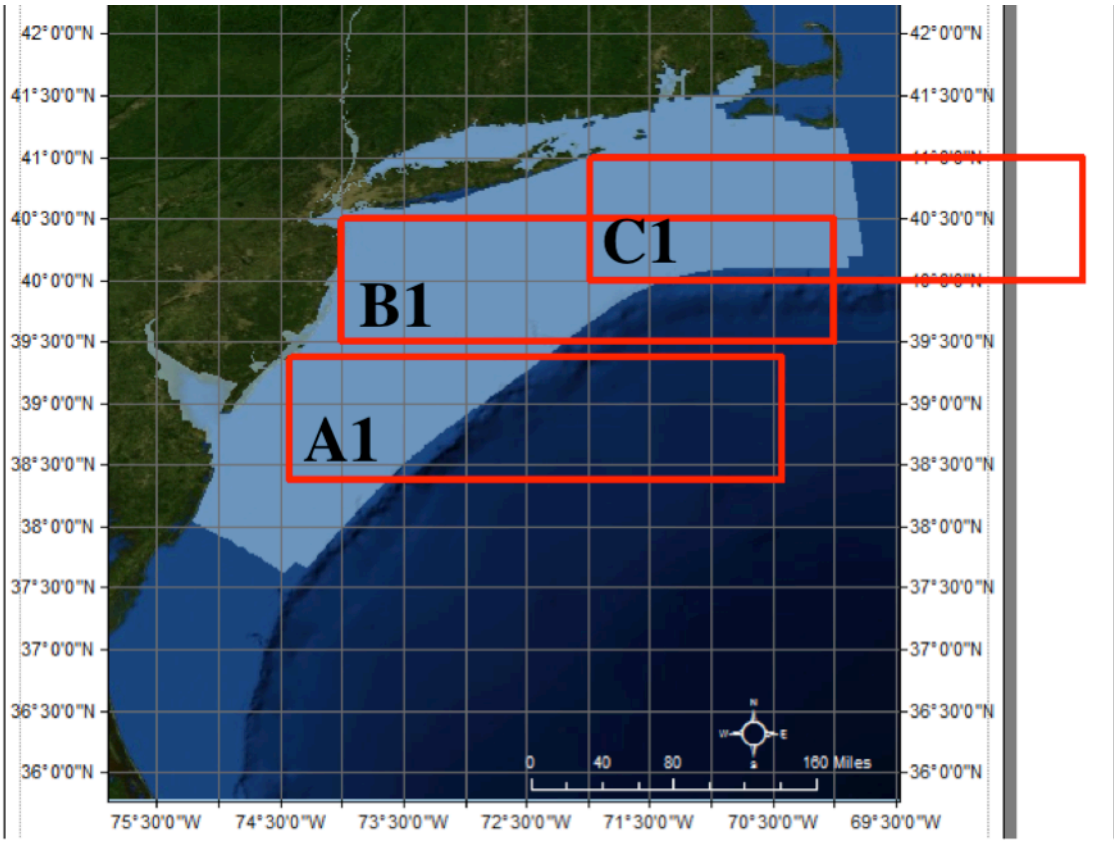


Figure 2.4: SIT-NYHOPS grid in light blue denotes the region of 1800 UTC NAM data available within box A1/B1/C1.

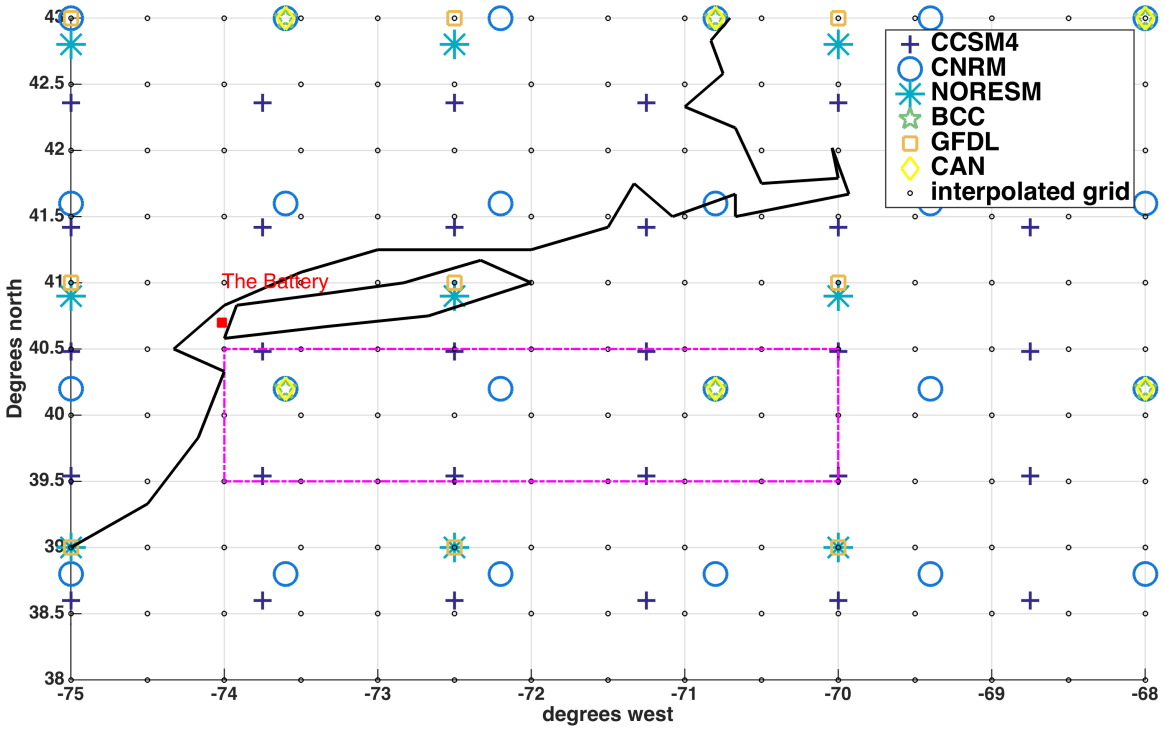


Figure 2.5: Grid points for each GCM overlaid with the 0.5° x 0.5° grid that GCM data was interpolated to. The predictor box is illustrated with a dash-dotted magenta line.

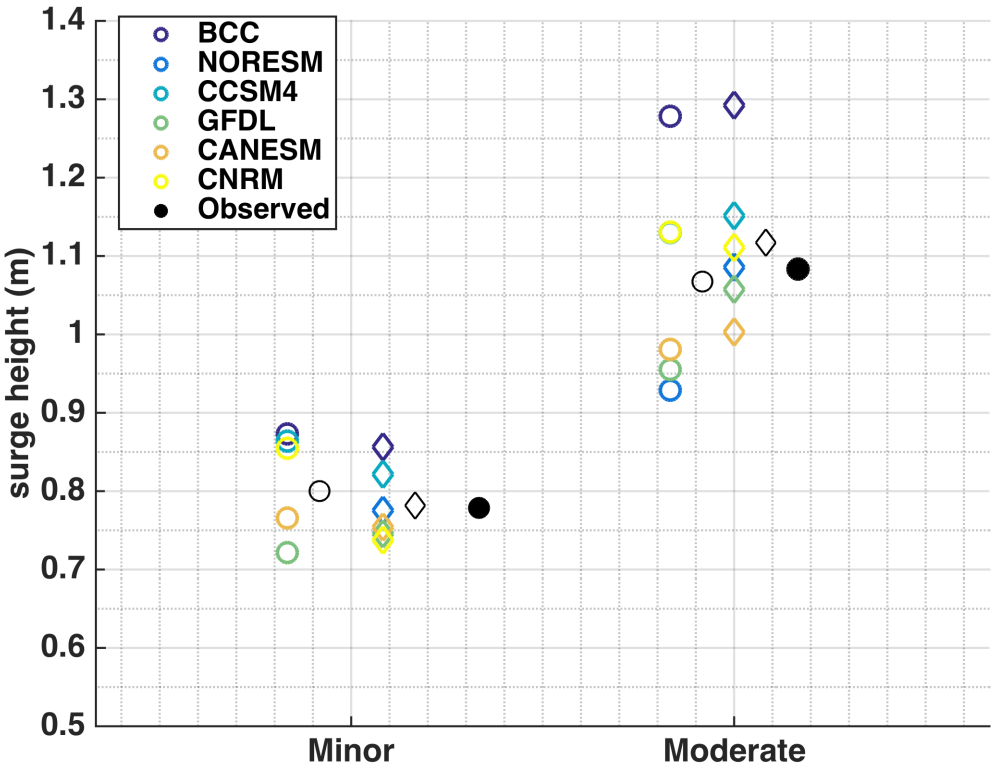


Figure 2.6: The mean of 3-h-predicted storm surge intervals ≥ 0.61 m (impactful) and surge intervals ≥ 1.0 m (moderate) using raw (hollow circles) and bias-corrected (hollow diamonds) atmospheric data from GCMs during the historical period (1979-2004). The black circle indicates the corresponding observed quantity at The Battery calculated over 1979-2004 Nov.-March and the black diamond next to the hollow diamonds indicates the ensemble mean.

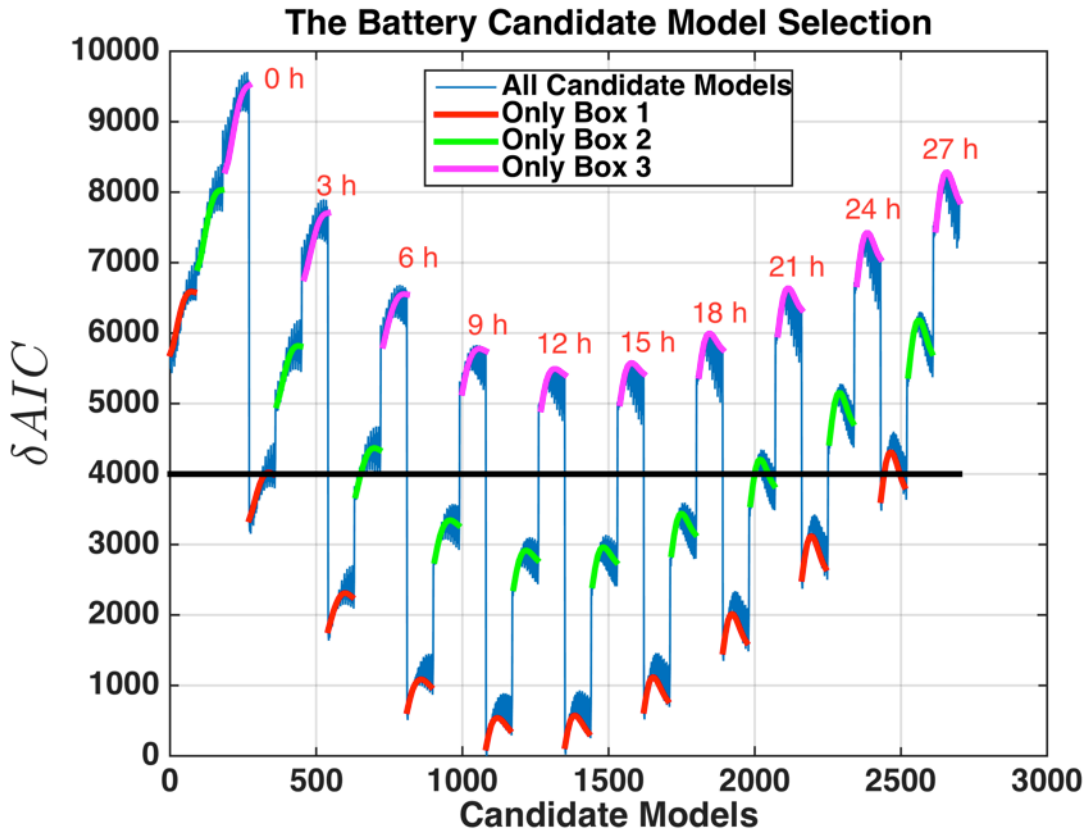


Figure 3.1: ΔAIC_i for the 2,700 candidate models calculated with NARR CV1 data for The Battery. The solid colored lines represent the ΔAIC_i for MLRs that are created with predictors from the same box (see legend). The red text above each of the 10 clusters indicate the *Lx* sum length in hours of the MLR used to calculate that cluster in the ΔAIC_i curve.

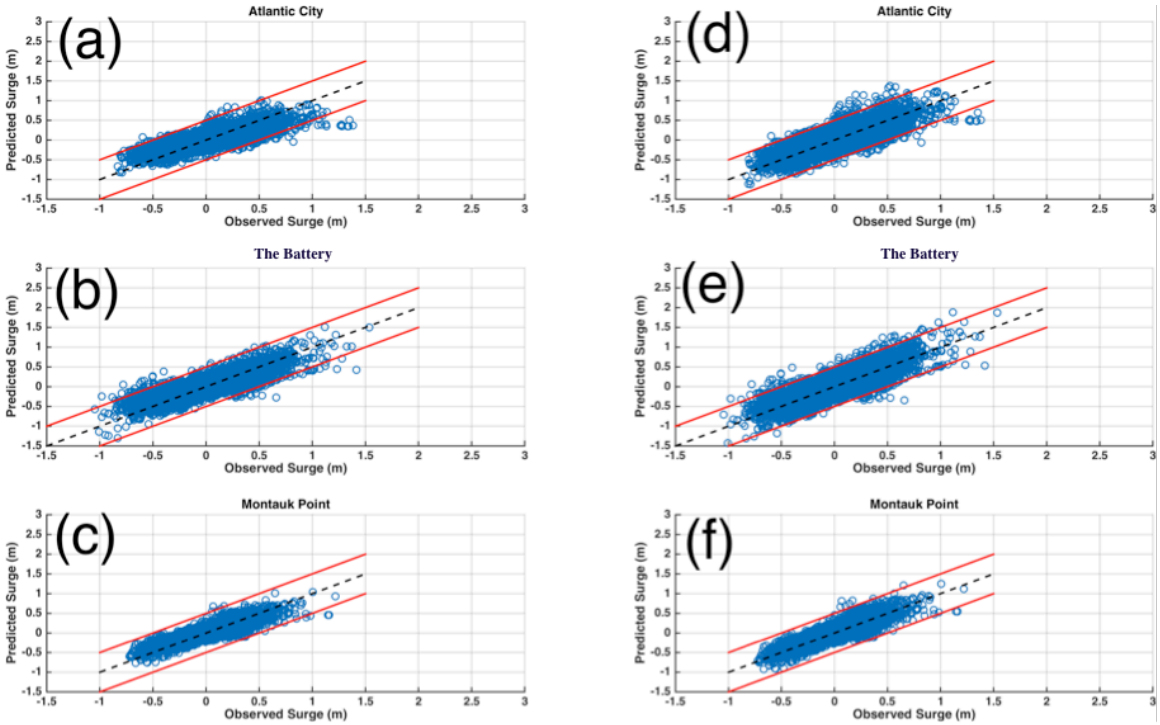


Figure 3.2: Scatterplots of observed surge on the abscissa and predicted surge on the ordinate in meters for the verification period using NARR CV1 data configuration for (a,d) Atlantic City, (b,e) The Battery, and (c,f) Montauk Point. The left column (a-c) are for MLR verification predictions and the right column (d-f) are for BC-MLR verification predictions. The dotted black line serves as the 1:1 reference line. The red lines parallel to the 1:1 line demarcate the region of error that are $\leq |0.5\ m|$

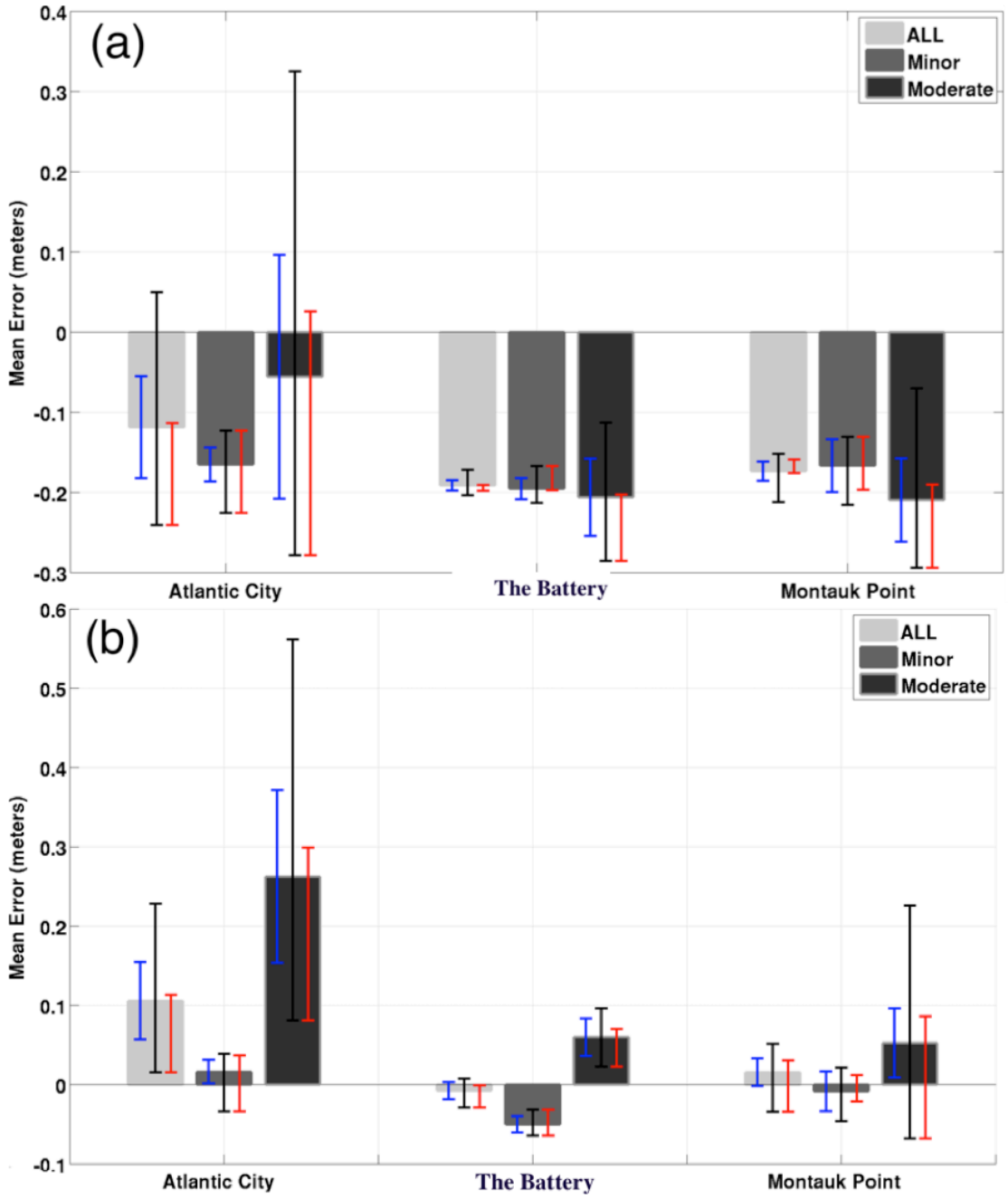


Figure 3.3: Binned ME for all, minor, and moderate events for verification predictions made with (a) MLR and with (b) BC-MLR at Atlantic City, The Battery, and Montauk Point. The red error bars represent the sampling uncertainty, blue error bars represent analysis uncertainty, and black error bars represent total uncertainty.

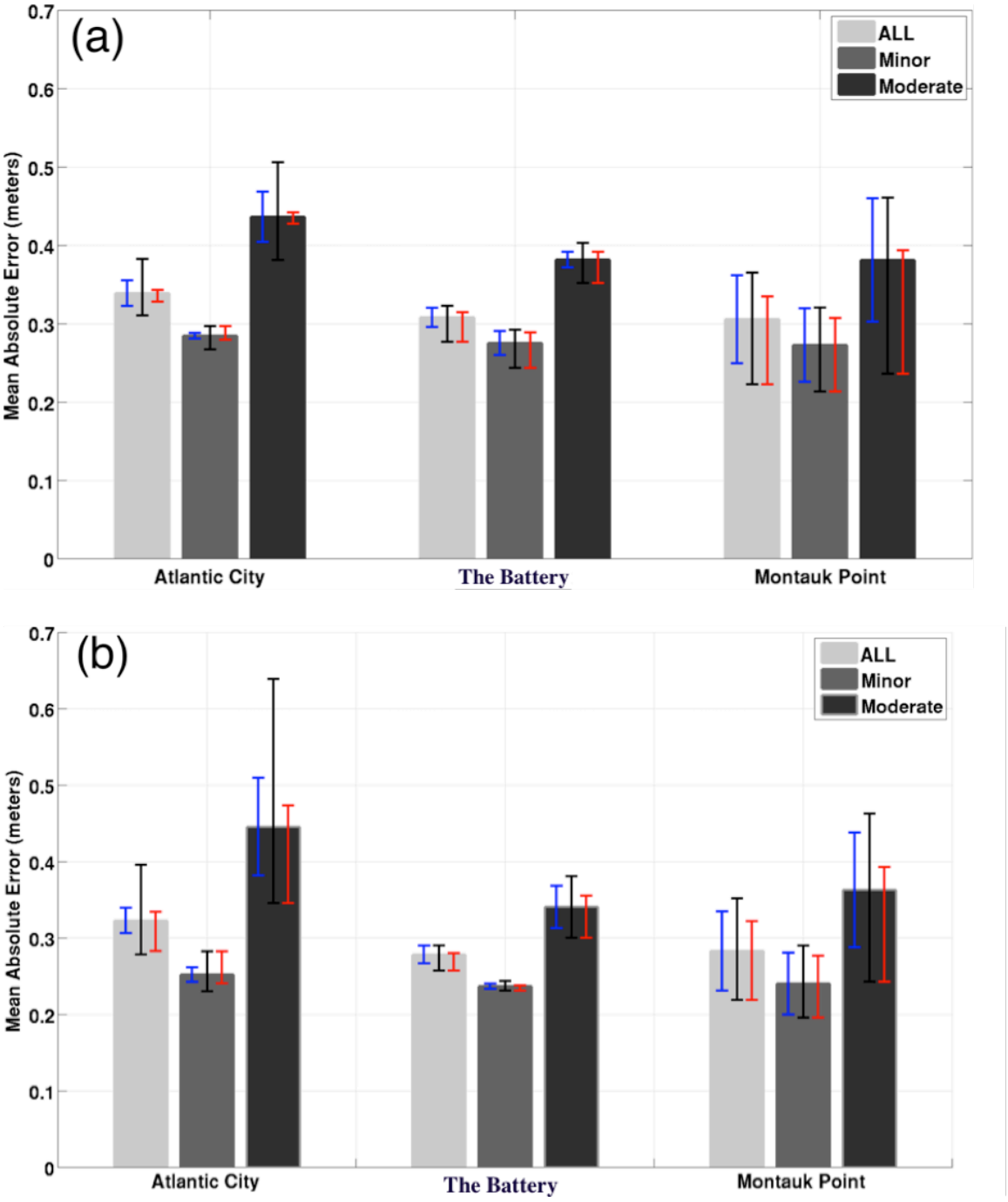


Figure 3.4: Same as 3.3 but for MAE.

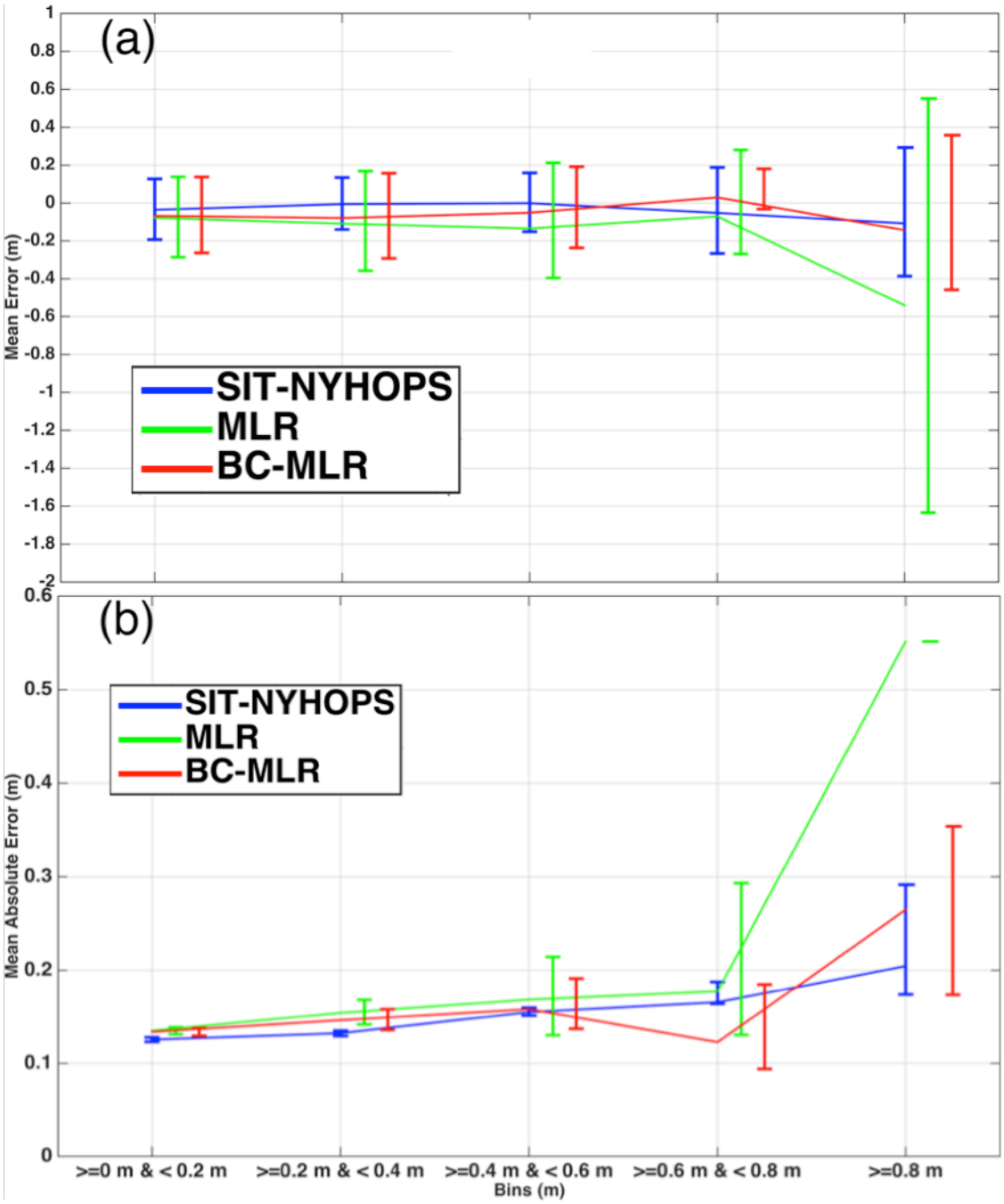


Figure 3.5: The (a) binned mean error and (b) mean absolute error calculated for raw SIT-NYHOPS (blue), MLR (green), and BC-MLR (red) to predict storm surge at The Battery for lead times between 3-12 h leads between Oct.-March 2010-2014. The 95th percentile ranges denoted by the error bars for each bin are calculated by bootstrapping 1,000 iterations.

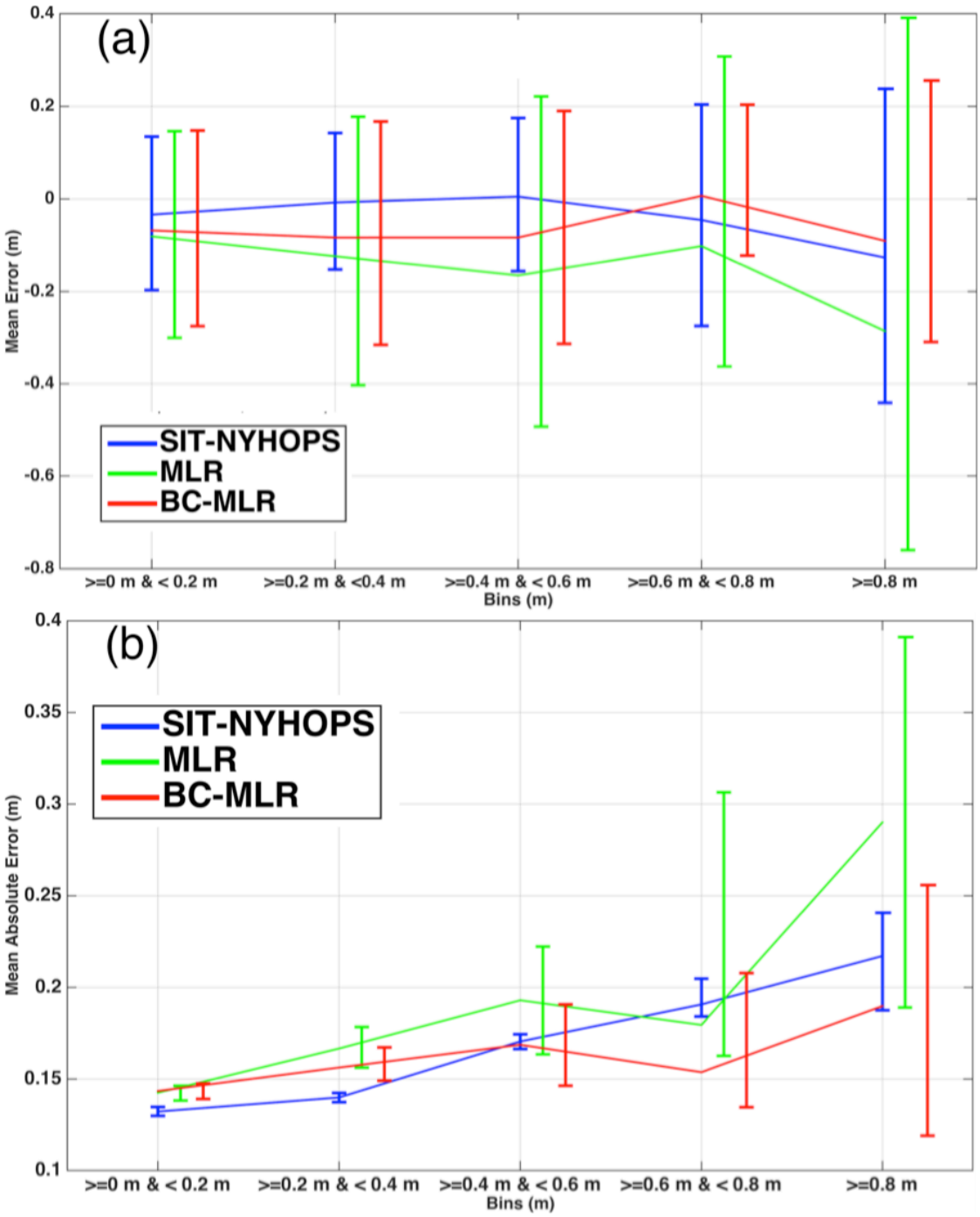


Figure 3.6: Same as figure 3.5 but for 15-24 h leads between Oct.-March 2010-2014.

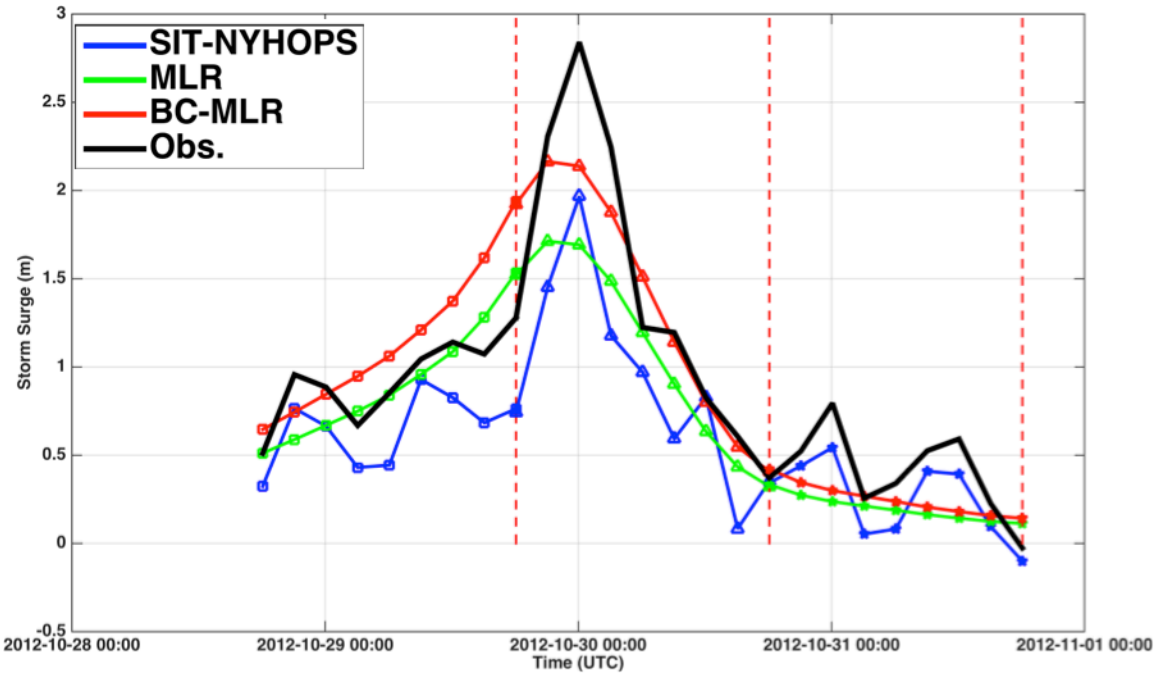


Figure 3.7: Raw SIT-NYHOPS (blue line), CFSR CV1 MLR (green line), and CFSR CV1 BC-MLR (red line) The Battery storm surge forecasts plotted every 3 hours for forecasts initialized on 1800 UTC on 28 October 2012 (lines with squares), on 1800 UTC 29 October 2012 (lines with triangles), on 1800 UTC 30 October 2012 (lines with pentagrams) each valid for 24 hours after the time of initialization, which is demarcated by the red-dotted vertical lines. Three-hourly observations of storm surge in meters at The Battery are plotted as thick black lines.

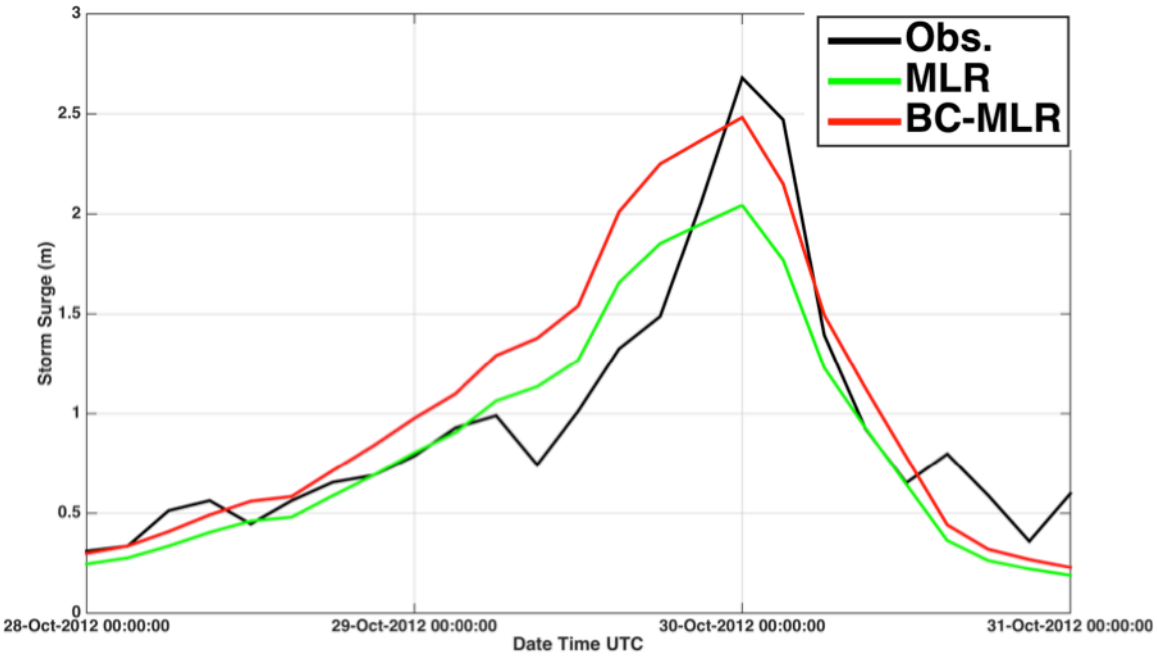


Figure 3.8: Three-hourly MLR (green) surge predictions, BC-MLR (red) surge predictions, and observed surge (black) for The Battery between 0000 UTC 28 October 2012 to 0000 UTC 31 October 2012. Storm surge predictions are calculated by forcing The Battery's NARR CV1 model with NARR data.

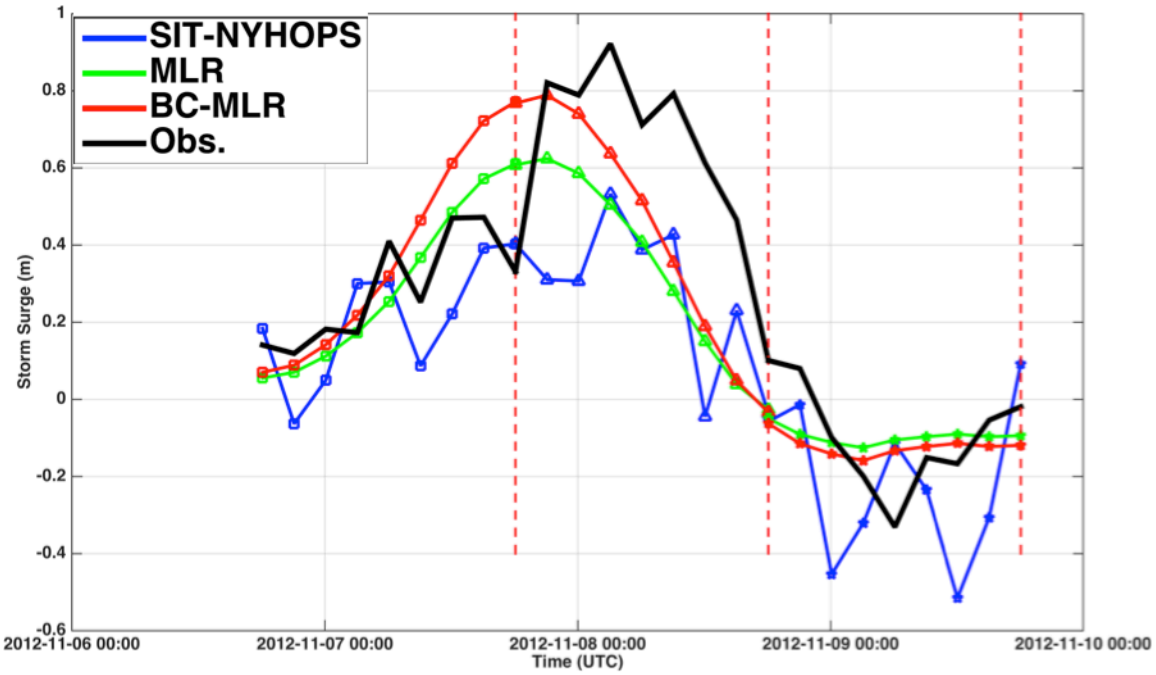
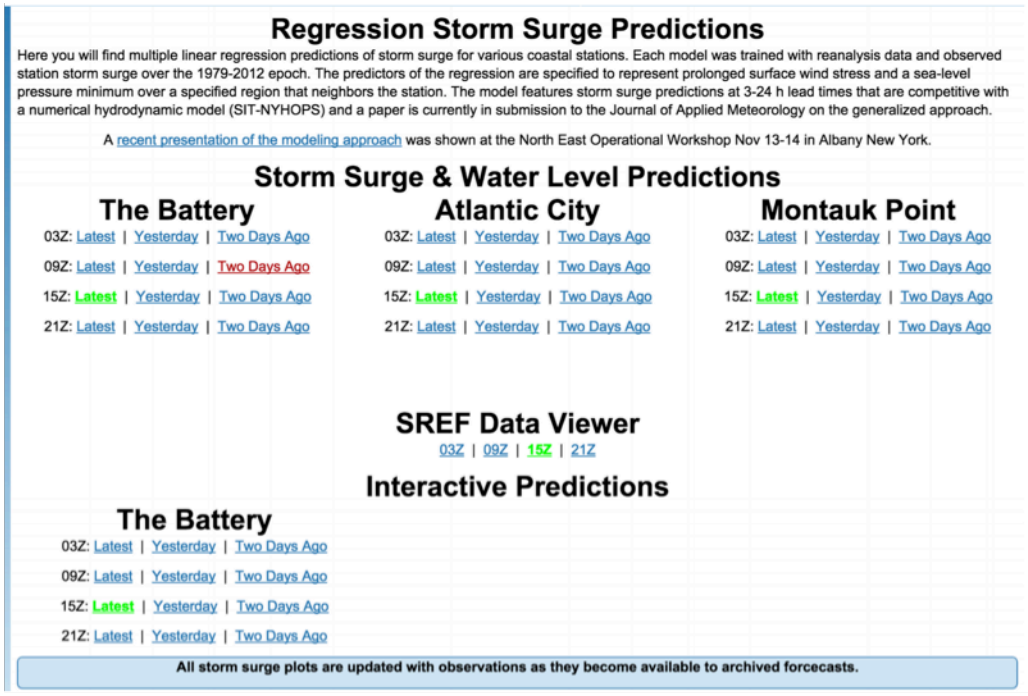


Figure 3.9: Raw SIT-NYHOPS (blue line), CFSR CV1 MLR (green line), and CFSR CV1 BC-MLR (red line) The Battery storm surge forecasts plotted every 3 hours for forecasts initialized on 1800 UTC Nov 6 2012 (lines with squares), on 1800 UTC Nov 7 2012 (lines with triangles), on 1800 UTC Nov 8 2012 (lines with pentagrams) each valid for 24 hours from the time of initialization, which is demarcated by the red-dotted vertical lines. Three-hourly observations of storm surge in meters at The Battery are plotted as thick black lines.



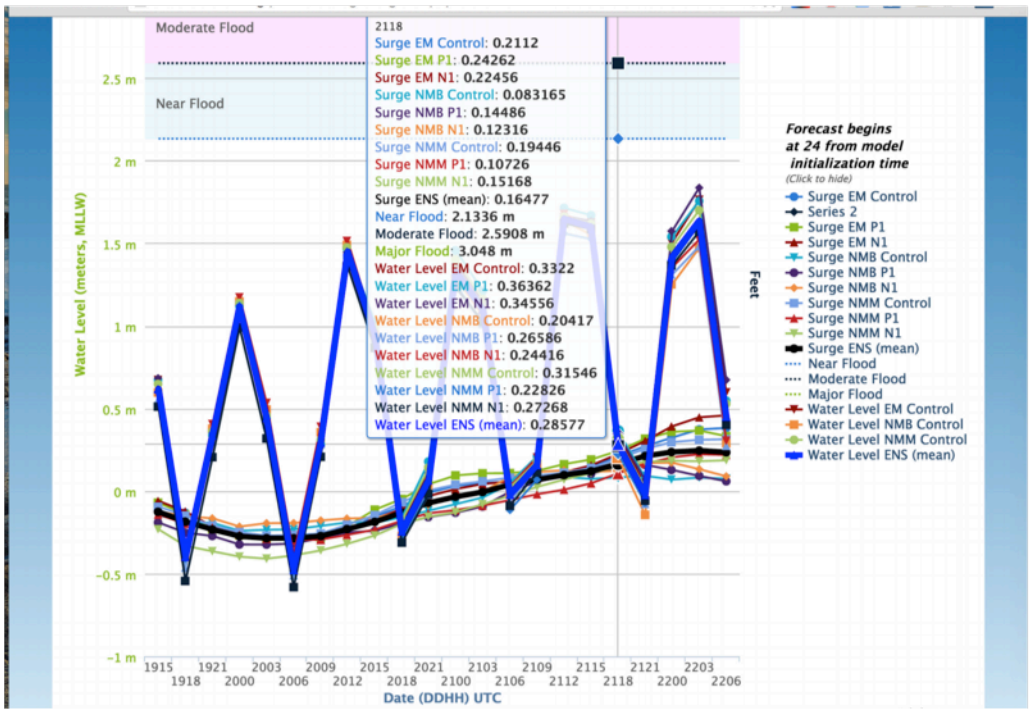


Figure 3.10: (a) A screenshot of the operational website illustrating the layout and stations predictions are made for. (b) An example of an interactive storm surge and water level forecast with flooding levels turned on.

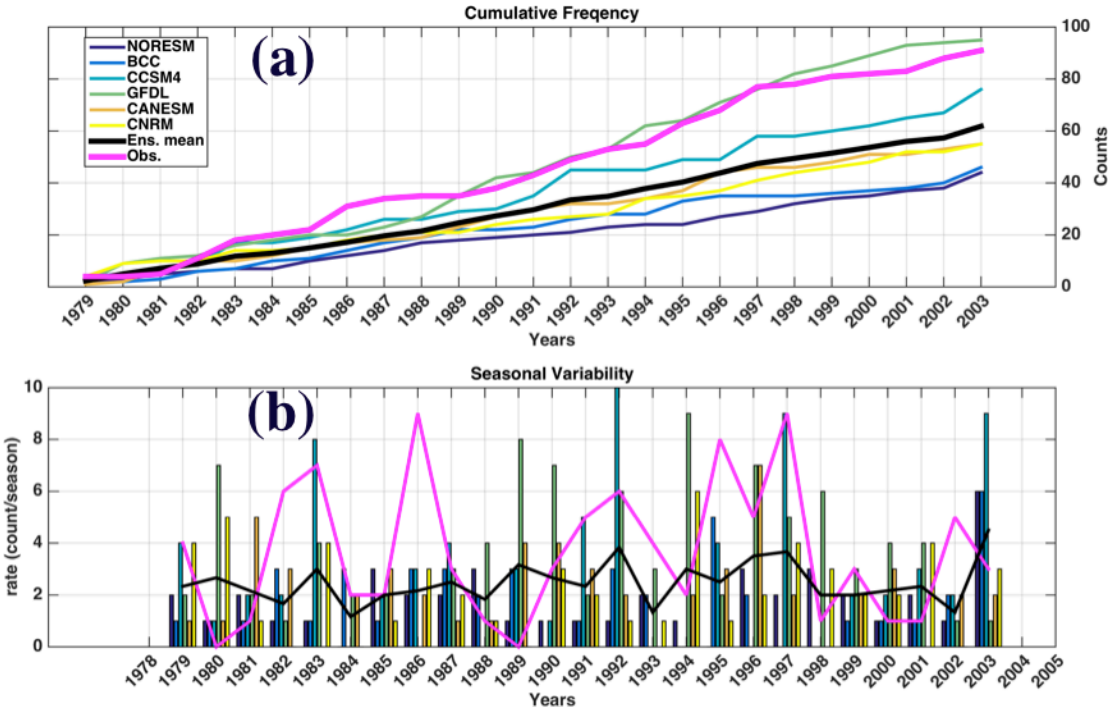


Figure 4.1: (a) The cumulative distribution of counts for modeled and observed data and (b) the interseasonal variability or rate (counts/season).

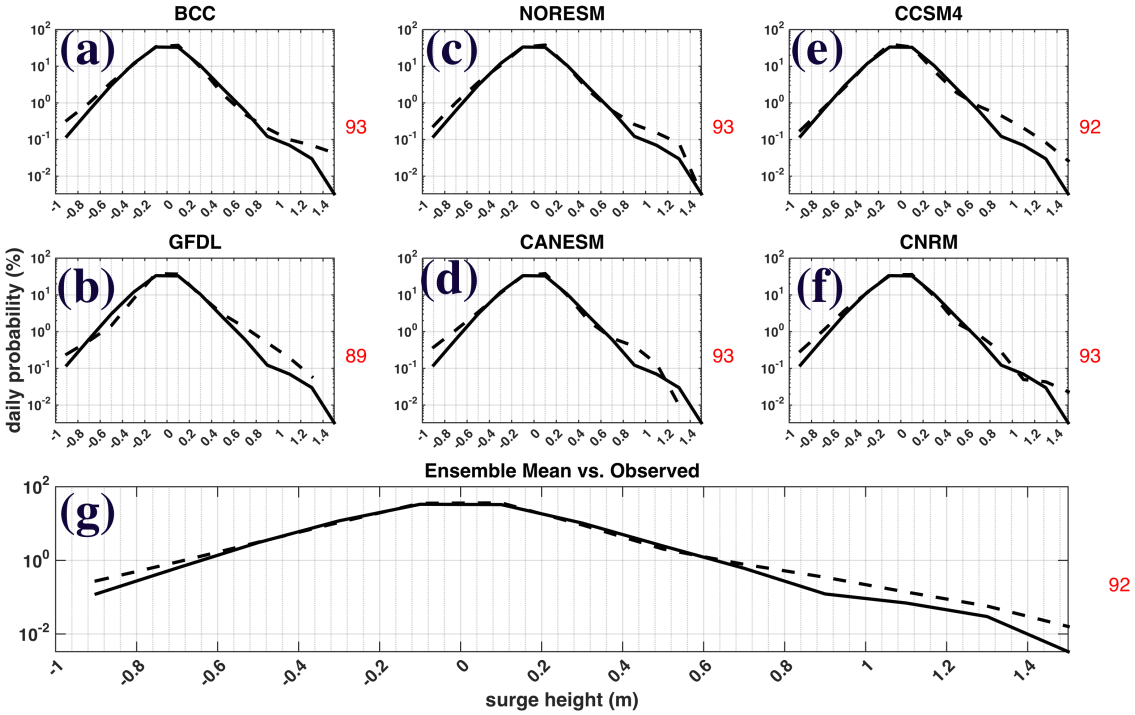


Figure: 4.2: The daily probability of a 3-h surge height for modeled (a-f) surge and (g) ensemble mean. In each plot, the solid black line represents observed surge and the dotted black line represents the modeled surge. Skill scores in percent are annotated as red text in the corner of each sub plot and calculated using the modeled data compared with the observed during the historical period.

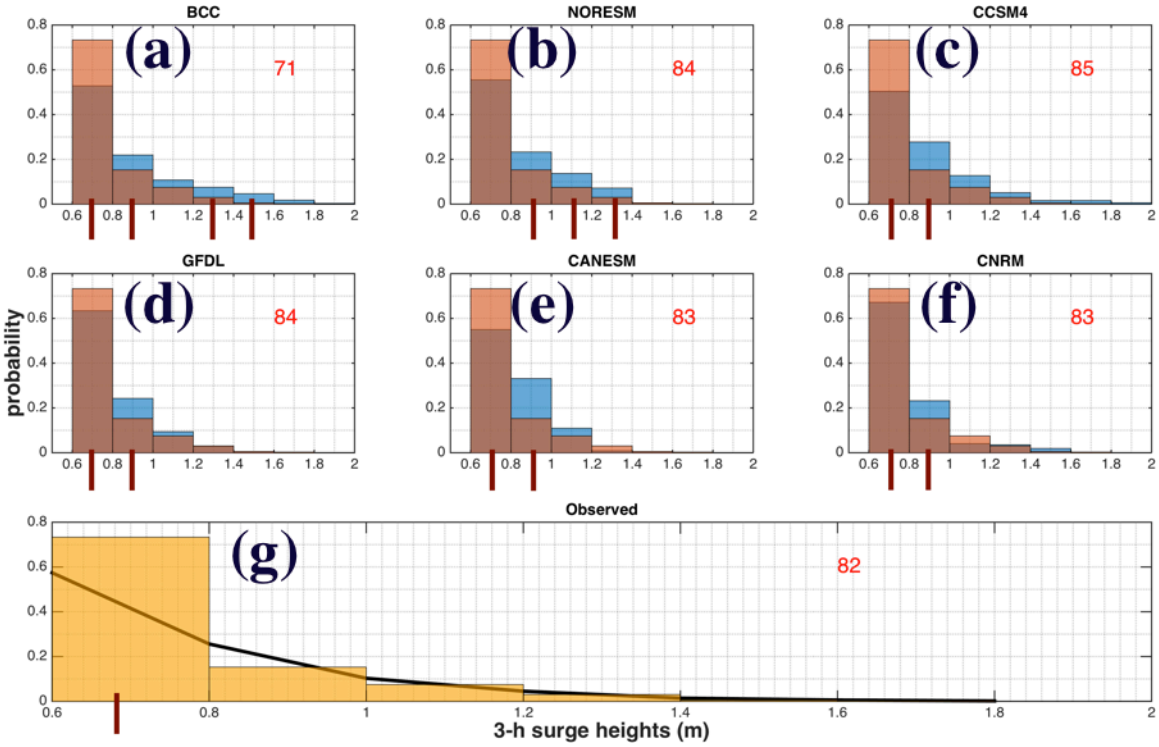


Figure 4.3: The probability of storm maximum surge meeting the bin surge height for modeled (a-f) and observed (g) impactful surge events. This plot is referred to as an intensity distribution. Each histogram (a-f) is overlaid in orange with the intensity distribution of observed data at The Battery during the historical period. The black line in (g) indicates the ensemble mean. Skill scores in percent are annotated as red text in the corner of each sub plot. Binned frequencies that are statistically insignificant to the corresponding observed binned frequencies at the $\propto = 0.05$ level are marked with a red vertical line underneath the bin.

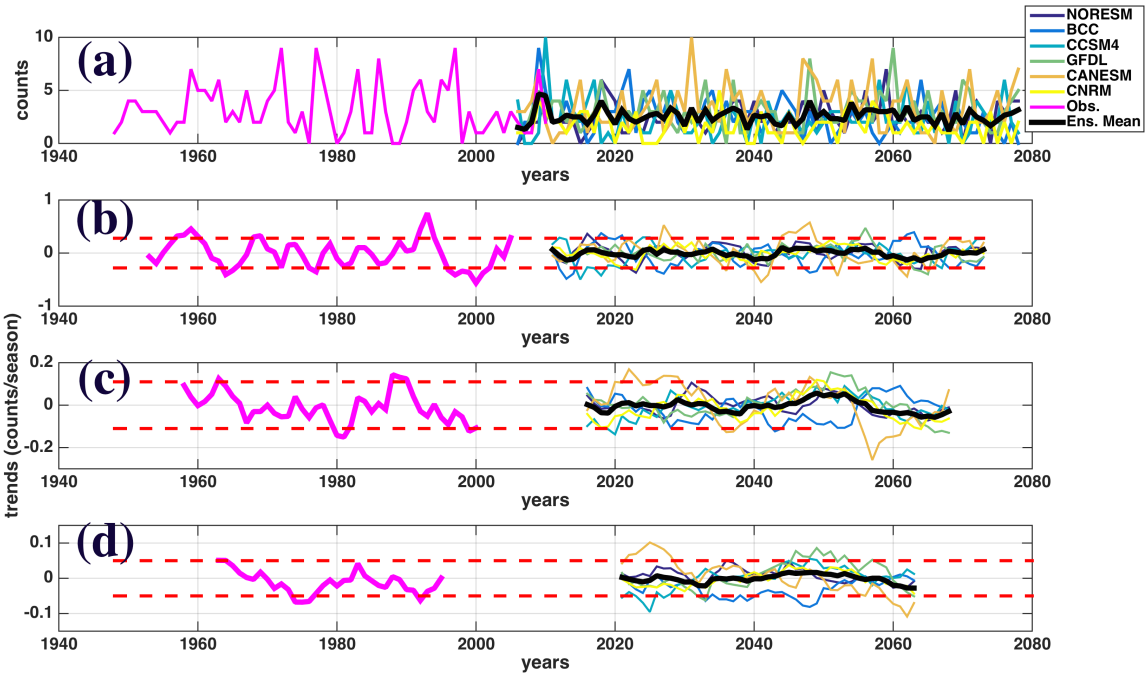


Figure 4.4: (a) Counts per season, (b) 10-y trends (counts/season), (c) 20-y trends (counts/season), and (d) 30-y trends (counts/season) for each GCM and observations during 1948-2079. Both (b), (c), (d) are calculated using centered sliding windows of length n years (where n is the sliding window length) where each rate corresponds to the preceding and following n/2 years. The trend necessary to be considered significant at the $\propto = 0.05$ is indicated with red-dotted lines in (b), (c), and (d). A Monte Carlo approach was used to determine the region of statistical insignificance or the linear trend that would reject the null hypothesis of a t-test given that the sample of the sliding window was less than 30.

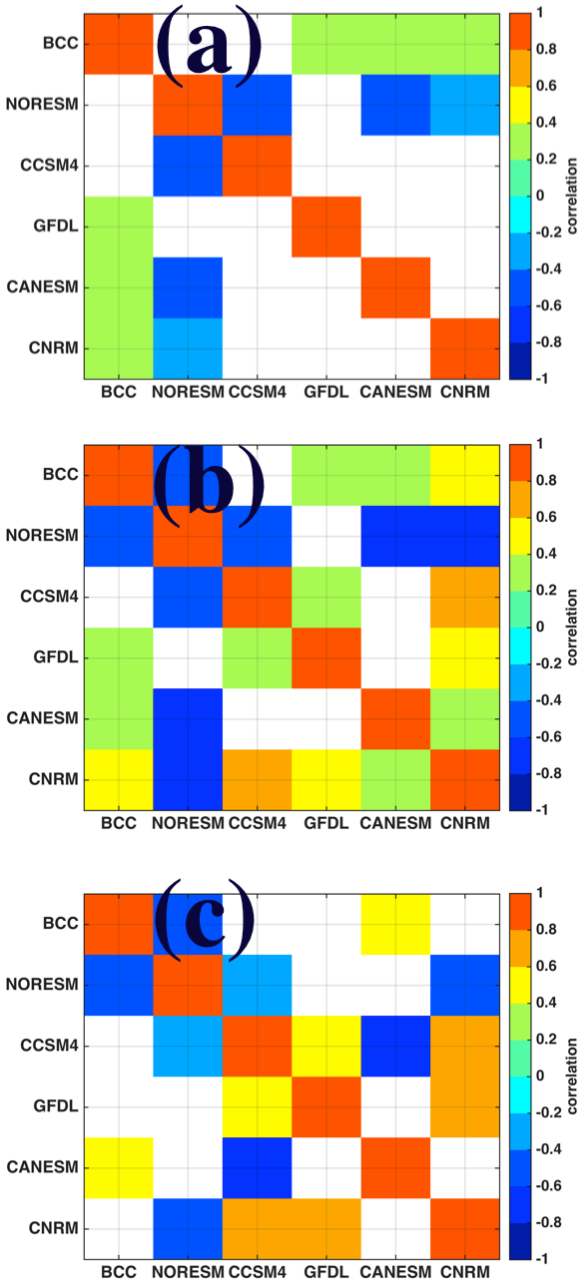


Figure 4.5: Psuedo-color checkerboards of the correlation matrix calculated between the six GCMs for (a) 10-y, (b) 20-y, and (c) 30-y centered sliding window trends of surge counts. Only statistical significant ($\propto = 0.05$) correlations are shown.

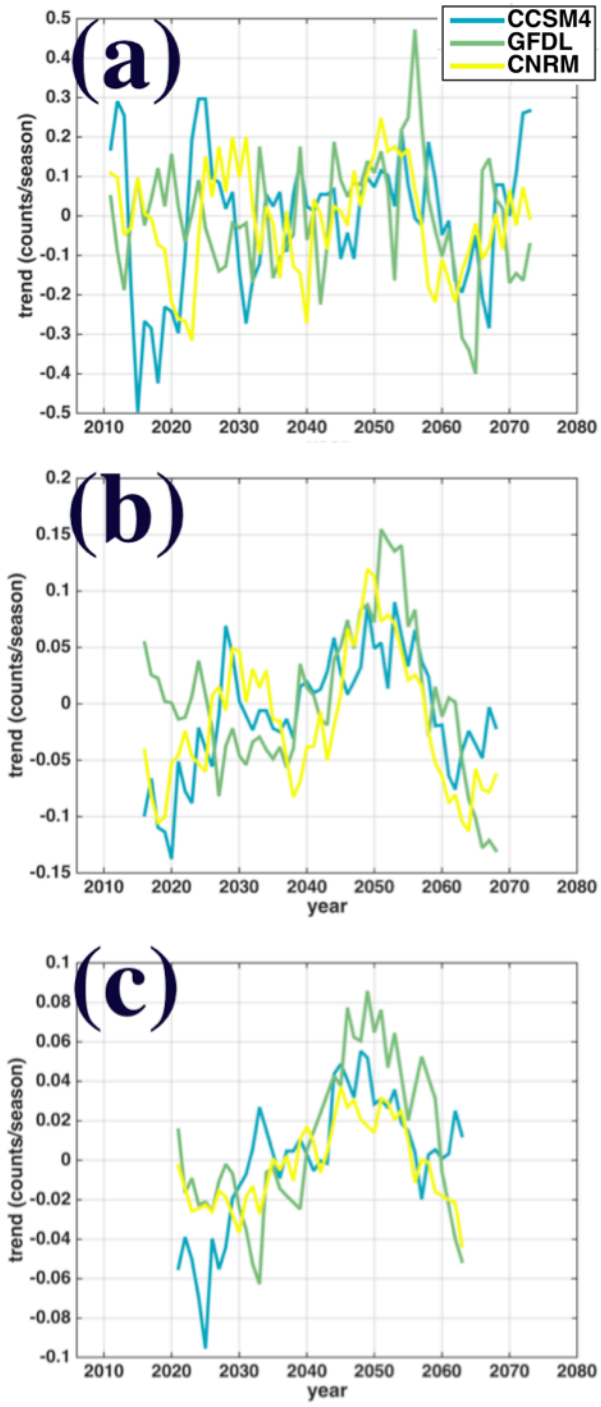


Figure 4.6: Select models (CCSM4, GFDL, CNRM) and their centered sliding window (a) 10-y trends, (b) 20-y trends, (c) 30-y trends in counts/season for the future period.

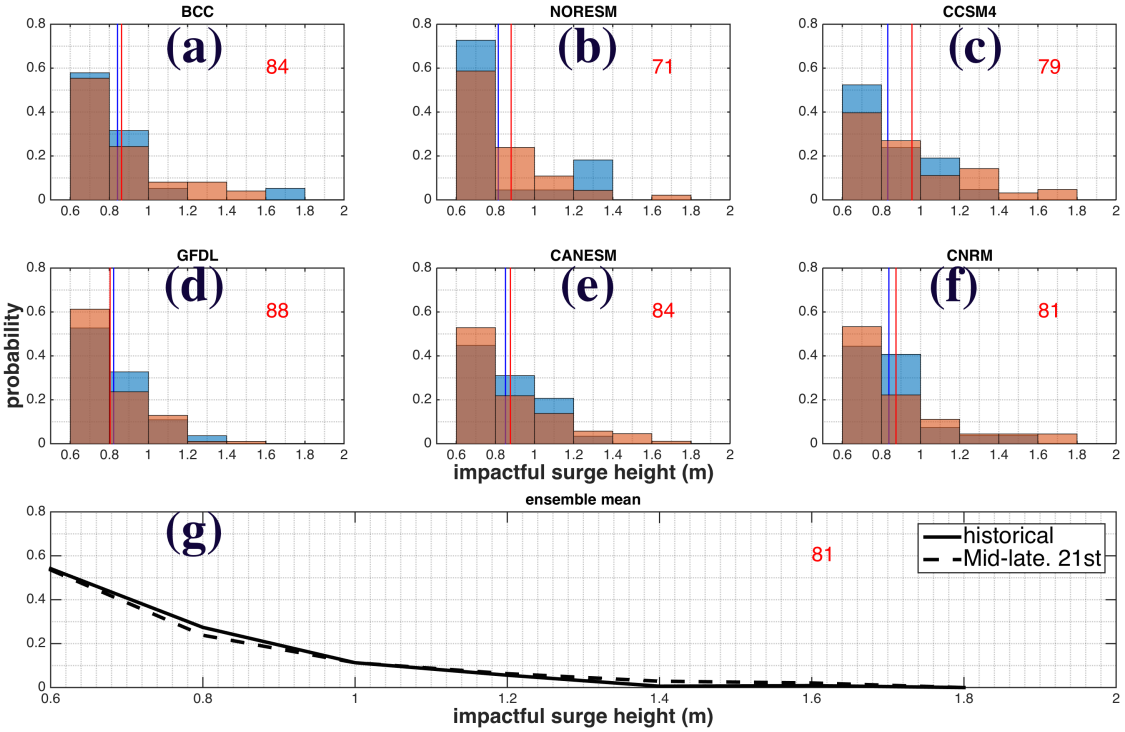


Figure 4.7: (a-f) Storm maximum surge intensities of impactful surge events for the historical period (1979-2004) compared to the mid.-late 21st century period (2054-2079) for all GCMs The skill score is annotated in the top right of each sub plot calculated by comparing the histograms of impactful surge events for the historical period against the late future period. (g) The ensemble mean intensity distribution of the historical period is compared with the ensemble mean of the late-middle 21st Century period. In each panel, the blue vertical line is the historical mean and the red vertical line represents the mid.-late future period mean.

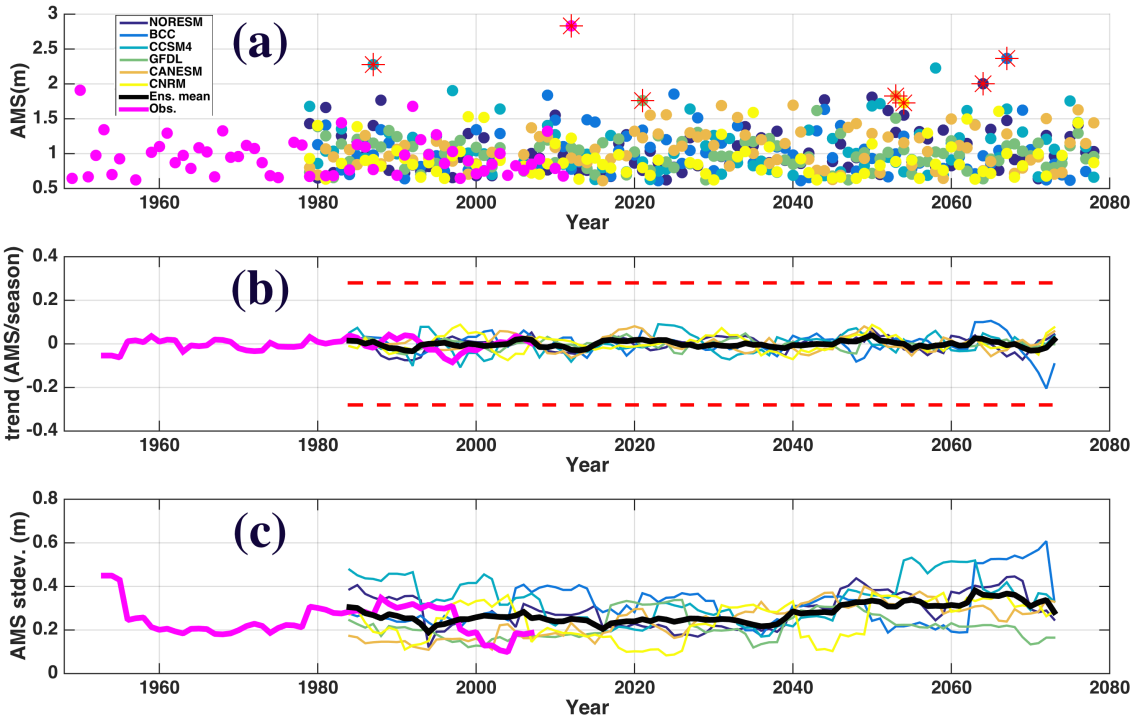


Figure 4.8: (a) The annual maximum surge (AMS) height (m) for each GCM and observed, (b) the 10-y trend calculated using a centered sliding window (AMS/season), and (c) the standard deviation in AMS calculated using a 10-y centered sliding window. The largest surges throughout 1948-2079 for each GCM and observed in (a) are marked with red asterisks.

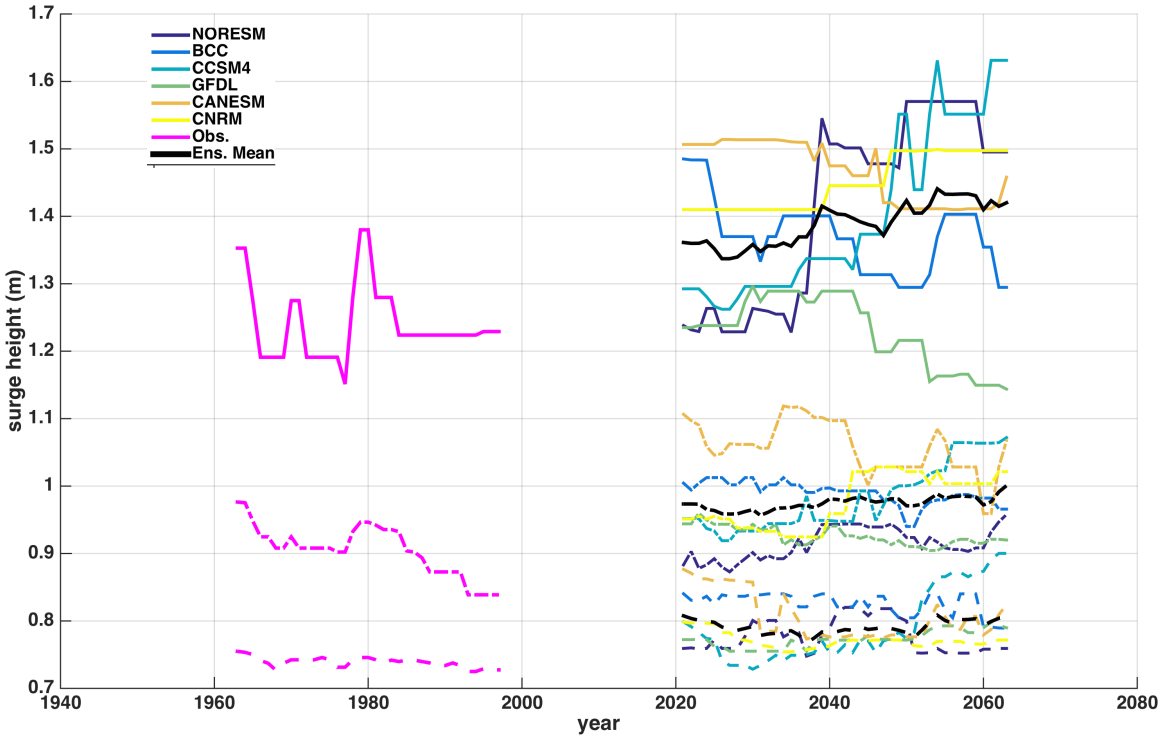


Figure 4.9: 30-y centered sliding window percentiles of observed and modeled storm maximum surges between 1948-2012. The 50th percentile is represented with dotted lines, the 75th percentile is represented with a dash dotted line, and the 95th percentile is represented with a solid line.

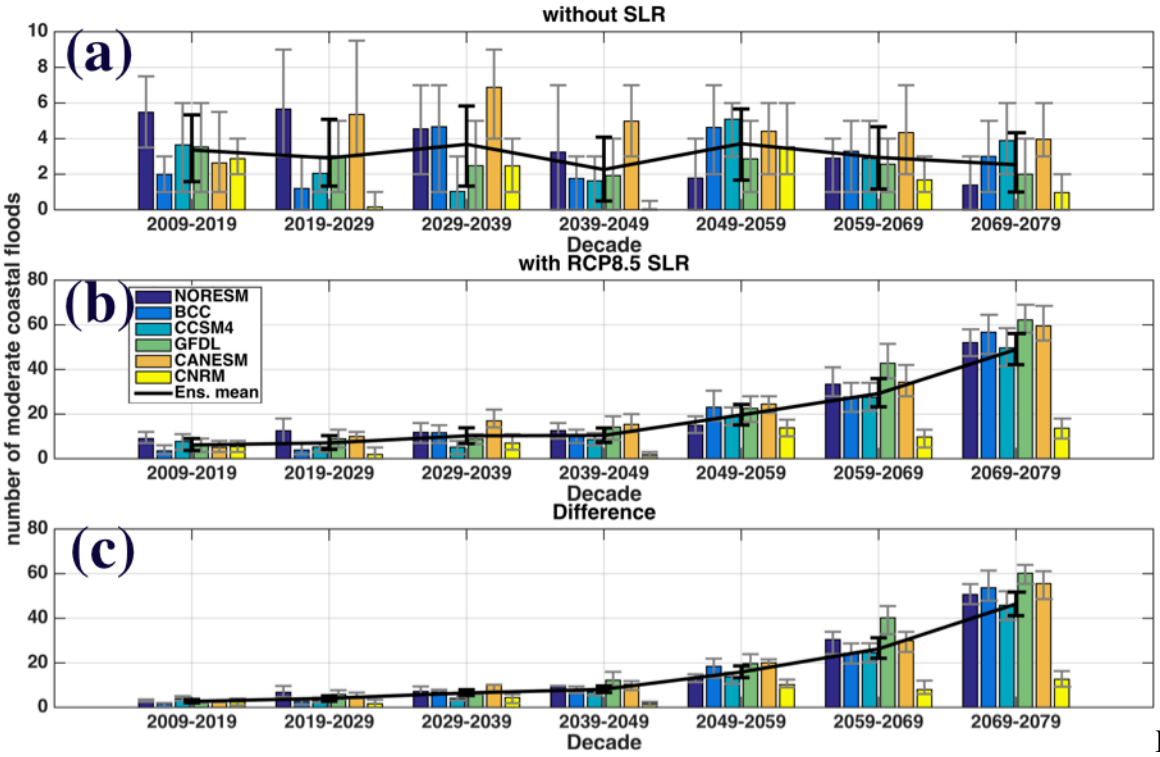


Figure 4.10: Predictions of moderate coastal floods per decade from 2009-2079 (a) without sea level rise (SLR), (b) with SLR consistent with RCP8.5 global warming, and (c) the difference between (b) – (a). Vertical bars represent 95th uncertainty due to tidal variations.

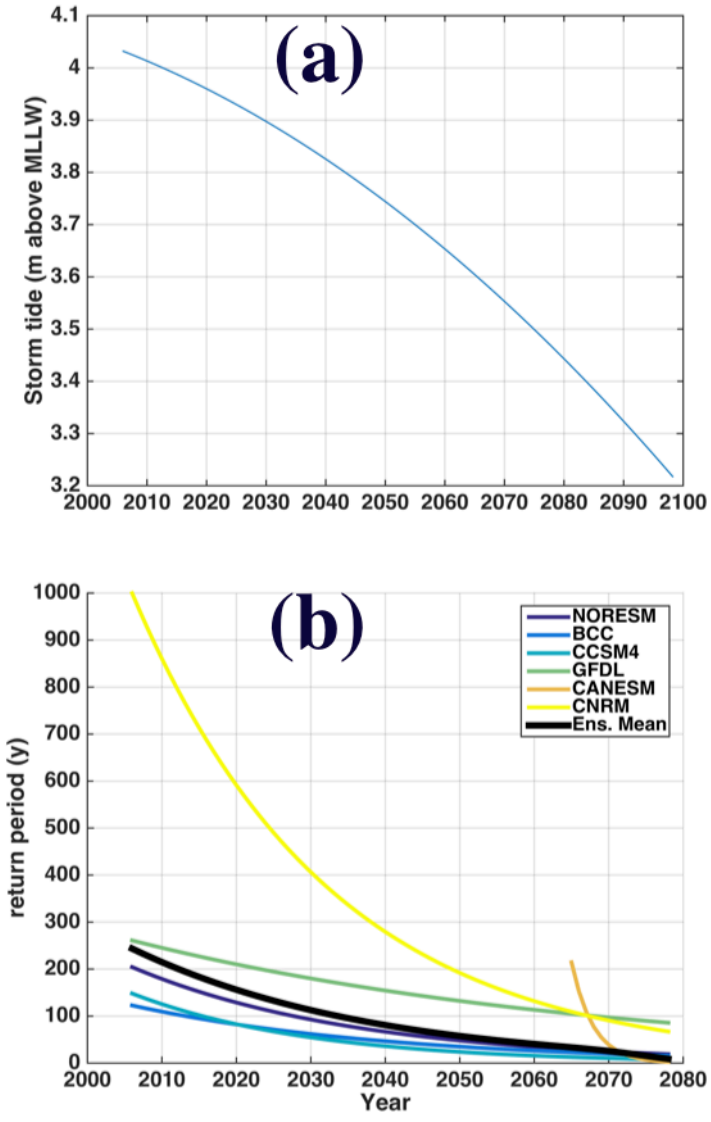


Figure 4.11: (a) The storm tide necessary to generate an equivalent depth coastal flood to that of Sandy's given regional sea level rise, tidal uncertainty, and global warming at The Battery into the 21st century. (b) The return period of an equivalent depth coastal flood to that of Sandy's.

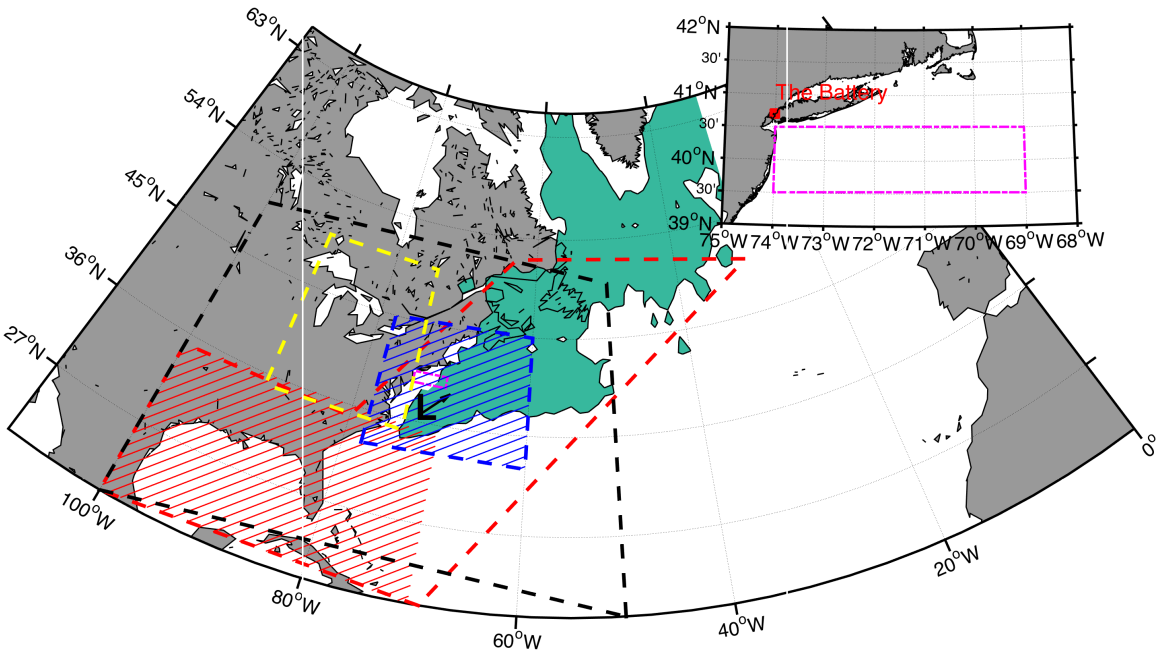


Figure 4.12: The green shaded is the north Atlantic cyclone track and demarcates the region where > 100 CFSR cyclones passed between 1979-2004 Nov.-March. The red polygon and hatched region denote spatial regions for the classification of Miller Type A cyclones and the yellow dotted and blue hatched polygon for Miller Type B cyclones. The inset is a close-up of the local region around The Battery and the blue highlighted coastline identifies the New York/New Jersey coastline. The rectangular magenta box in the inset is referred to as the predictor box and is used to collect data for the multi-linear regression's predictors.

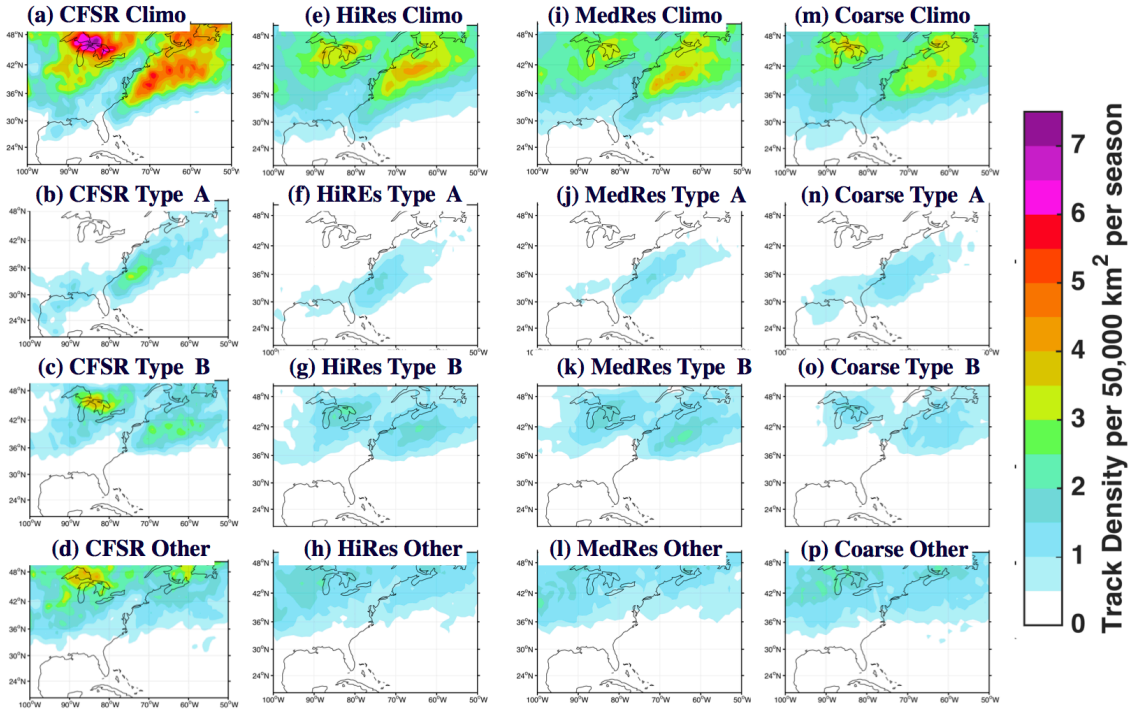


Figure 4.13: The cyclone track density per season calculated over the historical period (1979-2004 Nov.-March) for (a,e,i,m) climatology, (b,f,j,n) Miller Type A, (c,g,k,o) Miller Type B, and (d,h,l,p) other. HiRes represents the track density mean of CCSM4 and BCC, MedRes represents a mean of NORESM and GFDL, Coarse represents a mean of CANESM and BCC.

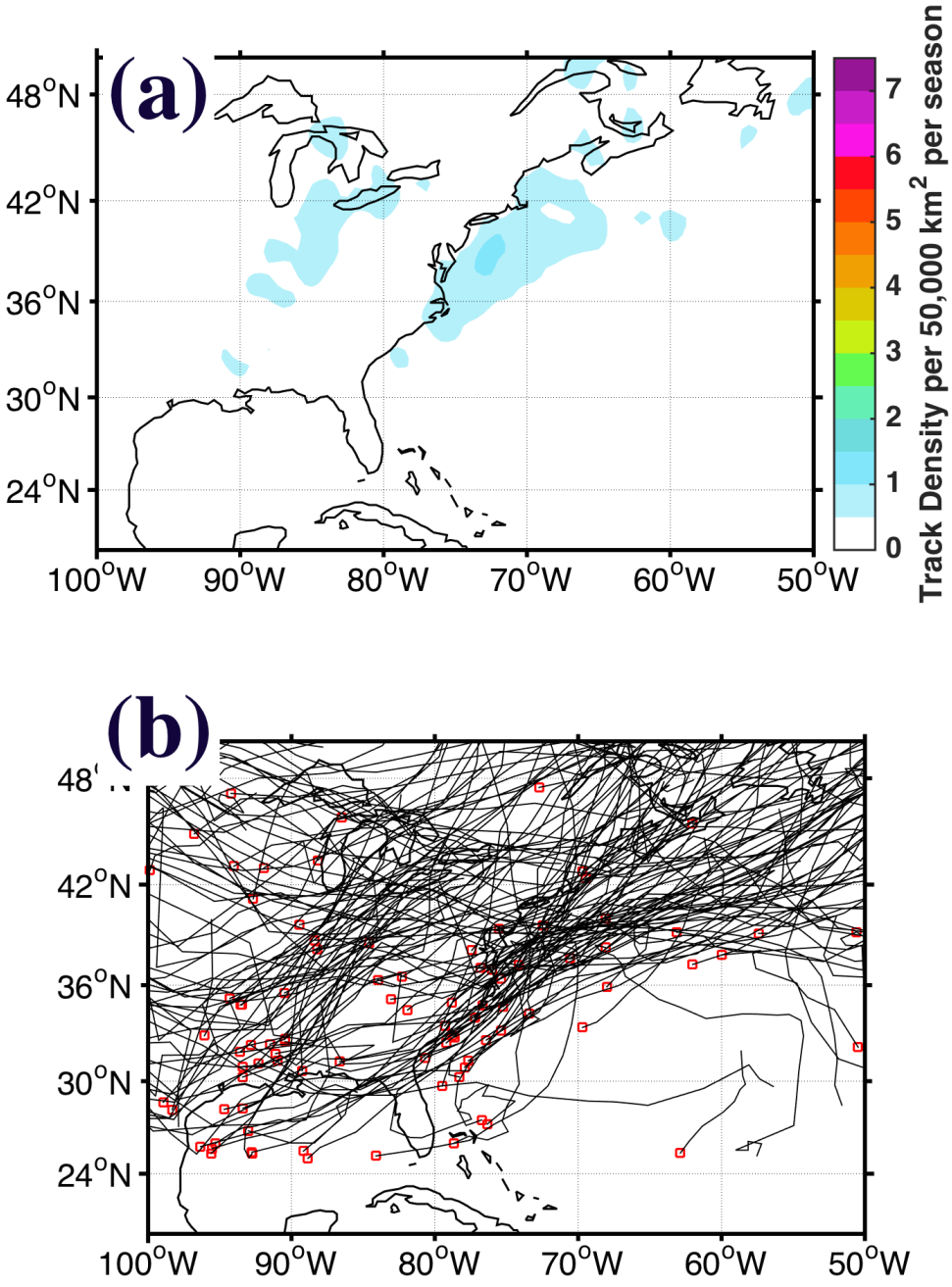


Figure 4.14: (a) The seasonal (Nov.-March) impactful surge-generating CFSR track density and (b) the 155 CFSR tracks that occurred during 75 observed impactful surge events. Red squares in (b) denote the first point tracked by the tracker.

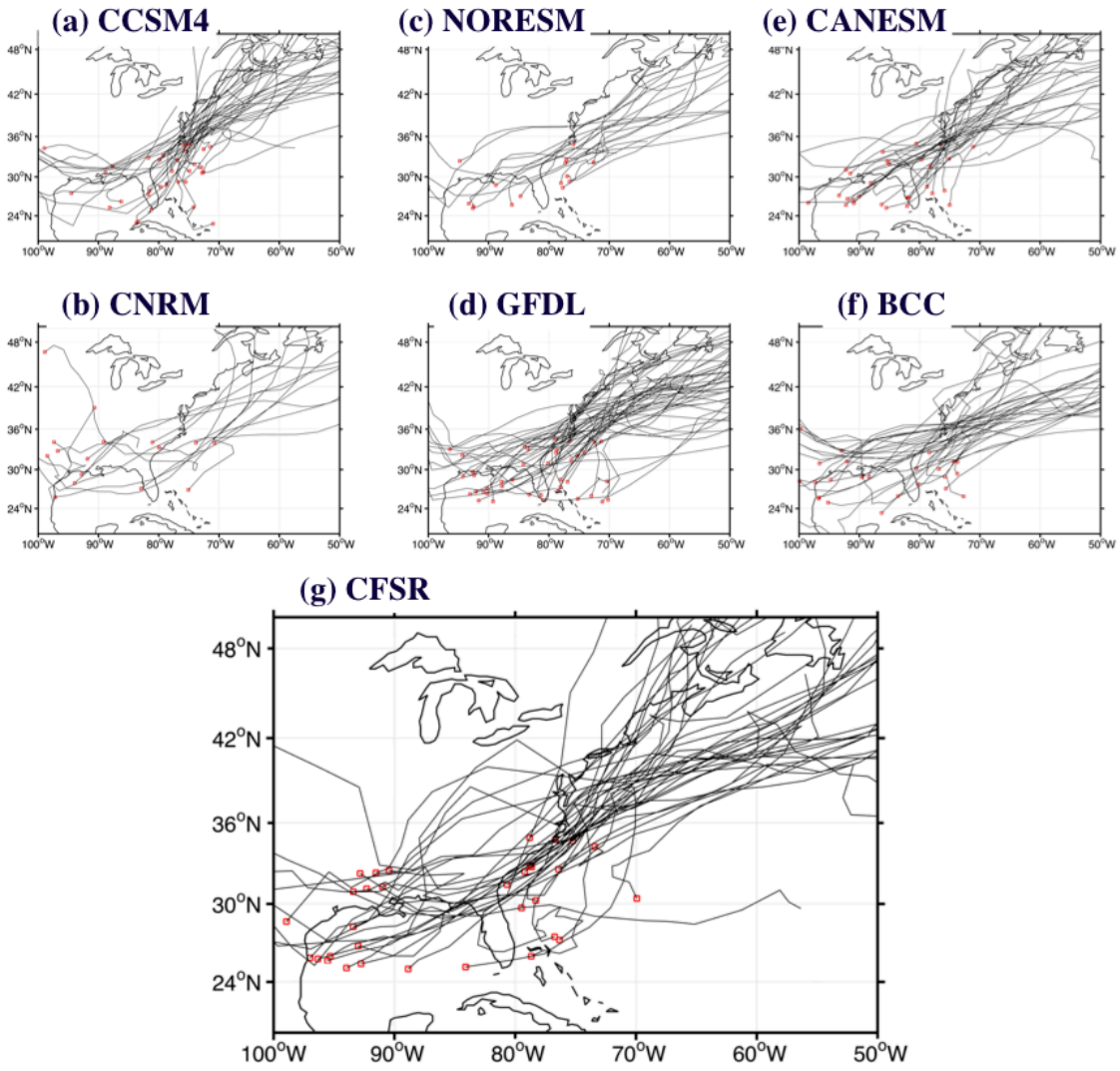


Figure 4.15: Miller Type A tracks that generated impactful surges (≥ 0.61 m) at The Battery during the historical period for (a-f) six GCM's modeled impactful surges and for (g) CFSR with observed surges. Red squares denote the genesis of the track.

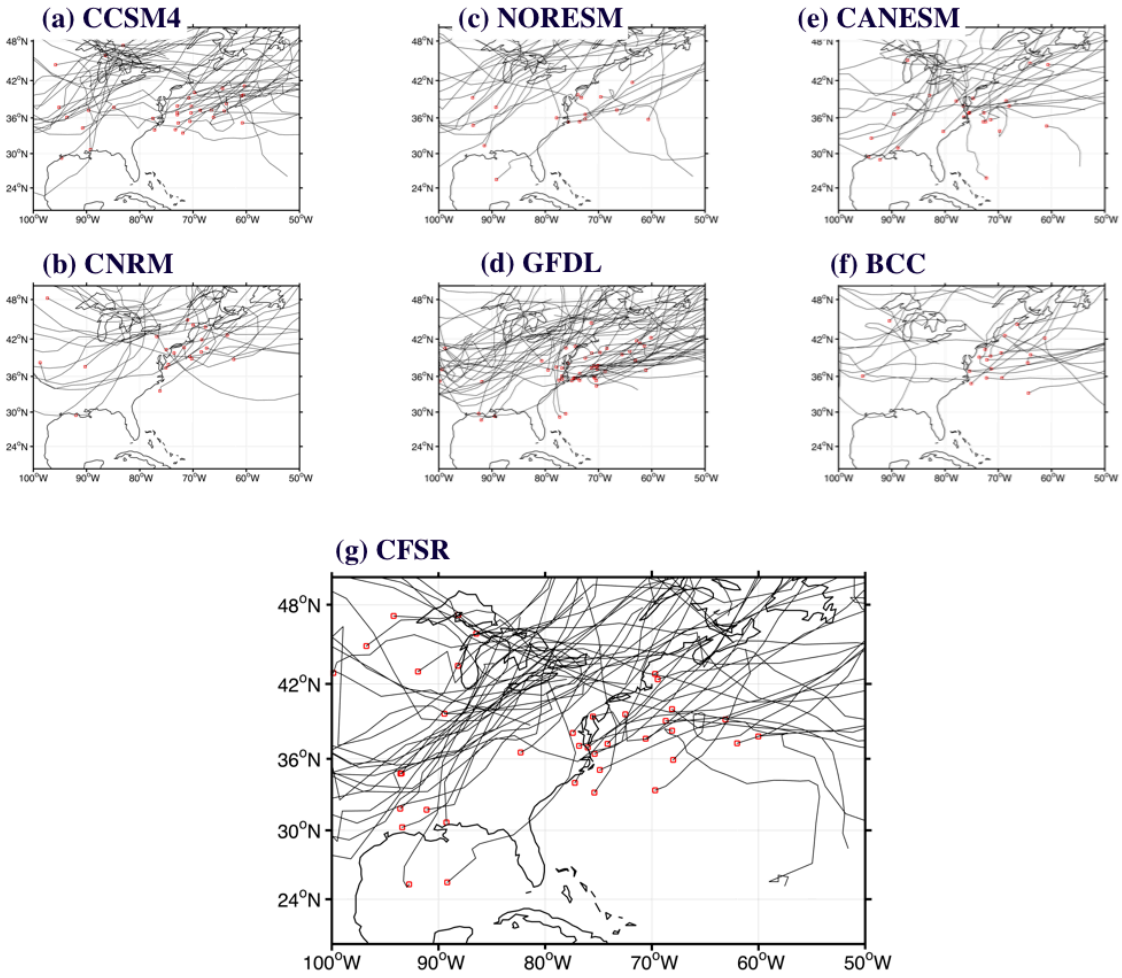


Figure 4.16: Same as 4.15 but for Miller Type B tracks.

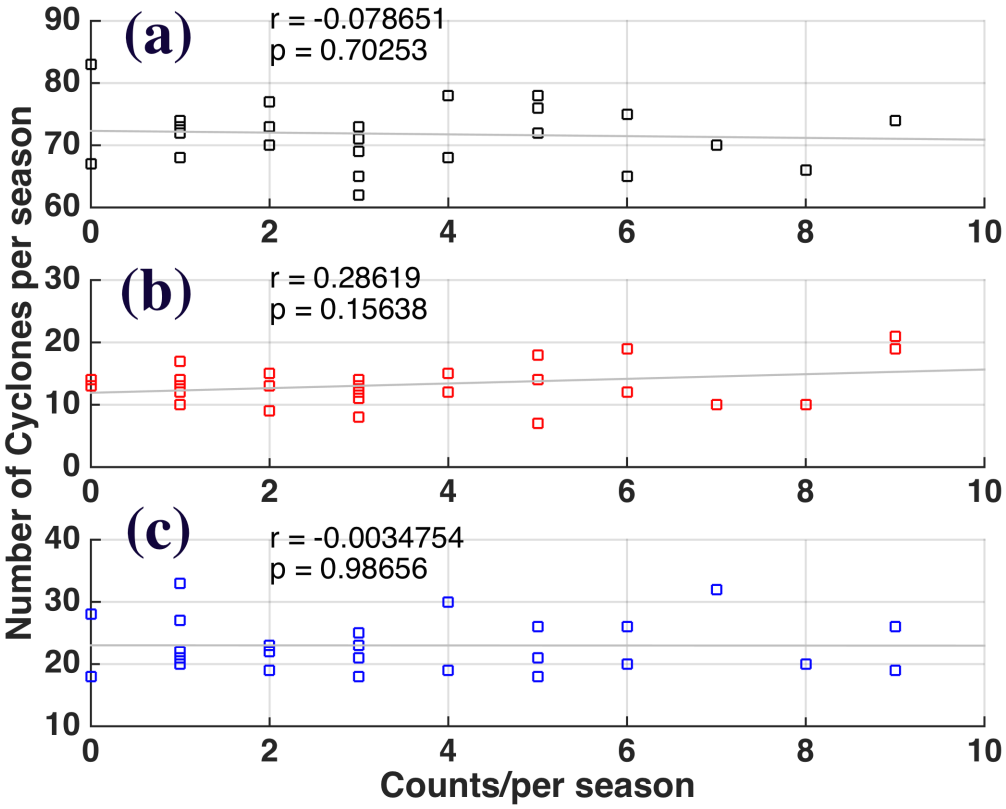


Figure 4.17: Scatter plots of the seasonal CFSR (a) climatological cyclones, (b) Miller Type A, and (c) Miller Type B on the ordinate to the number of observed impactful surge events per season (counts/season) on the abscissa. The correlation and p-value it is significant at is annotated in each panel.

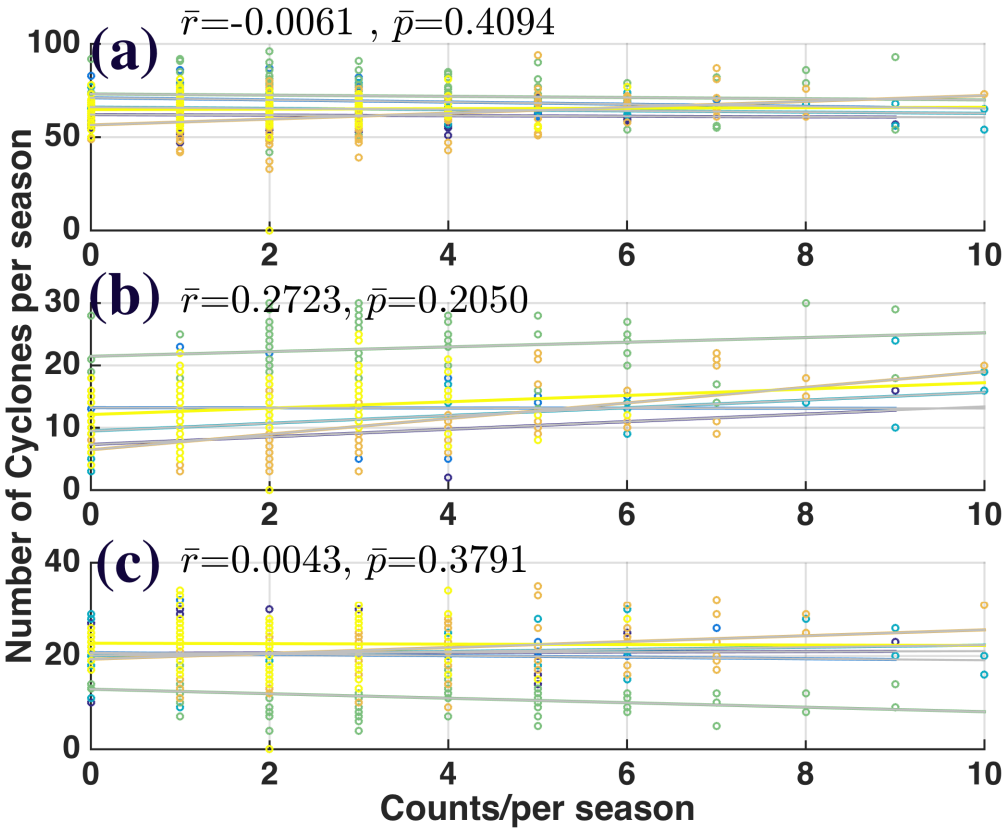


Figure 4.18: Same as 4.17 but for surge counts per season calculated using GCM data between 1979-2079. The average correlation and average p-value it is significant at is annotated in each panel.

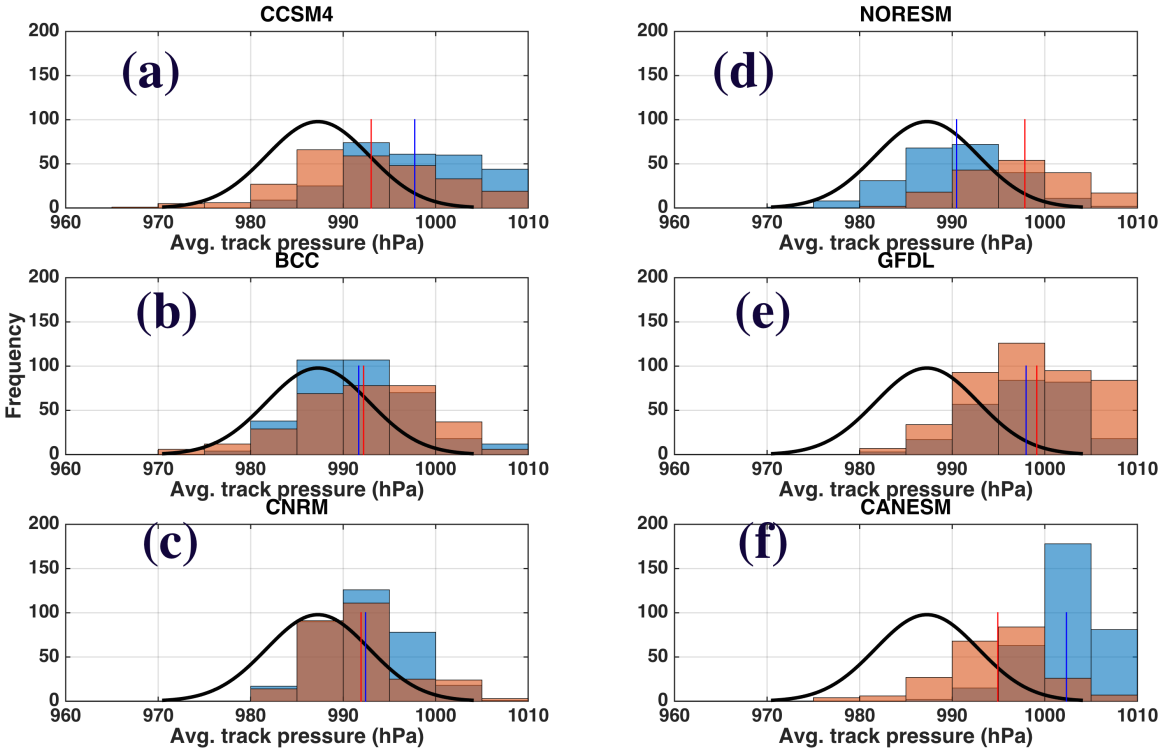


Figure 4.19: Histograms of average along-track mean sea level pressure (MSLP) for Miller Type A tracks during the historical period (1979-2004) and mid.-late 21st century (2054-2079) for six GCMs. The average along-track MSLP for Miller Type A CFSR tracks is drawn as a black line in each panel. In each panel, the blue vertical line is the historical mean and the red vertical line represents the mid.-late future period mean.

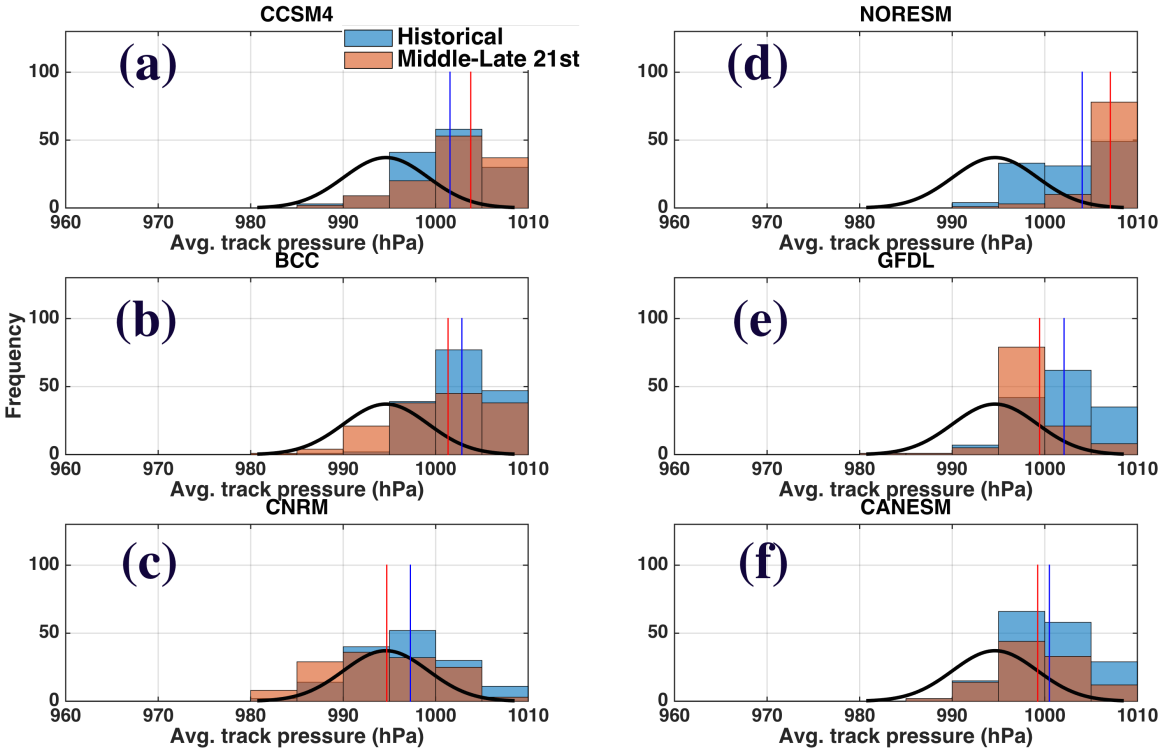


Figure 4.20: Same as figure 4.19, but for the section of Miller Type A tracks within the boxed region over the Mid-Atlantic, which is illustrated as the blue-hatched region in figure 4.12. The average along-track MSLP for the segment of Miller Type A CFSR tracks within the boxed region over the mid-Atlantic is drawn as a black line in each panel. In each panel, the blue vertical line is the historical mean and the red vertical line represents the mid.-late future period mean.

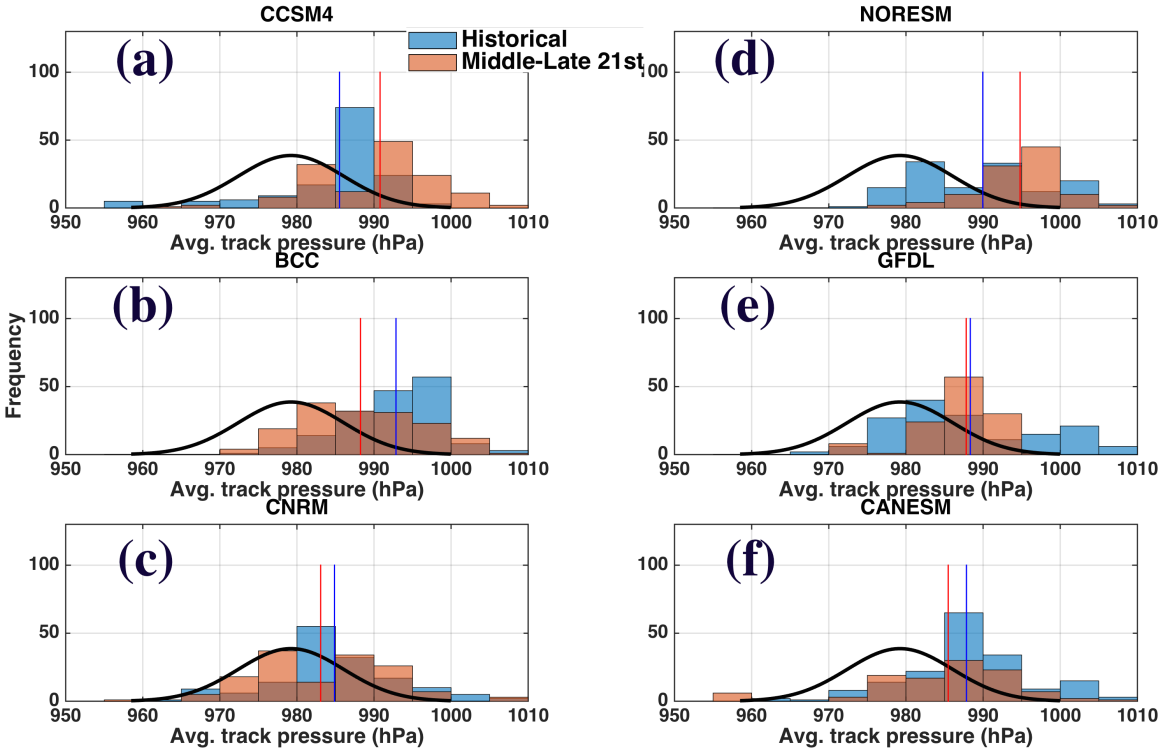


Figure 4.21: Same as figure 4.20 and 4.19, but for the minimum MSLP of Miller Type A track within the boxed region off the mid-Atlantic U.S. The minimum along-track MSLP for the segment of Miller Type A CFSR tracks within the boxed region over the mid-Atlantic is drawn as a black line in each panel. In each panel, the blue vertical line is the historical mean and the red vertical line represents the mid.-late future period mean.

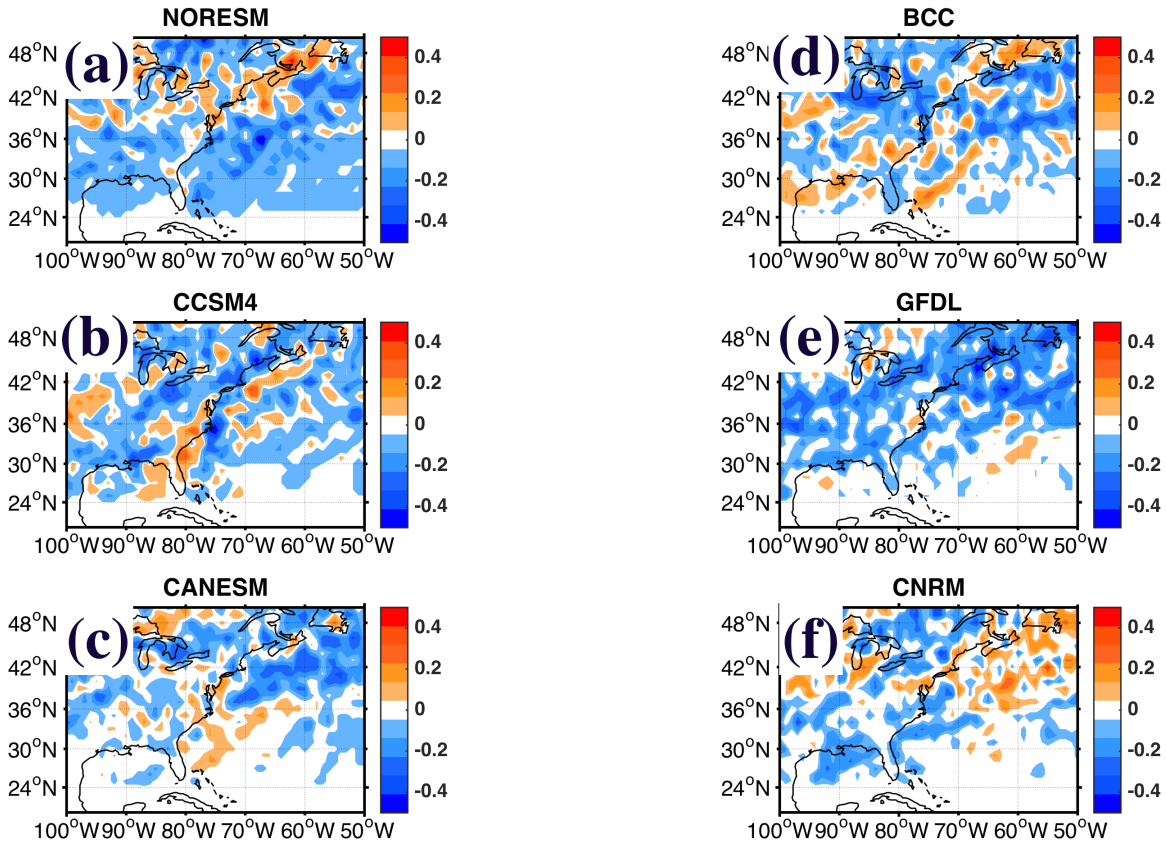


Figure 4.22: (a-f) The difference in the impactful surge-generating cyclone track density per 50,000 km^2 per season between 2054-2079 and 1979-2004 Nov.-March for six GCMs.

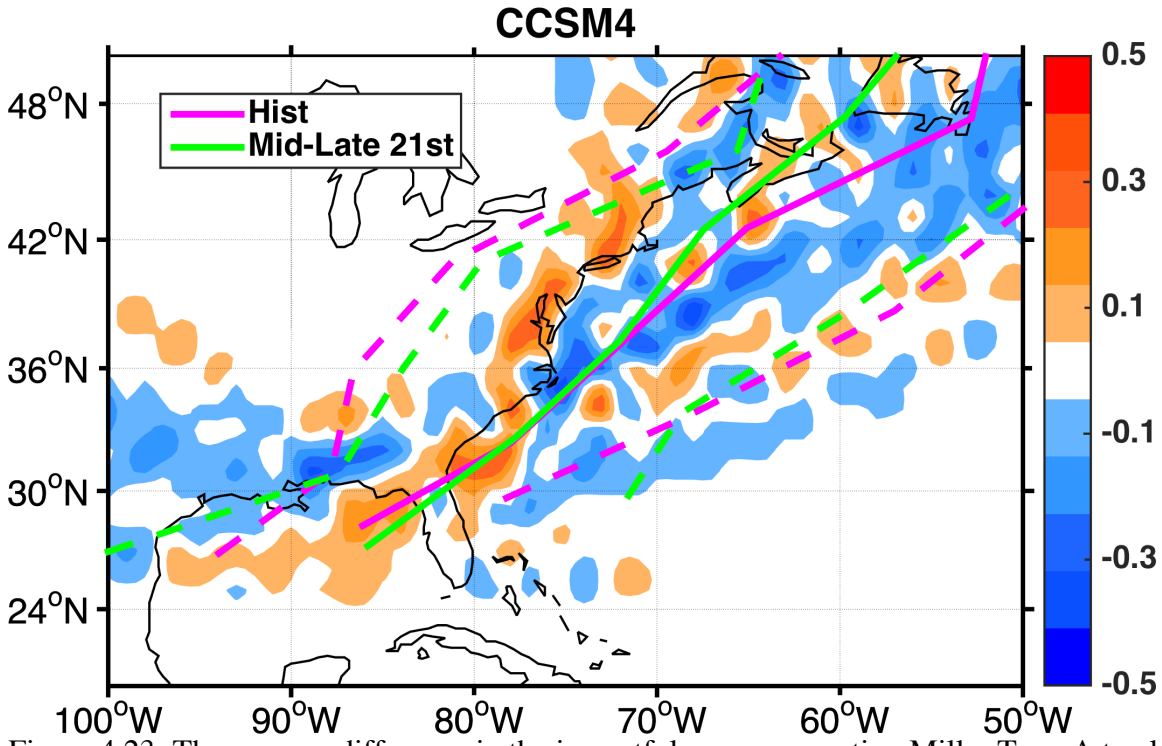


Figure 4.23: The average difference in the impactful surge-generating Miller Type A track density per 50,000 km^2 per season between 2054-2079 and 1979-2004 Nov.-March. The median Miller Type A track for each period is drawn along with the 2-sigma region, which is drawn as a dotted lines.