Stony Brook University

Verification of a Storm Surge Modeling System for the New York City – Long Island Region

A Thesis Presented

By

Thomas Di Liberto

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

August 2009

Stony Brook University

The Graduate School

Thomas Di Liberto

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
Associate Professor
School of Marine and Atmospheric Sciences

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

Dr. Edmund K.M. Chang, Thesis Reader
Associate Professor
School of Marine and Atmospheric Sciences

This thesis is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Thesis

Verification of a Storm Surge Modeling System for the New York City – Long Island Region

by

Thomas Di Liberto

Master of Science

in

Marine and Atmospheric Science

Stony Brook University

2009

Storm surge from tropical cyclones events nor' easters can cause significant flooding problems for the New York City (NYC) – Long Island region. However, there have been few studies evaluating the simulated water levels and storm surge during a landfalling hurricane event over NYC-Long Island as well as verifying real-time storm surge forecasting systems for NYC-Long Island over a cool season.

Hurricane Gloria was simulated using the Weather Research and Forecasting (WRF) V2.1 model, in which different planetary boundary layer (PBL) and microphysics schemes were used to create an ensemble of hurricane landfalls over Long Island. The simulations of the category 1 storm make landfall 20-50 km around the observed and roughly 30-60 min late. The ocean simulations using the Advanced Circulation ocean model (ADCIRC) shows large sensitivity to small changes in track and timing. The control (CTL) simulation, which made landfall 20 km west of the observed, produced a storm surge ~0.60 m higher at the Battery, NY than another WRF run, which has a slightly deeper cyclone but tracks the storm ~30 km east of the observed landfall. If the CTL simulation makes landfall one hour earlier, the peak water level is increased by 0.3 m at the Battery. If the CTL simulation makes landfall 5 hours earlier during high tide, water levels would have peaked 0.8 m above flood level at the Battery. The inclusion of

wave forcing in the simulations helps to reduce the storm surge error prior to landfall by 50 %.

Three real-time forecasting systems (Stony Brook ensemble, Stevens Institute, NOAA-extratropical surge model) are verfied for 75 available days during the 2007-2008 and 2008-2009 cool seasons. Stevens (SIT) has the lowest root mean square error (RMSE), while the NOAA-ET has the largest RMSE later in the forecast (36-48 h) as a result of relatively large negative bias. The Stony Brook storm surge (SBSS) ensemble also has a slight negative surge bias, which is not from a bias in surface wind forcing. The negative bias may be from running ADCIRC in two-dimensions and neglecting wave impacts. The SBSS ensemble is clustered together in terms of mean error (ME) and RMSE, since many of the atmospheric members have similar wind errors on average. The inclusion of an ensemble using all models (8 SBSS members, SIT, NOAA-ET) improves upon the SBSS ensemble mean. A bias correction applied to the ALL ensemble (BC-ALL) helps to reduce the RMSE and make the ME slightly positive over all forecast hours

Probabilistically, the SBSS and ALL ensembles are underdispersed and are made slightly more dispersed using a bias correction. The inclusion of model members using different ocean models improves the performance of the ensembles in terms of Brier score and Brier skill score for positive surge thresholds. A three member ensemble (ENS-3 = SBSS control member 9a, SIT, NOAA-ET) has a larger Brier skill score than the SBSS and ALL ensemble. The BC-ALL and BC-ENS-3 perform better than all ensembles. Reliability diagrams for surge events > 0.3 and > 0.4 m shows that the SBSS and ALL ensembles are underconfident and unreliable for forecast probabilities less than 80%. This work suggests more work is needed to improve these storm surge models

Table of Contents

List of Tables

List of Figures

Acknowledgements

Financial support for this research project was given by New York State Sea Grant.

I would like to thank my advisor, Dr. Brian Colle, for his continued support and help throughout this project. His encouragement during my years at SoMAS was greatly appreciated.

Special thanks also should be given to New York State Sea Grant who always were enthusiastic about my research and very helpful in providing funds for my trip to the 89th annual meeting of American Meteorological Society where I delivered a talk as well as a poster presentation

I also should thank Dr. Frank Buonaiuto, Dr. Malcolm Bowman, Dr. Robert Wilson and Dr. Charles Flagg for listening to my endless questions about physical oceanography and modeling.

A particular thanks goes to Yanluan Lin and Joseph Olsen who answered any and all of my questions about atmospheric modeling that I may have had even after they had graduated and left SoMAS.

A special heartfelt thanks goes out to my girlfriend, Gretchen Goldman, who has had to deal with the brunt of my stressful ways. She has been an anchor, keeping me grounded and focused on the goal at hand.

Lastly, I want to thank my parents, Jim and Marybeth Di Liberto for opening their home to a new college graduate so that he could attend Stony Brook without looking for housing. Their continued selfless support for my passions and endeavors has been superhuman.

Chapter I: Introduction

a. Background

Storm surge is a major hazard for those coastal areas exposed to tropical and/or extratropical cyclones. Storm surge is defined as the rise of seawater above mean tidal level as a result of surface winds around the storm and relatively low surface pressure (AMS Glossary 2nd edit. 2000), with winds having the largest impact on the surge. Hurricane Katrina (2005) produced a 6-9 m (20-30 ft) storm surge along coastal Mississippi and Louisiana, which illustrated the devastation that a major storm surge can create in a coastal metropolitan region. The surge flooded large sections of New Orleans and coastal Mississippi, causing $81 billion worth of damage (Knabb et al. 2005). Another example of a storm surge catastrophe during the cool season is the severe gale force winds that occurred over the North Sea on 31 January – 1 February 1953, which resulted in a storm surge that killed over 2,000 people in the Netherlands, United Kingdom, and Belgium. This led the Netherlands to implement the Delta Plan in 1953 (http://www.deltawerken.com/English/10.html?setlanguage=en), which involved building an intricate series of storm surge barriers throughout this low lying country (Gerritsen 2005, Baxter 2005). In 1970, a typhoon in the Bay of Bengal struck then eastern Pakistan (now Bangladesh) with a 6 m (20 ft) storm surge that killed 300,000 people (Frank and Husain 1971).

There are several other coastal urban areas vulnerable to storm surge. The director of the National Hurricane Center, Max Mayfield, testified before Congress on 20

September 2005, noting five other urban areas where hurricane storm surge could cause massive damage in the United States: Houston/Galveston, Tampa Bay, southern Florida/Florida Keys, New York City- Long Island (NYC-LI) and New England (http://www.legislative.noaa.gov/Testimony/mayfieldfinal092005.pdf). The New York City area, including Long Island (NYC-LI), is vulnerable, since much of this region is less than 3-m (9.8-ft) above sea level. About 260 km^2 of NYC is at risk for storm surge flooding by a 100-yr storm event (Bowman et al. 2005). The only protection surrounding lower Manhattan is a sea wall that is 1.5-2.5 m above mean sea level (MSL) (Colle et al. 2008), which is insufficient to protect much of the business district during major storm surge events.

A common severe storm surge scenario for New York Bight involves a cyclone propagating from south to north along the eastern seaboard (Colle et al. 2009), with winds from the northeast advecting water into western Long Island Sound and the New York Bight (Fig. 1.1). During periods of sustained northeasterly winds over several hours, Ekman transport to the right of the wind can also contribute to the setup of higher water along the south shore of Long Island (Bowman et al. 2005).

During the 11-12 December 1992 nor'easter, 30-40 kt (15-20 m s^{-1}) easterly winds over a 24 h period caused a water level of ~2.6 m (~8.5 feet) above MSL at the Battery in southern Manhattan (Colle et al. 2008), which surmounted the sea wall for nearly two hours. There was flooding 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 (~4 feet) of sea water in lower Manhattan (National Weather Service Disaster Survey Report 1994).

NYC-LI is also vulnerable to storm surge created by tropical storms. The region has a history of tropical cyclone landfalls, such as the hurricanes of 1821 (category 3), 1893 (category 2) near NYC (Scileppi and Donnelly 2007), and the category 3 Long Island Express in 1938 (Brooks 1939). More recently, tropical storm Floyd on 16 September 1999 created a maximum surge of ~1-m in western Long Island Sound (Colle et al. 2008), but luckily the peak surge occurred near low tide.

The most recent hurricane to make landfall on Long Island was hurricane Gloria, which passed between Islip and JFK Airport at 1600 UTC 27 September 1985 (Fig. 3.1), with 75 kt (~37 m s^{-1}) sustained winds and a landfalling central pressure of 961 hPa (Case 1986). The storm surge recorded at the Battery in New York City (~2.0 m) was the highest surge during the last 50 years (Colle et al. 2009). However, since the peak surge occurred at close to low tide, the total water level was not the highest on record at the Battery, NYC (Jarvinen and Gebert 1987), and only minimal flooding was observed around NYC.

b. Hurricane modeling

Hurricanes obtain their energy from surface moisture fluxes over the relatively warm ocean basins, which are released as latent heat within convective storms that surround the cyclone. A realistic simulation of a hurricane requires relatively high spatial resolution (<= 6 km grid spacing) (Davis et al. 2008), proper treatment of the surface fluxes, vertical mixing within the planetary boundary layer (PBL), and parameterization of cloud microphysical processes (Braun and Tao 2000, Fovell and Su, 2007). Liu et al. (1997), through Yau et al. (2004), in a 6 part study on hurricane Andrew (1992), were

able to produce realistic inner core structures of hurricane Andrew (1992) using a triply nested (54 km, 18 km, 6 km horizontal grid spacing) version of the Penn-State-National Center for Atmospheric Research Mesoscale Model (MM5; Grell et al. 1993), with the inner two nests moving along with the hurricane vortex. The inner nest helped to resolve the multi-scale interactions associated with hurricane Andrew. The 6 km horizontal grid spacing was still too coarse to resolve all the convective features in the inner core (Liu et al. 1997 and Yau et al. 2004).

Since the analyses used to initialize these models are typically rather coarse (> 30 km grid spacing), a realistic vortex has to be either spun up or inserted into the higher resolution model. For example, during the 1995 Atlantic hurricane season, Kurihara et al. (1998) ran the GFDL hurricane model down to 1/6° resolution, in which a relatively weak vortex was removed and replaced with a more realistic hurricane vortex. Using this approach, the GFDL model outperformed in track forecast the next best National Weather Service (NWS) hurricane guidance (A90E, a statistical-dynamical model, BAMD, beta advection model (deep layer version)) at forecast hour 36, 48 and 72-h by 14%, 19%, 25%, respectively (Kurihara et al. 1998). In more recent years, the GFDL model has been coupled with a high resolution Princeton Ocean Model (POM) to incorporate ocean response to hurricanes. Hurricane induced cooling of sea surface temperatures through mixing can negatively impact hurricane intensity. In addition, increases to vertical and horizontal resolution have been made (Bender et al. 2007). In 2006, the model outperformed the statistical hurricane intensity prediction system (SHIPS) (DeMaria et al. 2005) beyond 12 h for the first time.

There are some limitations in the technique of inserting a specified vortex, defined by an analytical expression, into the initial conditions for a model. Wang (1998) showed that track and intensity of a bogus vortex can be very sensitive to the vertical structure of the inserted vortex. Also, Xiao et al (2000) note that special care has to be taken to balance the model's initial state from the "initial shock" of inserting the vortex. Hurricanes have also been successfully spun up in an atmospheric model from a relatively weak tropical disturbance and then inserted into the initial conditions. For example, Liu et al (1997) used this spin-up approach to successfully simulate hurricane Andrew's (1992) track, propagation and rapid deepening using 6 km grid spacing. For example, Davis et al. (2008) also used this spin-up technique. Using a triply-nested setup (12-km, 4-km 1.33-km grid spacing), with the two inner domains moving with the vortex, Davis et al. (2008) utilized the Weather Research and Forecasting (WRF; Skamarock et al. 2005) model to investigate the capability of modeling five landfalling hurricanes over the western Atlantic during the 2005 warm season. The high-resolution inner domains allowed for a more realistic simulation of organized convection within the hurricane eyewall, a more rapid deepening of the cyclone, and more realistic rainbands compared to the coarser resolution domains (4 km and 12 km) during hurricane Katrina. With this technique, they found little improvement in track forecasts by using the high resolution nests (4 and 1.33 km nests). Also, they found no systematic improvement in intensity prediction, as judged by maximum sustained winds, with the addition of a 4 km grid. However, the authors indicated that further reduction of grid increment to 1.33-km is beneficial around the storm center as seen by the more rapid deepening of the storm similar to observations.

In the past, the influence of the ocean (two-way coupling of air- sea exchanges) was not taken into account when modeling hurricanes. However, Schade et al. (1999) showed using a simple coupled hurricane ocean model that interacts with the ocean can significantly reduce the intensity of hurricanes through ocean cooling under the storm. Bender et al. 2000 explored this relationship using real events. The authors used the GFDL hurricane model coupled with the high resolution Princeton Ocean Model (POM) for two cases, hurricane Opal (1995) and hurricane Gilbert (1988). Bender et al. found that the tropical cyclones caused a significant cooling of the sea surface which resulted in a substantial decrease in evaporation and boundary layer moist static energy over the cold wake. This had a large effect on storm intensity, especially when the storms moved slowly, reducing the intensity when compared to simulations which neglected the ocean. In each event, the inclusion of the ocean model lead to better intensity forecasts (Bender et al. 2000). More recently, Davis et al. (2008) coupled the advanced hurricane WRF to a simple mixed layer model during hurricane Katrina and found a similar effect, with the net effect of the ocean cooling on the maximum surface winds was a ~8 $m\ s^{-1}$ reduction in wind speed prior to landfall.

In addition to the ocean, hurricane models are also sensitive to planetary boundary layer (PBL) schemes and microphysics schemes (Braun and Tao 2000; Fovell and Su 2007). Braun and Tao (2000) showed that for hurricane Bob (1991), the MM5 had a strong sensitivity of differing PBL schemes for track and intensity. This was caused by differences in the way each PBL scheme treats surface fluxes and vertical mixing. For instance, the authors found that the MRF vertical mixing scheme produces a weaker storm due to an excessively large deep boundary layer which causes drying in the lower

portion of the PBL. However, the Burk-Thompson, Blackadar and bulk aerodynamics schemes produce storms with comparable intensity and structure. Fovell and Su (2007) found a strong sensitivity in hurricane track forecasts of hurricane Rita (2005) to the changing microphysics schemes. The changing of the microphysics schemes lead to changes in certain vortex characteristics such as size, winds at large radius as well as hurricane motion (Fovell and Su 2007).

c. Storm surge physics and models

Storm surge is determined by a number of factors, such as wind speed (surface stress, bottom stress (friction), duration and strength of the surface winds, wind fetch towards the coast, surface pressure, and wave breaking (Weisberg 2008). Equation 1.1 shows the momentum terms for a storm surge in a basin with a uniform bathymetry in the alongshore direction and a shoreline that is much longer than the width of the shelf (Signorini 1992). Also, this equation:

$$\frac{\partial \eta}{\partial x} = \frac{1}{\rho g h}\left(\tau_w + \tau_b\right), \qquad (1.1)$$

where η is the sea level height, x is the distance perpendicular to the shore, h is the depth of the water, τ_w is the surface (wind) stress and τ_b is the bottom stress, assumes a steady state and neglects the inverse barometer effect and the advection of momentum. In finite difference terms, equation 1.1 turns into:

$$\Delta\eta = \frac{K\Delta x \tau_w}{\rho g h} \qquad (1.2)$$

where K accounts for bottom friction when the cross shelf velocity is zero. The pressure gradient is equal to the sum of the two opposing stresses, surface (wind) stress and

bottom stress, divided by the average depth of the water column. Therefore, storm surge will be higher if wind setup occurs over a long distance (larger Δx) there is shallow water (smaller h) (Signorini 1992). Also, given the inverse barometer effect, for every 1 hPa drop in sea-level pressure, the open ocean surface rises .01 m (Harris 1963). Shelf width and depth also play a role in determining the height of the storm surge. A steep slope (large h) or narrow shelf (small Δx) will lead to a smaller surge development.

In addition, changes in friction and bottom stress, e.g., the differences between wetlands and open ocean, can also influence storm surge development. There are varying estimates related to the impact of wetlands on storm surge. A U.S. Army report based on seven storms from 1909 -1957 in southern Louisiana reported storm surge height attenuating 1 m after traveling through 14.5 km of marsh (Resio and Westerink 2008). However, the data had considerable scatter causing a wide range of possible attenuation rates (1 m per 20 km to 1 m per 7 km).

Resio and Westerink (2008) also describe the impact breaking waves have on storm surge. The flux of momentum transfer into the water column from waves (a radiation stress) can add an additional positive influence on water levels near shore (Resio and Westerink, 2008). This rising of the sea level near-shore because of breaking waves occurs due to a balance of momentum in the cross shore direction as explained in equation 1.3 (Komar 1988).

$$\frac{\partial S_{xx}}{\partial x} + \rho g(\overline{\eta} + h)\frac{\partial h}{\partial x} = 0 \qquad (1.3)$$

where S_{xx} is the radiation stress, x is distance perpendicular to the shoreline, ρ is the density of water, g is gravity, $\overline{\eta}$ is the difference between still-water level and the water level in presence of waves and h is the water depth (Komar 1988). S_{xx} is directly

proportional to wave heights (Komar 1988). Therefore, when waves are close to the shore and wave heights are increasing, $\frac{\partial S_{xx}}{\partial x}$ becomes > 0. In order to maintain balance, $\overline{\eta}$ must decrease. This is referred to as wave set-down (Komar 1988). However, when the waves break and their height decreases, $\frac{\partial S_{xx}}{\partial x} < 0$ and $\overline{\eta}$ must increase to maintain a momentum balance and sea levels rise. This is referred to as wave set-up and is the positive wave contribution to storm surge along the coast (Komar 1988). Thus, in order to properly address storm surge, waves should also be included as a forcing.

The National Weather Service (NWS) utilizes the Sea, Lake and Overland Surges (SLOSH) model developed by the Techniques Development Laboratory of the NWS (Jelesnianski et al. 1992) for its hurricane surge predictions. SLOSH ingests surface pressure, size of the storm, surface winds, cyclone track and forward speed to estimate storm surge heights for one of 40 SLOSH basins along the U.S. East Coast, including the New York City metropolitan region. SLOSH calculates the wind field assuming a symmetric wind field around the storm and using the observed storm radius, observed central and peripheral sea level pressures, and storm forward speed. SLOSH computes water height using Newtonian equations of motion (shallow water equations) without the advective terms, and the continuity equation integrated from the sea floor to the surface (Jelesnianski et al. 1992). SLOSH does not include wave run-up, wave setup, and rainfall flooding.

Other ocean models have also been used to simulate storm surge events for tropical cyclones. For example, Shen et al. (2005) used the Advanced three dimensional Circulation Model for Coastal Ocean Hydrodynamics (ADCIRC; Westerink et al. 1993)

to hindcast the surge for Hurricane Isabel (2003) for Chesapeake Bay. Utilizing a simplified wind model similar to SLOSH to calculate the atmospheric forcing for ADCIRC, the authors found that simulated surge (1.9-2.5 m) was within 0.3 m of observed, with a RMS error of ±0.19 m. Westerink et al. (2007) used ADCIRC to successfully hindcast the surges associated with hurricanes Betsy (1965) and Andrew (1992) in Southern Louisiana, and the model was within 10% at most gauge stations. For tropical storm Floyd (1999) impacting NYC-LI, Colle et al. (2008) used ADCIRC and the MM5 at 12 km grid spacing to realistically simulate the surge to within 0.1 m of observed. Weisberg et al. (2006) used the Finite Volume Coastal Ocean Model (FVCOM) to gauge the sensitivity of storm surge to track in the Tampa Bay area, and found that surge is sensitive to storm approach, and storm speed with slower moving storms having twice the storm surge of faster moving cyclones.

Weisberg et al. (2008) investigated the impacts of running FVCOM in a two-dimensional (barotropic) mode versus three-dimensional for storm surge predictions for Tampa Bay. The 2D model uses vertically-averaged velocities within the water column to calculate bottom stress, while 3D models are vertically independent and include a bottom boundary layer velocity. Weisberg et al. (2008) found that this leads to larger bottom velocities and larger bottom stresses in 2D models compared to 3D models. They showed that this overestimation of bottom stress in a 2D model can lead to an underestimation of surge heights.

Storm surge from non-tropical (extratropical) cyclones generates different challenges. The assumed symmetric wind field in SLOSH is less accurate, given the large asymmetries in the surface wind and pressure fields for cool season mid-latitude cyclones

and for tropical storms that become more extratropical at higher latitudes (Atallah and Bosart 2003; Colle 2003). As a result, the NWS developed the Extratropical (NOAA-ET) Storm Surge model during the mid-1990s (Shaffer et al. 1997, Blier et al. 1997). The NOAA-ET model is a vertically-averaged barotropic ocean model that includes a specified bottom friction (Tilburg and Garvine 2004), similar to the SLOSH model. The NOAA-ET surge model is forced with the surface wind and pressure fields from the NCEP AVN (currently Global Forecast System) model. Also, the NOAA-ET model is run on a large grid that covers much of the western Atlantic and for a few days given that extratropical surge events can persist at a location for more than 24 h (Blier et al. 1997).

Tilburg and Garvine (2004) verified the NOAA-ET model and a simple linear regression model for surge forecasts at Atlantic City, NJ from 1997-1998. They found that the NOAA-ET model explained 79% of the total observed subtidal frequency of water level as compared to the 74% in the regression model. Blier et al. (1997) used the NOAA-ET model along with winds and sea level pressure from the AVN analyses to hindcast storm surges along the south coast of Alaska. They found that the model realistically predicted the surge in an event 6 October 1992 to within 0.8 m, although for a short duration surge event (20 August 1993) the model did not have a peak surge and under-predicted the observed surge by ~1.0 m. This deficiency may have been from the low temporal and spatial resolution of the AVN wind fields, which could not resolve the coastally-enhanced winds along the Alaskan coast. Colle et al. (2008) successfully simulated the surge from the December 1992 nor-easter at Battery, NY using ADCIRC forced with 12-km mesoscale model (MM5) winds to within 5-10% of observed.

Over the course of several years, Blumberg et al. (1999) used the Princeton Ocean Model (POM) to forecast storm surge for the New York Bight. They found that the errors in water level were less than 10% of the local tidal range and correlations between the data and model exceeded 0.95 from Oct. 1994 – Sept. 1995.

d. Motivation

There have been several studies applying high resolution atmospheric models to hurricanes along the U.S. Gulf Coast and off the southeast U.S. coast (Davis et al. 2008; Liu et al 1997; Yau et al. 2004; Zhu et al. 2006). However, there have been no formal studies of simulated hurricanes over the Northeast U.S. Colle (2003) simulated the landfall of tropical storm Floyd (1999) over LI, but this storm produced only a ~1.0 m surge. Although hurricanes are less frequent over the Northeast U.S. than to the south, it is still important to simulate past landfalling hurricane events over the Northeast using state of the art technology in order to better understand the model capabilities.

There have been a few case studies of simulated storm surges caused by hurricanes along the mid-Atlantic and U.S. southeast coast (Shen 2005; Weisberg 2006, 2008; Westerink 2007), but there has been no formal modeling case studies of a hurricane storm surge over the Northeast U.S. Also, given the large population in the NYC-LI area, evacuation is paramount for a pending extreme hurricane event, especially in the flood prone areas. It is not clear how a small change in track and timing of the storm can impact the coastal water levels around NYC-LI, thus additional numerical experiments are needed.

This thesis focuses on hurricane Gloria (1985), which made landfall on Long Island as a category 1 storm on 27 September 1985. Gloria was a classic Cape Verde hurricane, originating as a tropical easterly wave on 15 September 1985 becoming a hurricane six days later at 1800 UTC 21 September. Over the warm waters of the tropical Atlantic, hurricane Gloria strengthened to a category four hurricane with a central pressure of 919 hPa at 0120 UTC 25 September (Case 1986). Hurricane Gloria began to turn northwest on 26 September, and to the northeast by 27 September (Fig. 3.1). At 0600 UTC 27 September, Gloria made its initial landfall over the Outer Banks, NC. Ten hours later, Gloria made landfall as a category 1 hurricane (961 hPa) at 1600 UTC near Islip, NY (Landsea, personal communication, 2008). Overall, Gloria killed 8 people and caused damage estimated at $900 million (Case 1986). There was widespread tree damage which caused power outages across New England but since the storm made landfall near low tide, coastal flooding throughout the northeast was minimal (Case 1986). Gloria has been a well researched storm from the ocean response (SST cooling) (Brink 1989), potential vorticity development (Shapiro and Franklin 1995), kinematic structure (Franklin et al. 1993) dropwindsonde and radar observations (Franklin et al. 1988) and modeling studies by Kurihara et al. 1990. However, there has not been a study modeling the storm surges across the northeast caused by hurricane Gloria.

There has been limited evaluation of operational storm surge models for southern New England and New York. Although Colle et al. (2008) used ADCIRC/MM5 to realistically simulate the water levels associated with tropical storm Floyd (1999) and the 1992 nor'easter around NYC-LI, ADCIRC needs to be verified for more events. With the relatively large NYC-LI population and a complex coastal geometry prone to storm surge

flooding, it is important to investigate the skill of storm surge models over a longer verification period than just a few surge events. There are currently several modeling systems from various institutions (NWS, Stevens Institute of Technology, and Stony Brook University) that forecast storm surge in real-time system for the NYC-LI region using different models; however, there has not been any formal verification or inter-model comparison of these models over the Northeast U.S.. Also, the Stony Brook system is run in an ensemble (8-member) configuration, so the benefits of this need to be quantified relative to using any single model as well as using all three (multi-model) operational models together. It is hypothesized that an ensemble of storm surge runs created by varying atmospheric forcing should perform better on average than a single atmospheric forcing and a selected ocean model.

This research will answer the following questions:

1. How well can the atmospheric model simulate a historical landfalling hurricane over the Northeast U.S.?
2. How accurate is ADCIRC in simulating storm surges for a significant hurricane event over the NYC-LI region using high-resolution wind and pressures from an atmospheric model?
3. How sensitive are the ADCIRC water levels to relatively small changes in hurricane track, strength, and timing?
4. What is the potential contribution from model surge errors in the atmospheric forcing as compared to uncertainties within ADCIRC?
5. How well does the ADCIRC/MM5-WRF storm surge model perform over a full cool season around NYC-LI, which includes several extratropical cyclone events?

6. What is benefit of an ensemble storm surge prediction and does the ensemble have some probabilistic skill?

Chapter 2 describes the data and methods used in the hurricane Gloria simulations using WRF and ADCIRC. Chapter 3 presents the observational and modeling analysis of Gloria, including some sensitivities of the predicted surge to different WRF runs and timing. The long term verification of ADCIRC as well as the NOAA and Stevens' surge forecast systems is presented in Chapter 4. Summary and conclusions are presented in chapter 5.

Chapter II: Data and Methods

a. Hurricane Gloria (1985) modeling

The North American Regional Reanalysis (NARR) (Mesinger et al. 2006), with a 32-km grid spacing, was utilized for initial and boundary conditions for a Weather Research and Forecasting (WRF) model hindcast of hurricane Gloria (1985). WRF is a terrain-following sigma-height coordinate model that solves compressible, nonhydrostatic Euler equations to predict regional circulations and precipitation (Skamarock et al. 2005). The WRF was run at 36-, 12- and 4-km grid spacing using a vortex following algorithm on the 4 km grid (Tenerelli and Chen 2001) (Fig. 2.1). Although it has been found that < 4-km grid spacing can improve the realism of the inner hurricane vortex (Davis et al. 2008), there were only small improvements in the Gloria wind field from 12-km to 4-km grid spacing in that study, so a 1.33 km grid run was not attempted. The moveable 4-km WRF grid is setup to follow the center of the 500 hPa circulation every 15 minutes (Tenerelli and Chen 2001). The Gloria simulation is initialized at 0000 UTC 26 September 1985 (41 hours before landfall occurred over Long Island), and is run to 2100 UTC 28 September 1985. The NARR sea surface temperatures (SSTs) did not have the wake cooling via mixing from Gloria, so multiple SSTs analyses were not used in WRF. Given the rapid movement of Gloria up the East coast, the impact of upwelling under the storm was likely small as well (Dickey et al 1998).

In order to realistically simulate Hurricane Gloria, the storm was inserted into the North American Regional Reanalysis (NARR) using the “bogusing” approach, originally

developed by Davis and Low-Nam (2001) and successfully applied to typhoon simulations in Chou and Wu (2008), Wu and Yen (2002) and Farfan and Zehnder (2001). This scheme replaces the poorly-defined cyclone in the NARR with a realistic vortex given some of the observed winds structure. First, the NARR vortex was removed for a 500-km radius away from the center of the storm, defined as the maximum vorticity near the surface in the NARR grid. Next, a rankine vortex is inserted at the observed location using the observed maximum wind speed and radius of maximum wind as input parameters. The horizontal wind profile is described as (Davis and Low-Nam 2001):

$$v(r) = A(z) * V(r), \tag{2.1}$$

where,

$$V(r) = v_m * \frac{r}{r_m} \; ; (r \leq r_m) \tag{2.2}$$

or,

$$V(r) = v_m * \left(\frac{r}{r_m} \right)^{\alpha} \; ; (r > r_m), \tag{2.3}$$

and v_m is the maximum tangential wind at the radius of maximum wind (r_m). V(r) represents the wind speed as a function of radius of the storm (r), where α changes the shape of the wind profile beyond the radius of maximum wind. The A(z) contains the amplitude and height wind dependence as a function of height. As in Davis and Low-Nam (2001), this A(z) term was set to one from the surface up to 850 hPa, 0.95 at 700 hPa, 0.9 at 500 hPa, 0.7 at 300 hPa, 0.6 at 200 hPa and 0.1 at 100 hPa. Lastly, v(r) represents the wind speed structure of the new vortex over all levels out to a prescribed radius of 500 km. The mass and wind field around the storm are balanced using non-linear balance (Davis and Low-Nam 2001).

Special consideration was taken to accurately insert hurricane Gloria into WRF at 0000 UTC 26 September 1985. When the α default parameter was used as in Davis and

Low-Nam (2001), (α = -0.75, RMW = 50 km, and radius of storm = 500 km), the winds 600 km to the west of the storm center over the Florida peninsula were four times (~40 kts) as strong as those observed (~10 kts) at initialization. Also, the flight level winds from the NOAA P-3 aircraft were overestimated by ~20 kts about 600 km southwest from the storm at around 4-h into the simulation. In order to better represent the observed winds around the storm, the α in eq.2.3 was changed to -1.2, which created initial surface winds within 5 kts of observed over Florida. Flight level winds 4-h into the simulation were also within 5 kts near the coast of Florida. Thus, an alpha parameter of -1.2 was therefore used for all of the simulations. It should be noted that this approach does constrain the central pressure of the storm during the initialization. Rather, the sea level pressure is derived after the new cyclone wind field is constructed.

The NARR at 0000 UTC 26 September also lacked enough clouds and moisture to fully develop into a hurricane during the first 24-h. In order to improve the Gloria cloud representation, the relative humidity was set to 95% from the surface to 200 hPa for the region where there were surface winds exceeding tropical storm force (33 kts).

The initialization time of 0000 UTC 26 September 1985 is chosen due to factors relating to the spatial extent of the NARR, strength of hurricane Gloria as well as the models ability to have hurricane Gloria make landfall across Long Island. Any earlier than the 25 September 1985 and hurricane Gloria would have been outside the domain of the NARR fields. In addition, hurricane Gloria reached its maximum intensity early on the 25th and weakened throughout the day. Inserting a vortex at Gloria's maximum intensity causes a large initial shock to the simulation. Simulations starting at 0000 UTC 25 September also fail to have hurricane Gloria make landfall across Long Island, which

is the goal of the simulations to test the surge model. The latest time period in which to simulate Gloria and have the storm make landfall across Long Island is the initialization time chosen.

Braun and Tao (2000) showed that hurricane simulations are sensitive to surface fluxes and vertical mixing parameterizations within the model planetary boundary layer (PBL) schemes. Fovell and Su (2007) showed that simulated tracks of hurricane Rita (2005) were also sensitive to varying bulk microphysical schemes in WRF. Thus, for Gloria (1985), a variety of WRF runs were completed using different planetary boundary layer (PBL) and microphysical schemes in order to test the surge model sensitivities to a spread of hurricane landfalls across Long Island. The PBL schemes used include the Yonsei University (YSU) (Skamarock et al. 2005 and Hong et al. 2006), GFS (Janjic et al. 2008), and Mellor-Yamada-Janjic (MYJ) (Janjic 2002) PBL schemes. The GFS scheme is a slightly modified version of the Medium Range Forecast (MRF) PBL scheme (Hong and Pan 1996). Both the YSU and GFS schemes are non-local mixing schemes. The GFS scheme is a first-order vertical diffusion scheme of Troen and Mahrt (1986) using similar counter-gradient flux terms as the YSU scheme. The PBL height is determined from an iterative bulk-Richardson number approach working from the ground up and using counter-gradient terms to represent fluxes in unstable conditions. The YSU scheme treats the entrainment layer at the PBL top using an additional entrainment flux term not existing in the original MRF PBL scheme. The YSU scheme uses a critical bulk-Richardson number of 0 is used to determine the PBL top, while the GFS uses 0.5. The MYJ scheme uses the Mellor-Yamada Level 2.5 turbulence closure model (Mellor and Yamada, 1982) for all atmospheric turbulent regimes.

Two different microphysical parameterizations were used in the WRF Gloria simulations. The Purdue Lin (Lin) microphysics scheme (Chen and Sun 2002) is based on Lin et al. (1983) and Rutledge and Hobbs (1983), with some modifications such as a saturation adjustment following Tao et al. (1989) and ice sedimentation. The Lin scheme has 6 classes of hydrometeors, including water vapor, cloud water, rain, cloud ice, snow and graupel. The Ferrier scheme is a double moment scheme that has four classes of hydrometeors: suspended cloud liquid water, rain, precipitation ice (snow, graupel, and sleet) and cloud ice (Ferrier 1994 and Ferrier et al. 2002). Both microphysical schemes include mixed phased processes. Meanwhile, the Kain-Fritsch (Kain 2002) cumulus scheme was used in the outer 36- and 12-km domains.

The approach of changing PBL and microphysics schemes instead of slightly changing the vortex insertion parameters to create an ensemble of landfalls was used for a couple reasons. The physics approach allows for the same initialization of the vortex inserted Gloria for all simulations. This way the most accurate vortex inserted Gloria is used for all runs. In addition, this approach allows us to look into some of the sensitivities that PBL and microphysics schemes have on track and intensity predictions similar to the work of Braun et al. (2000) and Fovell and Su (2007).

In order to determine how well the model simulated hurricane Gloria, observations from various sources were used as a source for comparison. In addition to surface observations of wind direction, temperature and sea level pressure across the United States for the duration of the cyclone, various aircraft reconnaissance flights were made through hurricane Gloria. Two WP-3D (P3) aircrafts operated by the NOAA for airborne hurricane research were flown through Gloria during 26 and 27 September 1985.

The aircraft measured pressure, wind speed, wind direction, temperature, dew point, surface pressure and potential temperature every second for the duration of the flight. Wind speed and direction from several passes through the eye of Gloria will be compared with model output at similar times and altitude. In addition, the P3 aircraft gathered radar images through radars attached to the lower fuselage..

b. Ocean and wave models

The Advanced three dimensional Circulation Model for Coastal Ocean Hydrodynamics (ADCIRC) solves time-dependent, free-surface circulation and wind-driven transport problems in a barotropic configuration on a finite element grid (Westerink et al. 2007). Hourly 10-m winds from the 12- and 4-km WRF grids starting at 0000 UTC 26 September were used to calculate a wind stress for ADCIRC using Garratt (1977)

$$\begin{aligned} \tau_{sx} / \rho_w &= W_d \times 0.001293 \times u \times U, \\ \tau_{sy} / \rho_w &= W_d \times 0.001293 \times v \times U, \end{aligned} \qquad (2.4)$$

where τ_{sx} (τ_{sy}) is the u (v) component of the wind stress, ρ_w is the density of water, U is the wind speed, u and v are the wind components. The 0.001293 is calculated from extrapolation of observations (Garratt 1977). The wind drag (W_d) is calculated as follows

$$W_d = 0.001 \times (0.75 \times 0.067 \times U), \qquad (2.5)$$

where $W_d = 0.003$, if W_d is calculated to be > 0.003. The constants were derived through observational studies.

The ADCIRC model simulates the water elevation and currents for the domain along the U.S. East Coast to offshore over the continental shelf (Fig. 2.2). Its grid has

triangular elements (~108,000 nodes) ranging from 70 km offshore to about 10 m around parts of Long Island and NYC (Fig. 2.2b), with the highest resolution around the coast, specifically New York Harbor. ADCIRC is spun up one day before the full WRF forcing is applied using tides forced along its open boundary. The M2, K1, O1, N2 and S2 tidal constituents obtained from an ADCIRC global model are applied along the boundary during the entire run as in Westerink et al (1993). In addition, wind and pressure forcing is linearly increased throughout the ramp day to reduce the impact of suddenly introducing new forcing into the ocean model.

Since the 4-km WRF domain is moving and not always within the full ADCIRC domain, any portion of the ADCIRC domain not inside the 4-km WRF domain used the wind stress and surface pressures from the 12-km WRF domain. In other words, once any part of the 4-km domain entered the ADCIRC grid, the 4-km wind stresses were used in that portion of the ADCIRC grid.

For selected ADCIRC runs, wave forcing (radiation stresses) was included in the ADCIRC model using the Simulating Waves Near-shore (SWAN) wave model (Booij et al. 1999). 15-minute 10-m winds (from 0600 UTC 27 September – 0600 UTC 28 September) and 1-h 10-m winds (0000 UTC 26 September – 0600 UTC 27 September) from the WRF were used to calculate U_* (frictional velocity) and the drag coefficient, as specified in Wu (1982), in order to force SWAN. SWAN estimates wave period and height based on wind speed, refraction and dissipation due to bottom friction, wave-breaking and non-linear interactions (Booij et al 1999). During the last 10 hours before landfall, the wave field is developed over a 15-minute period using the wind conditions for those 15 minutes. Each subsequent 15-minute run is “hot started” with the previous

15-minutes sea state to more accurately spin-up the wave field. This approach has been used in other studies (Ris et al. 1999)

Xu et al. (2007) used SWAN to accurately predict waves during Hurricane Juan (2003) across the Lunenburg Bay region of Nova Scotia. They found that SWAN tended to over-predict wave height to the right of the cyclone track due to too high surface drag coefficients. Recent studies (Donelan et al. 2004) have suggested that an upper bound of 0.003 should exist for drag coefficients in hurricane conditions, while in the hurricane Juan simulations, drag coefficients maxed out at 0.0044 on the right side of the storm. Wornom et al. also used SWAN along with WAM (Wave model; Hasslemann et al. 1988) to simulate waves along the North Carolina and Virginia coasts during hurricane Luis (1995). SWAN was able to accurately predict the peak wave heights along the coast as well as at buoys offshore.

The observed water and storm surge levels at the Battery in lower Manhattan, New York from the National Ocean Services of the National Ocean and Atmospheric Administration (NOAA) (http://tidesandcurrents.noaa.gov) were used to verify ADCIRC. Additional surge data along the south shore of Long Island was obtained by from the town of Hempstead on Long Island, NY. Unfortunately, there was no wave measurements to valid the waves around NYC-LI, but there was station CHLV2, 26 km east of Virginia Beach, VA.

c. Cool season storm surge model verification

ADCIRC and two other operational modeling systems were evaluated over one and a half cool-seasons (November 2007 – March 2008, October 2008 – December 2008)

around NYC-LI. Operational storm surge forecasts initialized each day at 0000 UTC were obtained from three modeling systems: Stony Brook Storm Surge (SBSS) model, Stevens Institute of Technology (SIT) model and NOAA E-T Surge model. Figure 4.1 shows the stations around NYC-LI used in the verification analysis. Table 2.1 lists the important configurations for the three modeling systems.

The Stony Brook Storm Surge (SBSS) system uses an ensemble approach by running 8 separate ADCIRC members to hour 48 with the surface wind and pressure from either the WRF or the MM5 model. The MM5 and WRF models are run using three domains (36, 12 and 4 km) (Fig.2.3), with the wind and pressures for ADCIRC originating from the 12 km WRF/MM5 grid. The members are "hot started" each day meaning the previous days forecast is used as the initial forcing for the current forecast. To avoid model errors continually being used as initial conditions, the ensemble is cold started, with a one day spin up period, every two weeks. The five MM5 members and 3 WRF members used in the ensemble are run using varying initial conditions and model physics listed in Table 2.2. Four different initial conditions are used (WRF-NMM, NOGAPS, GFS, and Canadian model), three different microphysics [Simple Ice (Dudhia 1989), Reisner (Reisner et al. 1998), Ferrier (Ferrier et al. 2002), and WSM3 (Hong et al. 2004)], four different planetary boundary layer schemes [MRF (Hong and Pan 1996), Mellor Yamada (Janjic 2002), Blackadar (Zhang and Anthes 1982) and YSU (Hong et al. 2006)], two radiation packages [Cloud Radiation, CCM2 (Hack et al. 1993) and RRTM (Mlawer et al. 1997)] and three different convective parameterizations [Grell (Grell et al. 1994), Betts Miller (Betts and Miller 1993) and Kain Fritsch (Kain 2002)] schemes. Member 1 (Table 2.2) is labeled as the control member for the SBSS ensemble, since it

uses a MM5 member that has been run for several years over the Northeast U.S. (Colle et al. 2003; Jones et al. 2007). The operational ADCIRC grid used the same grid as in the hurricane Gloria simulations (Fig. 2.2).

The Steven's Institute of Technology's (SIT) storm surge modeling system uses a modified version of the Princeton Ocean Model (POM) called the Estuarine, Coastal and Ocean Model (ECOM). The ECOM is a three dimensional, time dependent model that solves equations regarding the conservation of mass, momentum, heat and salt (Blumberg et al 1999). The model is driven by surface wind, sea level pressure, surface heat flux, freshwater heat fluxes as well as 268 point sources of sewer overflows and storm runoffs. The hourly NCEP 12-km resolution NAM is used for the wind and pressure forcing. The domain extends from southern New Jersey north to Cape Code, MA, including the Hudson River up until the Troy Dam. (Fig.2.3b) (Blumberg et al. 1999). The resolution of the grid runs from 500-m inside the rivers to 42-km farther offshore. The M2, S2, N2, K2, O1, P1, and Q1 tidal constituents are forced along the boundary along with salinity, temperature, wind stress, heat flux and freshwater inflow (Blumberg et al 1999).

The SIT ECOM system is run in a three-dimensional configuration, with 10 sigma levels scaled to the local water column depth. In forecast mode, the model performs a 72-h simulation. This includes a 24 hour hindcast of the previous day using NAM analysis and short-term forecasts available every 6-h and a 48-h forecast of salinity, temperature, currents, and water level. (Bruno et al. 2006).

The NOAA Extratropical Storm Surge Model (NOAA-ET) (Fig. 2.3c) is based on the same quasi-linear depth-integrated shallow water equations as the Sea, Lakes and Overland Surges from Hurricanes (SLOSH) model (Shaffer et al. 1997; Jelesnianski et al

1999). However, the NOAA E-T model does not use a parametric wind model as in SLOSH, but rather the NCEP GFS model for hourly wind forcing. Figure 2.3b shows the domains for the NOAA-ET model, which is run in a 2-D barotropic mode to save computation resources. Unlike the SBSS and SIT modeling systems, the NOAA-ET model only predicts storm surge and does not include tides (Shaffer et al. 1997). The NOAA-ET model uses a reference sea level to determine the storm surge produced with the surface wind and pressure forcing. The North American Vertical Datum of 1988 (NAVD-88; Zilkoski et al 1992) is used as the reference sea level (0 m). The storm surge levels are then added to the NOS predicted tides to obtain the total water levels.

Observed storm surge levels were determined by subtracting the observed astronomical tide from the total (measured) water levels at each station. The tide was derived by using the T-tide program (Pawlowicz et al. 2002) which performs a harmonic analysis on the observed water level; in this case the Nov. 2007 – March 2008 / Oct. 2008 – Dec. 2008 NOAA observed water levels at each individual station around NYC-LI. The derived tides were used, since the astronomical tides provided by NOAA are referenced to the 1983-2001 datum, which is several years old and therefore can be susceptible to underestimating water levels given slowly rising sea level (~3 mm yr^{-1}). Several error metrics were used to determine the accuracy of each modeling system during the cool seasons (Nov-Mar 2007 and Oct-Dec 2008). The root mean square error (RMSE in eq.2.6), and mean error (ME in eq.2.7) were used in the deterministic verification.

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(y_k - o_k\right)^2} \qquad (2.6)$$

$$ME = \frac{1}{n}\sum_{k=1}^{n}\left(y_k - o_k\right), \qquad (2.7)$$

where y_k is the forecasted value, o_k is the observed value at the same time, and n is the total number of values.

In order to determine how well the SBSS ensemble system and all models together performed probabilistically, the Brier Score (BS in eq.2.8), and Brier Skill Score (BSS in eq. 2.9) were calculated.

$$BS = \frac{1}{n}\sum_{k=1}^{n}(t_k - w_k)^2 \qquad (2.8)$$

$$BSS = 1 - \left(\frac{BS}{BS_{ref}}\right) \qquad (2.9)$$

where t_k refers to the probability forecast and w_k refers to whether the event occurred (w_k = 1) or not (w_k = 0). In essence, the Brier score is the mean squared error of the probability forecasts (Wilks 2006). Furthermore, the Brier score was separated into reliability (REL in eq.2.10), resolution (RES in eq.2.11) and uncertainty (UNC in eq.2.12) where BS = REL – RES + UNC. Reliability evaluates the conditional bias of the forecasts and is a weighted average of the squared differences between the forecast probabilities and the relative frequencies of the forecast event in each subsample (Wilks 2006). For reliable forecasts, the observed relative frequency is equal to the forecast probability in each subsample. Resolution looks at the ability of the forecasts to tell subsample forecast periods with different relative frequencies of the event. Statistically, resolution is the weighted average of the squared differences between the observed subsample relative frequencies and the overall climatologically relative frequency. Therefore, for an ensemble with good resolution, the forecasts will sort the observations into subsamples having different relative frequencies than the overall sample climatology (Wilks 2006). Uncertainty depends only on the observations and cannot be changed by

the forecast. If the event occurs very often or very rarely, the uncertainty is small. In essence, the uncertainty is equal to the brier score of the climatology (Wilks 2006).

$$REL = \frac{1}{n}\sum_{i=1}^{I} N_i\left(d_i - \overline{f_i}\right)^2 \qquad (2.10)$$

$$RES = \frac{1}{n}\sum_{i=1}^{I} N_i\left(\overline{f}_i - \overline{f}\right)^2 \qquad (2.11)$$

$$UNC = \overline{f}\left(1 - \overline{f}\right) \qquad (2.12)$$

where N_i refers to the total number of times each forecast value is used, d_i is each possible forecast probability, $\overline{f_i}$ is the observed relative frequency of occurrence given a possible forecast probability, and $\overline{f}$ is the overall sample frequency of the event or the sample climatology.

Rank histograms (Talagrand diagrams) were also calculated for several stations across the New York metropolitan region to evaluate the dispersion of the ensemble members (Wilks 2006). Ideally, an observation should have an equal chance of falling above, below, or between any ensemble member. This is done by sorting the ensemble predictions at each forecasted time from lowest to highest and creating bins that represent values between each ensemble member as well as bins that represent values higher and lower than all members. Every time an observation falls into one of these bins a rank of 1 is added to that bin. After every observation is sorted, a bar graph of the sum of the ranks is made which will show how well dispersed the observations are within the ensemble. For instance, if observations tend to be either lower or higher than all of the ensemble members, the ensemble is under-dispersed (U- or L-shaped histogram), which means that the ensemble members are too similar to each other. On the other hand, if the

observations tend to fall in the middle of the ensemble and very rarely outside the ensemble spread, the ensemble is said to be over-dispersed (inverted U-shape histogram). Random noise (~0.02 m) was added to the observed surges, as suggested by Hamill (2001) in order to represent uncertainty in the observed values

The surge forecasts were also compared after using a simple bias correction of the SBSS ensemble members, NOAA E-T models and SIT model. To compute the bias correction, two regression lines are calculated for each model member for storm surge cases > and < 0 m from the observed storm surge and the model error. This is because the sign of the error depends on the sign of the storm surge. The bias correction that is applied is then calculated for each model member using the latest storm surge observation. The RMSE and ME metrics comparing the bias and non-bias corrected forecasts will look into the benefit of this approach.

Finally, in order to test for statistical significance, a bootstrapping approach was used to resample the data and obtain proper confidence intervals around the means (Zwiers 1990). For each parameter (e.g., RMSE of surge error), a new sample of the same size was obtained by randomly selecting from the original sample and allowing for repeated selections. The mean was calculated and this process was repeated 1000 times. The 90% confidence intervals around the mean were determined by finding the 5^{th} and 95^{th} percentile of the means of all 1000 resamples. If the confidence intervals of two particular regions do not overlap, then they are significantly different at the 90% level.

Chapter III: Hurricane Gloria

a. Background

Hurricane Gloria originated as a tropical easterly wave off the west coast of Africa on 15 September 1985 (Fig. 3.1), and it became a tropical depression on 1200 UTC 16 September and a tropical storm on 1200 UTC 17 September. Gloria remained a minimal tropical storm for several days as it traveled west across the tropics from 14.6 °N, 28.3 °W to 17.7 °N, 54.2 °W. Gloria became a category one hurricane on 1800 UTC 21 September 1985, and began moving northwest in response to a weakness in the western portion of the Atlantic ridge caused by two previous hurricanes, Fabian and Henri (Case 1986). Gloria continued to strengthen over the warm (27-30 °C) Atlantic waters, and by 0120 UTC 25 September at ~ 24.2 °N, 70 °W, was a category 4 hurricane with a central pressure of 919 hPa, the lowest recorded pressure ever in the Atlantic at that time (Case 1986).

This study begins the analysis of Gloria starting at 0000 UTC 26 September when it was 940-hPa (~85 kts or category 2). Subsequently, Gloria began to weaken as it moved to the northwest and north and was ~944-hPa by 2000 UTC 26 September. However, a buoy (41002) 100-km east northeast of Gloria at 32.4 °N, 75.3 °W recorded a wave height of 14.3 m, which at the time was a record for a NOAA data buoy for a tropical cyclone (Case 1986). Gloria, with a central pressure of 943 hPa, made landfall at Cape Hatteras, North Carolina at 0600 UTC 27 September. Gloria continued north at ~

18 m s^{-1} (35 kts) and made landfall over central Long Island at 1600 UTC September as a category 1 hurricane and a central pressure of 961 hPa (Chris Landsea, personal communication, 2008). Although wind reports were sparse over Long Island during landfall, a peak gust of ~74-kts (38 m s^{-1}) was recorded at Islip, Long Island (cf. Fig. 1.1 for location). A storm surge of 1 to 2 m was recorded from North Carolina to Massachusetts (Case 1986).

b. Gloria observations and simulations

1. Synoptic evolution

At 0000 UTC 26 September, hurricane Gloria was situated around 600-km to the east of the central Florida coast. A 500 hPa trough extended southward from Minnesota to Texas (Fig. 3.2a), while an upper-level ridge was centered just east of Maine. Gloria was in an environment with little vertical wind shear, and southwesterly 500-mb winds were relatively light wind (~5 m s^{-1}) ahead of the storm at the coast.

At the surface at 0000 UTC 26 September, Gloria had an observed central pressure of 940 hPa (not shown), while the WRF 36 km domain was initialized at 947 hPa (Fig. 3.2b). Surface high pressure (~1021 hPa) was located over the Northeast U.S., while another surface high was located over eastern Wyoming (1027 hPa). Meanwhile, an inverted surface trough was situated over the lower Mississippi Valley, with an associated surface baroclinic zone over the southern Plains. This baroclinic zone was weaker to the east over the Appalachians and mid-Atlantic coast. The surface trough over the central U.S. developed into a 1006 hPa closed cyclone over southern Michigan by 0900 UTC 26 September (not shown), while the 500 hPa trough moved slowly eastward

given the relatively strong ridge over the western Atlantic. The more southwesterly flow aloft helped to steer Gloria more north-northeastward along the coast at this time (not shown).

At 0600 UTC 27 September (Fig. 3.3), the 500-hPa trough in the 30-h WRF simulation had become quasi-stationary over the Mississippi Valley region (Fig. 3.3a), while the 500-hParidge persisted over the North Atlantic. Both of these simulated features were close to the observed position (within 300-km) and magnitude (within 30-m) (not shown). Gloria at 500 hPa was embedded within the 15-20 m s^{-1} (30-40 kt) southwesterly flow along the East Coast, with the 500 hPa trough axis located ~1500 km to the west of Gloria at 0600 UTC 27 September. There was a larger separation between Gloria and the mid-level trough than tropical storm Floyd in 1999 (~750 km), which followed a similar track along the East Coast (Atallah and Bosart 2003; Colle 2003). As will be highlighted below, Unlike Floyd, Gloria did not dynamically couple with the mid-level trough, and thus, did not produce as much precipitation near the coast.

At the surface at 0600 UTC 27 September (Fig. 3.3b), surface high pressure (1017 hPa) was located over northern Texas (Fig. 3.3b), which was within 2 hPa of the surface observations (not shown). Meanwhile, there was a broad baroclinic zone exists at the surface along the East Coast. Gloria was located over Cape Hatteras, NC in both the observed and 36-km WRF, although the 36-km WRF (987 hPa) is much weaker than observed (942 hPa). Given these model errors, a more realistic simulation of Gloria requires higher resolution, which is highlighted in the next section.

2. Mesoscale evolution

Figure 3.4.a shows the observed track of Gloria from 0000 UTC 26 September to 0600 UTC 28 September and the tracks from 4 km WRF runs using different PBL and microphysical parameterizations, as well as a run using 12-km grid spacing (Table 1). All the WRF runs using the vortex initialization are within 150 km of the observed throughout most of the 54-h forecast, thus, the cyclone track is only slightly sensitive to model resolution and physics. The average WRF track has an error at hour 24-h (~46 km) and 48-h (~88 km) that is close to the average track errors from the National Hurricane Center in 2007 (24-h – 51.2 km, 48-h – 91.9 km) (Franklin 2008).

The remainder of this section focuses primarily on the control 4-km WRF (CTL) run (CTL on Table 1), defined simply as the WRF simulation with the smallest track and central pressure error at landfall over Long Island, which is needed to get the best storm surge simulation in section 3.d. The CTL run exhibits an eastward track error of ~100-km during the first 24-h simulation. Both the observed track of Gloria and the CTL run brush Cape Hatteras, NC at 0600 UTC 27 September (30 h), and the two tracks are located within 50-km of each other. Three hours before landfall over Long Island at 1600 UTC 27 September, the CTL run turns Gloria in a more northward direction, resulting in a landfall ~20-km to the west of the observed and ~30 minutes too slow (Fig. 3.4b).

Since central pressure is not a specified parameter in the Davis and Low-Nam (2001) initial vortex scheme (section 3), Gloria is initialized ~13 hPa too weak in WRF 4 km domains (Fig. 3.5). As a result of model spinup of precipitation around the storm, the WRF simulation also weakens Gloria too rapidly (5-6 hPa) during the first six hours of the simulation. As Gloria crossed the Gulf Stream between 1800 UTC 26 September and

0300 UTC 27 September (Fig. 3.4a), the observed storm deepens by ~5 hPa, which is likely the result of the additional latent heat fluxes over the relatively warm waters. All the WRF runs also gradually intensify the storm during this period, but the deepening rate and timing of intensification varied substantially. Thus, by 0300 UTC 27 September, the WRF central pressures for the various PBL and microphysics range from 931 to 958 hPa, while the observed is 941 hPa (Fig 3.5). After 0600 UTC 27 September, Gloria began weakening, with the observed central pressures increasing more rapidly than the WRF runs. As a result, by the time of landfall (~1500 UTC 27 September), the CTL run is nearly equal to the observed (~955 hPa). The other WRF runs range from 944 to 958 hPa at this time. Some of the reasons for these differences between the WRF runs will be discussed in section 3.c.

Using the flight-level observations from the NOAA P-3 aircraft, some of the hurricane wind structures from the WRF simulation were evaluated. At 0900 UTC 26 September, Gloria in the 4-km CTL WRF simulation is centered ~600 km east of central Florida (see X on Fig. 3.4a) with a central pressure of 959 hPa. The observed cyclone (945 hPa) is slightly southwest of the CTL simulation (Fig. 3.4a). The 4-km WRF run generates a semi-closed eyewall on the northern side of the storm at this time (Fig. 3.6a), with primary and secondary rainbands spiraling outward from the center[1]. Between 0830 and 1030 UTC 26 September, the NOAA P-3 aircraft flew two transects through the eye of hurricane Gloria (observed of ~945 hPa at 0900 UTC) at ~2800 m above sea level (ASL) (Fig. 3.7a). The strongest simulated surface winds (35-40 m s^{-1}) are just east of the storm center at this time (Fig. 3.7b). Although the WRF peak winds at 2800 m ASL are

[1] Simulated reflectivities were calculated using empirical relations based on the model cloud and precipitation mixing ratios as in Koch et al. (2005)

within 5 m s^{-1} of the observed just east of the eye during this period, the WRF winds 100-200 km east of the center were 5-10 m s^{-1} too weak. To the west (100-200-km) of the storm center, the WRF peak winds are 10-20 m s^{-1} too weak (Fig. 3.7b). Thus, the hurricane force winds in the WRF are smaller than observed at this time, especially to the west of the center.

By 2200 UTC 26 September, Gloria was located a few hundred kilometers south of Cape Hatteras, NC and it was moving northward. The speed of Gloria increased to ~10 m s^{-1} and the cloud shield expanded northward to southern New England (Fig. 3.8a). There is no well pronounced eye apparent in the satellite imagery at this time given the increasing mid-level winds (shear) over the storm (Fig. 3.8a). The 12-km CTL WRF run also produces mid- and high-level clouds that extend from South Carolina to New England (Fig. 3.8b), although the cloud shield coverage is ~ 300-km less than observed to the west of Gloria over Pennsylvania and West Virginia. The 4-km CTL is able to realistically simulate the cold cloud tops near the inner core of Gloria as well as some of the advection of dry air to the east of the center (Fig. 3.8b). The simulated radar reflectivity[2] from the CTL run also shows the asymmetry developing in the precipitation shield, with the more defined band to the west and north of the cyclone center (Fig. 3.6b). This is consistent with Chen et al. (2006), who showed that the tropical cyclone precipitation tends to occur more downshear (to the north in this case) due to the storm motion and vertical wind shear creating asymmetries in the precipitation field.

By 0600 UTC 27 September, Gloria was located near Cape Hatteras, NC (Fig 3.3b). A NOAA P-3 lower fuselage radar snapshot from Franklin et al. (1988) at 2.7 km

[2] Simulated reflectivities were calculated using empirical relations based on the model cloud and precipitation mixing ratios as in Koch et al. (2005)

ASL at 0538 UTC was compared with the 4-km CTL WRF run (Figs. 3.9.a,b). A well defined eyewall precipitation band exists north of the cyclone center, with the strongest reflectivities (38 dBZ) in the northwest portion. The maximum wind speeds (~50 m s^{-1}) at this level occurred ~60 km to the east of the center. The simulated dBZ at this same time (0545 UTC) and level also shows stronger reflectivity bands to the north and little precipitation to the south as dry air is wrapping into the storm (Fig. 3.9b). The simulated maximum winds are also ~60 km to the east of the center and are 5-10 m s^{-1} stronger than that observed. However, winds just east of the center of the storm are 5-10 m s^{-1} weaker in the simulation. The observed storm has a tighter core of faster winds at this time compared to the modeled storm.

Along the mid-Atlantic coast, Gloria increased its forward speed to ~35-kts, and by 1400 UTC 27 September, the center of Gloria is located just east of the central New Jersey coast, with a central pressure of ~954 hPa (Fig. 3.10a). The simulated Gloria in the CTL is located ~50 km further south along the New Jersey coast and a central pressure of 950 hPa (Fig. 3.10b) Overall, the sea-level pressure pattern along the East Coast is similar between the CTL and observed, although the simulated pressures were a few hPa too low around the Long Island and coastal Connecticut areas. Both the observed and CTL simulation have an enhanced surface temperature gradient over northeastern Pennsylvania (6°C per 100km) at this time. However, this boundary did not merge with Gloria circulation along the coast, thus the frontogentical forcing as was weaker than for tropical storm Floyd (Colle 2003). During Floyd, an intense baroclinic zone setup just inland of the coast (Colle 2003; Atallah and Bosart 2003). Only 20-23 cm of precipitation

fell across eastern PA and New Jersey during Gloria (Fig. 3.11a), while some of these same regions received 31-41-cm of rain during Floyd (Fig. 3.11b).

Two NOAA P-3 transects across Gloria at 1.5-km ASL as it moved along the central New Jersey coast between 1300 and 1400 UTC 27 September are compared with the 4-km CTL run (Fig. 3.12). Both the observed and CTL show a relatively large and strong area (45-50 m s^{-1}) winds at 1500-m, extending ~200-km to the east of the storm center (Fig. 3.12 a,b). At ~1300 UTC, the CTL simulation wind speeds are much stronger (>~28 m s^{-1}) closer to the coast and are ~7.5 m s^{-1} weaker ~200-km east of the storm center. While at ~1400 UTC, the strongest observed winds (~50 m s^{-1}) are located just south of central Long Island, with the CTL the strongest winds (~50 m s^{-1}) are located slightly further west (~25-km). These differences have occurred since the simulated Gloria was ~20 km closer to the New Jersey coast at the times of comparison. The P-3 flew directly through the observed Gloria's eye off the central New Jersey coastline but the simulated storm is ~ 20 km west of this location grazing the New Jersey coast. This track displacement also explains the discrepancy in wind direction between the CTL WRF and observed along the coast. However, wind direction further away from the New Jersey coast is out of the southeast at ~1300 and ~1400 UTC, which is close to that of the observed. Further east of the storm center (by 80-180 km), the CTL and observed winds are within 5 m s^{-1} of each other (Fig. 3.12b).

During landfall over central Long Island (~1600 UTC 27 September), there is an absence of convection on the south side of the storm in both the observations and CTL run (Fig. 3.13a,b).The cloud field has a distinct anti-cyclonic comma head to the north of Gloria, associated with a mid-level ridge building downstream of the cyclone (not

shown), which has been noted in other extratropical transition (ET) cyclone events (Klein et al. 2000) A radar snapshot at 1600 UTC 27 September of the CTL run shows most of the precipitation of Gloria is north and west of the cyclone at landfall (Fig. 3.6c), which is characteristic of an ET event (Klein et al. 2000). The eye is indistinguishable, and the south side of the storm is void of any precipitation as a result mid-level subsidence and dry air entering the system (not shown). The maximum surface winds were ~70 kts just east of the undefined eye, ~ 30 km to the south of Long Island.

c. WRF physics sensitivities

Previous studies have shown that the track and intensity of simulated hurricanes can be sensitive to the model physics (Braun and Tao 2000, Fovell and Su 2007). To illustrate some of these sensitivities for hurricane Gloria, a small 5-member ensemble of WRF runs were completed down to 4 km grid spacing by varying the planetary boundary layer (PBL) and microphysics parameterizations (Table 1). In addition, the output from the control physics run at 12 km grid spacing was also included to test the impacts of horizontal grid resolution on the evolution of Gloria.

During the first 30 h of the simulation as the storm moved northward towards Cape Hatteras, the cyclone tracks for the various 4 km ensemble members are clustered within ~20 km of each other and are located about 60 km to the east of the observed track (Fig. 3.4a). Meanwhile, the cyclone in the 12-km WRF is situated ~30 km to the west of the 4-km CTL run and ~30 km east of the observed. At 0600 UTC 27 September (30 h), all the simulations converge close to the observed track over Cape Hatteras, NC. At one hour before observed landfall (1500 UTC 27 September), all the WRF runs except the 4 km CTL are located between ~10 km (12 km CTL run) and ~40 km (4 km YSU-LIN) to

the east of the observed (Fig. 3.4b). Four of the 4-km simulations (CTL, GFS-FERR, YSU-FERR, MYJ-LIN) are close to the observed timing of landfall (within around 1 h), while the 12 km run as well as the 4-km YSU-LIN run are 1-2 h slower. Another difference between all runs and the observed is that the WRF runs slow Gloria down upon landfall to 16-18 m s^{-1}, while the observed storm maintains its forward speed of ~20 m s^{-1}. As a result, by 1800 UTC 27 September, the observed storm is over northern Connecticut (CT), while a few of the simulations (CTL, YSU-FERR, MYJ-LIN) are in southern CT and the GFS-FERR and YSU-LIN runs are over Long Island Sound. The 12-km simulation is the slowest, with the storm still over Long Island three hours after the observed landfall. Meanwhile, a simulation not using a vortex insertion technique has hurricane Gloria making landfall across Long Island four to five hours after the observed storm (not shown).

The central pressure of Gloria in the WRF runs varies by 10-20 hPa during the first 12-24 h of the simulation (Fig. 3.5). During the initial 6 h spin-up period, the simulated central pressures increase 6-9 hPa. The MYJ-LIN deepens rapidly from 959 to 951 hPa between 6-18 h, such that it is only 4 hPa weaker than the observed by 1800 UTC 26 September. The YSU-LIN pressures (959-957 hPa) during this same period are between the MYJ-LIN and other simulations. From 1500 UTC 26 September to 0600 UTC 27 September (15-30 h), as Gloria crossed over the Gulf Stream, the central pressures in the 4-km WRF runs deepen by ~10-hPa in the CTL, GFS-FERR, YSU-FERR, ~15 hPa in the YSU-LIN, and ~20 hPa in the MYJ-LIN, which are all greater than the observed (~4 hPa). Gloria also deepens by 12 hPa in the 12 km CTL run during this same period.

Over the cooler SSTs to the north of the Gulf Stream, the observed central pressure (942 hPa at Cape Hatteras) for Gloria steadily increases to 961 hPa at landfall at 1600 UTC 27 September The MYJ-LIN scheme weakens the cyclone at a similar rate to the observed (~17 hPa over 10 h), while the YSU-LIN run does not weaken as fast (~10 hPa). This indicates that there is also PBL sensitivity to the weakening of the hurricane. The other WRF simulations are 3-6 h slow with their weakening. Subsequently, the YSU-FERR run weakens similarly to the observed, while the CTL, 12 km CTL and GFS-FERR weaken very slowly (5 hPa) until landfall.

At landfall (~1600 UTC 27 September), the MYJ-LIN and CTL are ~15 hPa and ~10 hPa too deep, respectively. The YSU-FERR scheme is the closest (within 2 hPa) to the observed (~959 hPa). After landfall, the YSU-LIN and the CTL do not weaken (~25 hPa) as fast as the observed (~30 hPa), while the two FERR and the MYJ-LIN simulations weaken similarly to the observed.

Nolan et al. (2009) and Hill and Lackmann (2009) found that the MYJ PBL scheme in WRF also produced a deeper cyclone than the other PBLs, and they attributed this to a larger area-averaged latent heat flux across the vortex in the MYJ than the YSU PBL. Figure 3.14a shows a time series of maximum latent heat flux around the storm (every 3 h) for each Gloria WRF run. On average, the maximum latent heat flux in the MYJ simulation was consistently ~500 W m^{-2} larger than the other simulations. There are two three-hour spikes (~2250 W m^{-2}) in latent heat flux (starting at 1800 UTC 26 September and 0300 UTC 27 September) that coincide with more intense convection surrounding the storm center (not shown) and a more rapid deepening of the storm in the MYJ simulation These spikes that coincide with the more intense convection explain

most of the MYJ pressure difference with the other runs (Fig. 3.5). At 1800 UTC 26 September, the time of the first spike in latent heat flux (~2100 W m^{-2} MYJ-LIN vs. ~1000 W m^{-2} YSU-LIN and GFS-LIN) (Figs. 3.14b,c), the near surface (<500 m) inflow and outflow near the storm center in the MYJ-LIN is 10-15 m s^{-1} stronger than both the YSU and GFS PBL scheme runs (not shown). Hill and Lackmann (2009) also showed latent heat fluxes up to 1000 W m^{-2} (50%) larger in the MYJ simulations compared to the YSU simulations. They demonstrated that this difference between YSU and MYJ runs is the result of a larger surface exchange coefficient for moisture in the MYJ scheme. In the YSU scheme, this coefficient is nearly independent of wind speed, while in the MYJ scheme it increases linearly with wind speed and reaches a maximum at ~55 m s^{-1}. Thus, in the MYJ PBL, greater surface wind speeds yield larger latent heat fluxes and a deeper cyclone. The GFS PBL run is similar to the YSU, since they share the same surface layer physics. As a result of the similar fluxes in the GFS and YSU PBL, their central pressures are within a few hPa of each other throughout the simulation (Fig. 3.5)

At 1400 UTC 27 September (Fig. 3.15), two hours before landfall of the observed storm, all simulations except the YSU-FERR have the fastest 10-m winds on the eastern half the storm 30-60 km from the storm center. Weaker surface flow is favored on the western side, since the winds are decelerated by the coastal land areas. The surface winds in the YSU-FERR and the MYJ-LIN (43 m s^{-1}) are ~ 5 m s^{-1} stronger than the CTL and YSU-LIN at 1400 UTC. The horizontal extent of the wind field is largest in the MYJ-LIN run, with 60 kt winds extending ~140 km to the east of the center as compared to ~115 km in the other four simulations. The only observations of wind speed at this time comes from aircraft flight level data at ~1.5 km ASL, which when compared with the CTL from

1335 – 1415 UTC show a similar extent of the strong winds (45-50 m s^{-1}) 40-50 km south of Long Island (Fig. 3.12a,b). The other simulations except the MYJ-LIN also have a similar extent of strong winds. The MYJ-LIN has stronger winds (45–50 m s^{-1}) that extend outward an additional ~50 km (not shown).

At 1500 UTC 27 September (not shown), only one hour before observed landfall, all the simulations have the fastest wind speeds to the east of the storm center. Only the YSU-FERR has a secondary wind speed maximum located south of the storm center. The surface winds in the MYJ-LIN scheme (~41 m s^{-1}) are ~2-4 m s^{-1} stronger than the non-CTL simulations and 5 m s^{-1} stronger than the CTL. The horizontal extent of the surface winds is largest in the MYJ-LIN, GFS-FERR and CTL simulations, with the 60 kt winds extending ~115 km to the east of center compared to ~95 km in the YSU-LIN and YSU-FERR simulations (not shown)..

To evaluate the simulated surface wind and SLPs over the coastal waters near NYC, hourly observations at the Ambrose lighthouse (AMB on Fig. 1.1) at 30 m ASL were compared with the 4 km WRF run at this same level (Fig. 3.16). All runs except the CTL have slightly more northeasterly surface winds prior to landfall (1000 – 1500 UTC 27 September) than observed, and the winds are weaker than observed by 5-10 m s^{-1} in the FERR simulations and 2.5-5 m s^{-1} in the YSU-LIN and MYJ-LIN. The CTL winds are within 2.5 m s^{-1} of the observed, and the CTL winds are more east than the other members that tracked further to the east. The relatively weak (15 m s^{-1}) observed winds at AMB at 1600 UTC 27 September suggests that the eye of Gloria passed nearby at this time, while the strongest winds (~58 kts) associated with the northwesterly flow occurred two hours later as Gloria moved inland (Fig. 3.16b). At 1600 UTC, the CTL run

maintains an east southeast wind, while the winds shift to a more northwesterly (northerly for the YSU-LIN) direction in the observed and other simulations. All the other WRF runs that track to the east of the observed track do not have a wind speed reduction as the eye passed since their storm centers are too far east. The CTL run, which tracks ~20 km to the west of the observed, follows the observed AMB winds better than the other runs; however, its secondary wind maximum is 10-15 m s^{-1} too weak at 1800 UTC.

At John F. Kennedy Airport (JFK on Fig. 1.1) (Fig. 3.17a), all runs except the MJY-LIN are 10-15 kts weaker than observed leading up to (1100 – 1500 UTC) and after landfall (1800-2200 UTC). Meanwhile, the MYJ-LIN is 5-10 kts weaker than observed during these time frames (1100 - 1500 UTC and 1800 - 2200 UTC) and the peak wind in the MYJ-LIN at 1700 UTC is ~1 kt weaker than observed at 1800 UTC. The wind difference between the MYJ and other simulations could be due to several reasons. The MYJ winds are 5-10 kts stronger as it deepens Gloria 5-10 hPa more than other runs. The MYJ also mixes down faster winds more easily than the other simulations. For instance, the 10 m wind speeds are ~60% of the 30 m wind speeds for the GFS and YSU schemes; however, the 10 m wind speeds in the MYJ scheme are 90-95% of its 30 m wind speed. This is also seen at Islip, NY (ISP) (ISP on Fig. 1.1) (Fig. 3.17b), where the 10 m wind speeds from the non-MYJ PBL simulations were ~15 kts lower than the observed.

At Block Island (Fig. 3.17b), an island ~25 km south of the RI-CT border (BID on Fig. 1.1), the 4-km WRF simulations all have wind speeds 5-10 kts stronger than the observations prior to landfall (1400-1600 UTC) and after landfall (2000-2100 UTC). All members had simulated 10-m wind speeds within 10 kts of the observed from 1100-2200 UTC 26 September.

Figure 3.16c shows the SLP evolution at AMB. Gloria in the CTL run passes over AMB at 1600 UTC with central pressure of 954 hPa, which is 14 hPa deeper than that observed at this location. The other simulations to the east of the observed are up to 6 hPa weaker than observed at AMB. After landfall, the CTL and YSU-LIN do not weaken as rapidly and are slower than observed, thus the simulated SLPs at AMB are ~10 hPa lower than observed. The other simulations are also ~5 hPa lower than observed by 2200 UTC.

d. ADCIRC storm surge hindcast of Gloria

The ADCIRC was integrated using the hourly surface winds and pressures from the 12 km WRF and the 4-km movable WRF when over the ADCIRC grid starting at 0000 UTC 26 September. Although only the tides are forced along the model boundary, a ~1 m storm surge develops along the North Carolina coast an hour after the storm entered the ADCIRC domain at 0500 UTC 27 September (Fig. 3.18a). In reality, storm surge slightly above ~1 m (~3-4 feet) was observed at Cape Hatteras and ~110 km to the north (Jarvinen and McDuffie 1987).

As the storm moved northward along the coast, the area of ~1.0 m surge is located just northwest of the center of Gloria. By 1200 UTC 27 September, ~1.3 m surge is simulated along the coastal regions of Delaware and at the mouth of Delaware Bay, with the storm located just east of the Delaware/Maryland border (cf. Fig. 3.4a). Observations at Lewes, DE (LEW on Fig. 1.1) have storm surge levels around 1.2 m (~4 feet) (Jarvinen and Gebert 1986), which is close to that predicted by ADCIRC. Further north along the New Jersey shore, as the easterly surface winds forced water towards the coast (Fig.

3.18b), surge values of 0.25-1.0 m spread northward to Sandy Hook, NJ (SHK on Fig. 1.1).

At 1400 UTC 27 September (Fig. 3.19a), the storm is centered near the coast of southern NJ, and the surge has increased to 0.75-1.0 m near New York Harbor. The highest surge (~1.75 m) is still located over southern NJ, just northwest of the storm center at this time. Across the south shore of Long Island, the surge varies from ~0.75 m south of Brooklyn to ~0.5 m to the south of central Long Island (Fig. 3.19a). Meanwhile, surge levels around 0.5 m are located further into the harbor surrounding Manhattan (Fig. 3.19b).

At 1500 UTC 27 September (Fig. 3.19c) when Gloria is located to the east of northern New Jersey, maximum surge values of ~1.6 m are located just north of the storm center along the New Jersey coast, while the surge ranges from ~1.3 m near NYC to 0.5-m to the east at Montauk Point (MT on Fig. 1.1). At 1600 UTC (Fig. 3.20a.b), Gloria is located just south of Long Island, with a 1.5 m surge extending as far east as Suffolk county (SUF in Fig. 4.1), while there are ~1.75 m surge values near Brooklyn (BKLYN on Fig. 3.20b) and entering New York Harbor. Inside New York Harbor, a ~1.6 m surge exists from near the harbor between Manhattan and Long Island to the east (Fig. 3.20b). Meanwhile, a negative surge of ~0.5 m has developed along the NJ coast (Fig. 3.20a) as strong (~25 m s^{-1}) offshore winds (W-NW winds) push water away from the shoreline. At 1700 UTC, a high surge (~2.0 m) has moved into New York harbor surrounding Manhattan, while surge levels decrease quickly at the entrance of New York harbor and along the south shore of Long Island as the surface winds blow offshore (Fig. 3.21a,b).

Easterly to southeasterly winds also increase the surge to ~0.75-1.0 m within Long Island Sound at this time (Fig. 3.21a), which peaked at 1.25-1.75 m at 1900 UTC (not shown).

An ensemble of ADCIRC runs was created using the different WRF members listed in Table 1. The objective is to determine the sensitivity of the ADCIRC water level hindcasts to the changes in the predicted Gloria track, timing, surface winds, and WRF horizontal resolution. Figure 3.22 shows the surge and water level above Mean Low Low Water (MLLW) at the Battery for the CTL run, CTL-12 km, YSU-LIN, YSU-FERR, GFS-FERR, and MYJ-LIN WRF simulations. For several hours preceding the peak surge at ~1630 UTC, all simulations except YSU-LIN (1730 UTC) are ~0.15-0.3 m lower than the observed (Fig. 3.22a). This is probably due to weaker surface winds in the simulation ahead of landfall noted above for AMB and JFK (Figs. 3.16a and 3.17a). It is also hypothesized that the lack of wave forcing in ADCIRC may also lead to the surge under-prediction, which will be tested in section 3.e below. The CTL simulation produces a peak surge of 2.1 m, which is within 0.1 m of the observed peak surge; although the CTL storm surge peak is somewhat inflated at the Battery (by 0.05–0.1 m) due to the inverse barometer effect caused by differences in SLP between the observed (Fig. 3.16c). Subsequently, there is a rapid change from positive to negative surge from 1700-1900 UTC 27 September after the storm crosses over Long Island.

The other WRF-ADCIRC simulations, which are to the east of the observed track, do not perform as well as the CTL at the Battery. The peak surge in the CTL-12 km run is 0.32 m less than the CTL run and 0.25 m less than the observed. The GFS-FERR and LIN-FERR simulations produce lower peak surges, ~1.32 and 1.36 m, respectively, than the observations and other simulations. These runs also have slower development (by ~ 1

h) in the surge within New York Harbor leading up to landfall. The peak surge (~1.81 m) in the YSU-LIN simulation occurs an hour after the observed and the CTL run and it is 0.29 m less than CTL and 0.22 m less than the observed). This is likely because it tracks 30 km too far east and is one hour slow. The peak surge in the MYJ-LIN run is 0.62 m lower than the CTL, since it also tracks ~45 km to the east of the CTL run. The above results suggest that 20-50 km changes in storm track play a large role in the storm surge inside New York Harbor. This has potential big implications and challenges for forecasting storm surge in this region for landfalling hurricanes.

The total water level (storm-tide) needs to be accurately predicted for coastal flooding. As a result, the timing of the various WRF-ADCIRC forecasts is also important. Fig. 3.22b shows the water level (above MLLW) during hurricane Gloria at the Battery. Moderate flooding occurs at the Battery when the water level exceeds 2.4 m above MLLW (Jeffrey Tongue, National Weather Service, personal communication, 2008). Gloria made landfall close to a low tide, thus any predicted storm making landfall an hour or two later than observed will have a lower water level than observed. Even though the peak surge in the CTL run and the observed occur at the same time (~1700 UTC), their peak water levels do not due to the influence of the tides. The CTL simulation produces its peak water level at 1700 UTC, which is ~1 h later and 0.12 m less than the observed. This is due to the lower surge values prior to landfall which keeps the water level lower than the observed. The CTL water levels actually decrease slightly before landfall (1300 – 1500 UTC), while the observed water levels steadily increase. This occurs because the tide is decreasing, yet the CTL surge levels are less than the observed. The other simulations also have lower water levels before landfall as a result of an underprediction

of the surge. The YSU-LIN simulation, before landfall, has an even greater difference in water levels from the observed than the CTL, since Gloria in the YSU-LIN run makes landfall one hour later than observed and closer to low tide.

Even though Sandy Hook, NJ is outside the NY harbor (cf. Fig. 1.1 for location), the model results (Fig. 3.23) are similar to the Battery. The closest surge simulation to the observed is the CTL simulation, which predicts a 2.1 m surge, ~0.12 m lower than the observed (Fig. 3.23a). This is in contrast to the Battery where the CTL overpredicted the surge by < 0.1 m. Also, the MYJ-LIN simulation, which has 5 kt (10 kt) stronger winds and a 15 hPa (5 hPa) deeper storm than the observed (CTL) at landfall, has a peak surge .56 m lower than the observed at Sandy Hook as opposed to 0.62 m lower at the Battery. As at the Battery, this result is due to the track of the MYJ-LIN Gloria as it tracks east of the observed. All other simulations show similar results to those at the Battery. The similarities to the Battery continue in the water level evolution at Sandy Hook, NJ (Fig. 3.23b). The CTL again shows a peak water level ~1 h later and ~0.30 m less than the observed.

In contrast to the Battery and Sandy Hook, the models severely underpredict the water levels at Freeport, NY. Freeport, located along the south shore of Long Island, is protected by barrier islands (FRPRT in Fig. 1.1). While the timing of peak water level is consistent with that seen at Sandy Hook and the Battery, the closest simulation (CTL) to the observed peak water level (1.35 m) is too low by 1.34 m (not shown). The other simulations are around 1.7 m lower than the observed (not shown). This suggests that the current ADCIRC grid, which has grid spacing of 50 m in this back bay region of southern Long Island, is still not sufficiently resolved to properly simulate hurricane storm surge.

Gloria only produced minor flooding across the eastern end of Long Island even though its peak surge was the highest recorded at the Battery during the past 50-years. If Gloria made landfall five hours earlier during a high tide, the flooding may have been more extensive. To illustrate, several additional simulations were completed using the CTL wind field, but the wind forcing was shifted one hour ahead and behind the CTL run, and 5 hours before the CTL run (Fig. 3.24). There is a small difference (~0.05 m) in peak water level at the Battery between the CTL run and the one hour delay run. However, the difference in peak water level between the one hour early and CTL run is much larger (~0.30 m), since the time rate of change of the tidal water level is doubled compared to the one hour delayed run. As a result, even a one hour change in landfall could mean the difference between no flooding and flooding in some areas. For the worst case scenario, landfall 5 hours earlier during high tide increases the peak water level to 3.1 m above MLLW, which is large enough to create moderate flooding to areas bordering Inner New York Harbor comparable to the 1992 event (Colle et al. 2008). The five hour shift creates water-level differences around Manhattan and New York harbor that are ~0.7-0.9 m higher in the high tide landfall then the CTL simulation (Fig. 3.25 a,b) However, these plots do not show flooding around NYC, since unlike Colle et al. (2008), wetting and drying was not applied within ADCIRC

e. Wave impacts

Resio and Westerink (2008) noted that waves can increase storm surge through the momentum they impart into the water column through wave breaking. This transfer of moment cause by waves is termed a wave radiation stress. To determine the impacts of

waves on the Gloria water level hindcasts, the Simulating Waves Near Shore (SWAN) wave model was used (Chapter 2 Section 2.b). Surface winds every hour from the 12 km CTL run was used to force SWAN from 0000 UTC September 26th – 0600 UTC September 27th. After 0600 UTC September 27th, when the storm was located over Cape Hatteras NC, SWAN was run using 15 minute wind forcing to capture the wave field more accurately during landfall. SWAN was run using two domains, a coarser 8 km resolution domain that mimicked the size of the WRF 12 km domain and a second much finer grid (150 m resolution) situated over New York Harbor extending out past Sandy Hook, NJ (Fig. 3.26a). The only wave observations available were from buoys off Chesapeake Light, VA, near the mouth of Chesapeake Bay (CHLV2 in Fig 1.1). The highest recorded significant wave height occurs at 27 September 1000 UTC at a height of 6.2 m. In the SWAN, the maximum significant wave height is 4.14 m at 0900 UTC 27 September (Fig. 3.26a). After landfall across Long Island (1700 UTC), areas just outside New York Harbor experienced larger ~6 m waves in the SWAN run, with wave height decreasing northward into the harbor as the water depth becomes shallower and the waves break (Fig. 3.26b). Although SWAN may have underpredicted significant wave height near the coast, it still allows one to investigate the impact wave forcing on storm surge.

The impact waves have on the storm surge level is seen in Fig. 3.27. The peak surge value actually decreases by 0.07 m (within 0.01 m of the observed, although delayed peak by ~30 min) in the simulation including waves (CTL-Wave). However, the major impact (benefit) is in the storm surge levels leading up to peak surge (Fig. 3.27a). In the preceding 4 hours before landfall (~1300-1600 UTC), CTL-Wave has 0.15-0.2 m

higher surge on averaged compared to the CTL (Fig. 3.27b). During these hours prior to landfall, 3-5 m waves were located east of Sandy Hook and south of Jamaica Bay. The wave impact on storm surge during this time helped to increase the storm surge levels preceding landfall.

Chapter IV: Validation of Three Storm Surge Models and Ensembles

a. Background

Three storm surge modeling systems were verified in this study: the Stony Brook Storm Surge ensemble (SBSS), the Stevens Institute of Technology model (SIT), and the NOAA Extratropical Storm Surge model (ET). In addition to these modeling systems, two other ensembles were created using all 3 surge models and the SBSS ensemble (multi-model "ALL" ensemble) and the SBSS control member 9a, ET, and SIT models ("ENS-3" ensemble). The details of these models can be found in chapter 2.

Since the ET model was only available hourly, which differed from SBSS (15 minute) and SIT (6 min), the surge verification was hourly to be consistent between models. Also, the SIT began their forecasts at local midnight, while SBSS and ET began at 0000 UTC, thus all models were compared for the same forecast hour rather than the time of day. During the first 4 or 5 hours (depending on daylights saving), the ALL ensemble consisted of only the SBSS and ET members, since the SIT starts 4 or 5 hours later than the SBSS and ET. The models were also only inter-compared for days in which all three models had data. In total, 75 days were available (Table 4.1), including the full 8 member SBSS ensemble. The first 12 hours are left out due to missing data in the SBSS ensemble. Including only these available days would have reduced the number of available days to unacceptable levels for statistical significance. The verification was completed for five stations across Southern New England: The Battery, NY, Sandy Hook, NJ, King's Point, NY, Bridgeport, CT, and Montauk Point, NY (Fig. 4.1)

In addition, a separate set of ensemble members were verified after bias corrections were applied to them. It was found that the sign of the bias (mean error) depended on sign of the storm surge; thus, a simple bias correction of averaging the surge error for the past 7or 14 day forecasts did not work well (not shown). Rather, a bias correction surge errors was calculated depending on the magnitude of the observed surge using at least 30 days of previous forecasts as larger surges tend to have larger errors than surges closer to zero. After 30 days, a regression line is calculated from the storm surge observations and the storm surge errors. This line is then used to determine the bias correction from the storm surge observation. The observation from the previous time closest to the current forecast time is used to determine the bias correction (in this case the observation 15 minutes before is used to calculate the bias correction). After applying the bias correction, a new ALL ensemble mean storm surge forecast was calculated using the bias corrected individual members (BC-ALL).

b. Deterministic verification

The surge predictions are binned in twelve hour intervals from 12-48 h to increase the sample size. Figure 4.2a shows the mean error (in m) for storm surge versus forecast hour for all stations, which were averaged together to increase the sample size. All models have a persistent negative bias (ME < 0) over all forecast hours. The SIT bias is statistically similar to the SBSS ensemble (-0.02 to -0.03 m) for hours 12-24 and 36-48. The SBSS ensemble and the ALL ensemble have the lowest mean error during the 24-36 hr period. The NOAA-ET has the largest negative bias (-0.09 to -0.11 m), which is significantly different than the other models and its bias becomes slightly more negative

at greater forecast hours. Because all members have a negative bias, the ALL ensemble is also negative and it is similar to the SBSS ensemble mean given the large number of SBSS members in the ALL. After bias correction, the negative bias for the ALL ensemble is replaced by a slight positive bias for the BC-ALL at all forecast hours (0.01 to 0.03 m).

Figure 4.3a shows the mean error for the individual members of the SBSS ensemble. All of the members and ensemble mean are clustered together (within -0.02 to -0.03 m). The K2MY member has an error closest to 0 of any model member. From hours 36-48, the K2MY and GRBLK errors (~ -0.01 m) are statistically different than the other ensemble members at the 90% confidence interval. There is some suggestion that the WRF members (gray lines) have a slightly more negative bias at all hours, but the results are not statistically significant.

All surge members except the NOAA-ET system have a slight fluctuation of the mean error with forecast hour. This oscillation represents the models inability to correctly predict the semi-diurnal and diurnal tide correctly in some of these complex near-coast environments (Fig. 4.1). This error was carried over into the storm surge predictions when the modeled tide was subtracted from the total water level. This issue is not seen in the NOAA-ET since it does not predict the tide.

Figure 4.2b shows the root mean square error (RMSE) for the SBSS, SIT, and NOAA-ET members as well as the ALL and BC-ALL ensemble means. At hours 12-24 the 9a member and SBSS mean have errors similar to the NOAA-ET (~0.16 m). Meanwhile, the SIT system has the lowest RMSEs on average of the non bias corrected modeling systems before hour 24, including the ALL ensemble. After hour 24, the

NOAA-ET errors (~0.17 m) are the largest of all members and increase with forecast hour, while the SIT has smaller RMSEs than the SBSS, ALL and it is statistically similar to the ALL ensemble. The bias corrected ALL ensemble (BC-ALL) has the lowest RMSE error (0.11-0.12 m) of all members, statistically significant from all other members for hours 12-48, which clearly shows the positive impact removing the negative bias has on the accuracy of the ensemble mean (Fig. 4.2b). The original bias correction of removing a 7 to 14 day forecast error has RMSEs indistinguishable from the ALL ensemble, supporting the notion that this is not the right procedure (not shown).

As with the bias, the SBSS members are also clustered together for the RMSE for storm surge (RMSE's within 0.02 m) (Fig. 4.3b). Across all stations, the SBSS ensemble mean and the 9a member have slightly lower RMSEs all members, but the differences are not statistically significant. Generally speaking, the GRBLK has the largest error, which is significant at hours 12-36 compared to the ensemble mean.

Interestingly, the model accuracy for the SBSS and ALL ensembles does not decrease with increasing forecast lead time. Meanwhile, the NOAA-ET model and to a lesser degree, the SIT between 12-48 h, shows increasing RMSEs with forecast hour (Fig. 4.2b), which is expected given that atmospheric wind and pressure errors typically increase with greater lead time (Yu and Gerald 2004). In fact, the RMSEs in the ALL and SBSS decrease with forecast hour from 12-48 h. This suggests that the use of the previous day's forecast to initialize the SBSS members had a detrimental impact early in the forecast. Analysis of the available SBSS forecasts for hours 0-12 suggests that the error is also large for this period (not shown). All models except the NOAA-ET show little error growth between hours 24-48. This suggests that similar errors can occur at one

or two-day lead times given the challenges in predicting the tide and water levels in ocean models.

Overall, the BC-ALL accounted for ~80% of the explained variance of the observations (not shown) The SIT accounted for ~70% with decreasing percentages with increasing forecast time. The ALL ensemble accounted for ~65% of the observed explained variance slightly larger than the 61% in the SBSS ensemble. The NOAA-ET accounted for the least percentage of the explained variance ~50% with sharply decreasing percentages for forecast hours past 36 h (not shown).

In order to understand the spatial distribution of errors across the region, the MEs and RMSEs are calculated for the 12-48 h period in all models (Fig. 4.5). The SIT model outperforms the SBSS ensemble mean, NOAA-ET and ALL ensemble mean at stations along the Atlantic Ocean coast (Battery, Sandy Hook, and Montauk), since SIT has a relatively small (~ -0.02 m) bias (Fig. 4.5a). The SBSS and ALL ensembles have mean errors closer to zero at stations along Long Island Sound (King's Point and Bridgeport); thus, the ALL ensemble mean outperforms the other models in terms of RMSE, while the SBSS has the lowest ME. The NOAA-ET has the most negative ME and highest RMSE of any model at all stations. Also, the ALL ensemble mean has lower RSME than the SBSS ensemble mean at all stations, even outperforming the SIT at the above mentioned Bridgeport and King's Point. This shows that adding different ocean models into an ensemble, the accuracy can be improved relative to an ensemble of atmospheric models forcing one ocean model (ADCIRC).

The various models and ensemble were also evaluated for different surge thresholds to determine how their performance changes for smaller and larger events

(Fig. 4.6). The MEs were calculated at each station when either the model or observation exceeded the following surge thresholds (> 0.4 m, > 0.3 m, > 0.2 m > 0.1 m, > 0 m, < 0 m, < -0.1 m, < -0.2 m, < -0.3 m, < -0.4 m), while the RMSE's were calculated only for cases when the observation exceeded the threshold to keep the samples similar for each model (Fig. 4.6). The five stations are then combined to increase the sample size. For the ME, there are, on average, 720 hours when the observations/ model forecasts are < -0.4 and 455 hours when > 0.4 (Fig. 4.6a). For the RMSE, there are 715 hours when observations are < -0.4 and 450 hours when observations were > 0.4 (Fig. 4.6b). For the negative surge events (water less than tide) (Fig. 4.6a), all models except the NOAA-ET and BC-ALL have a similar positive ME (0.02-0.07 m). This error increases linearly as the negative surge increases so that for events < -0.4 m, the errors have increased by 0.1 m. The NOAA-ET bias is negative (~ -0.075 m) for negative surge events with the error increasing linearly to -0.02 m for events < -0.4 m. Meanwhile, the BC-ALL has errors slightly above 0 (0.03 – 0.01 m) for all negative events. Given the clustering among members, the ALL ensemble shares the same errors as the SBSS and SIT.

For the positive surge events (water level greater than tide), the models, except for the BC-ALL have negative ME which get more negative for larger storm surge events (Fig. 4.6a). The SBSS and ALL models have a negative bias that increases linearly with greater surge (-0.11 to -0.24 m). Meanwhile, the NOAA-ET mean error also increases linearly with increasing storm surge although to a lesser degree (-0.14 to -0.21 m) The BC-ALL mean error actually increases slightly with increasing storm surge events (0.02 to 0.05 m). For the largest surge events (> 0.4 m) the SIT and BC-ALL performs the best at all stations, with the SIT having a ME of ~-0.16 m and the BC-ALL having a ME of

0.05 m (Fig. 4.6a). The BC-ALL is slightly overcorrecting its members. This is potentially due to large error outliers which help to create a larger bias correction. While these large error events are then corrected, events with smaller errors are overcorrected leading to a positive average mean error.

Overall, these results suggest that the SBSS and SIT models underpredict the amplitude of the negative and positive surges (more so for the SBSS), while the NOAA-ET tends to underpredict the water levels for both positive and surges negative surges. However, the BC-ALL overpredicts for all events suggesting that the bias correction over corrects the original forecast.

Meanwhile, the RMSE errors are the lowest for SIT and BC-ALL as compared to the other models for surge events < 0 m (significant at 90% level for all cases < -0.1 - < -0.3 m) (Fig.4.6b). The BC-ALL has lowest RMSE for all surge events > 0 m (~ 0.1 m) (significant at 90% level), while the SIT has the second lowest error. The rest of the models have higher errors and are statistically similar to each other.

These results illustrate why a different bias correction was used rather than the conventional addition of the mean error for the past 7-14 forecasts, as done in many atmospheric ensemble studies (Eckel and Mass 2005; Jones et al. 2007). The bias is dependent on the magnitude of the storm surge event for some models, with larger negative (positive) surges lead to large positive (negative) mean errors. However, it does appear that bias correcting storm surge forecasts is a difficult problem as the use of a regression line slightly over corrects the forecasts leading to a slight positive error for all surges, positive or negative.

c. Relationship of surge errors to wind errors

The clustering of the MM5 and WRF members in ADCIRC suggests that the SBSS ensemble is under-dispersed, and that most of the bias is driven by the storm surge modeling system rather than the different wind and pressure atmospheric forcing. To show this, the surface winds were verified at buoy 44017 (44017 on Fig.1.1) for the NAM (used by Stevens) and all WRF and MM5 members (Fig. 4.4). Due to missing data in both the NCEP-NAM (Dec. 2008) and the SBSS ensemble, comparisons are made for only 50 days. At buoy 44017, the surface wind errors are more negative (~ 0.5 m s^{-1} lower) in the NAM then the MM5/WRF (Fig. 4.4a), yet the SIT ocean model mean errors are statistically similar than ADCIRC (Fig. 4.2a). The NAM has one of the largest RMSEs (Fig. 4.4b); however, the SIT ocean model has statistically significantly lower RMSE than the SBSS ensemble members (Fig. 4.2b). Upon closer inspection of the surface stress formulations used in ADCIRC (Garratt 1977), and SIT (Large and Pond 1982), it was revealed that SIT has 15-25% less stress than ADCIRC for wind speeds 11-25 m s^{-1}, and similar for other winds. Thus, the higher water levels in SIT are not the results of surface stress formulations, but rather from ocean model setup (spinup procedures, model three-dimensionality, etc . . .). The ADCIRC forecasts errors are relatively clustered even with some diversity in the wind errors (Fig. 4.4), which suggests using more atmospheric members and one ocean model favors similar biases/errors and oes not improve its deterministic performance.

d. Probabilistic Verification

An important objective of an ensemble is to improve probabilistic skill as compared to deterministic (single) model. In this section, the verification of the SBSS, ALL and BC-ALL ensembles are performed. Ideally, each of the ensemble members should have similar accuracy, since they are weighted evenly for the mean. Figure 4.7 shows the percentage of the time that each member performed better than the other members for hours 12-48. In the original SBSS ensemble (8 members), the SBSS control member (9a) is consistently the best performing model 18-25% of the time at the various stations (Fig. 4.7a). The K2MRF for the MM5 and GFS-WRF and NOGAPS-WRF (see Table 2.1 for members) are the second set of highest percentages (10-20%) for many stations.

With the addition of the ET and SIT to the ensemble, the best member percentages in the ALL ensemble are much different (Fig. 4.7b). By far, the best member is the SIT model at all stations (25-47%), while the NOAA-ET member and MM5-K2MY are the second highest set of members (10-15%). The inclusion of the SIT and ET model does reduce the number of times the control SBSS member 9a performs the best relative to the other SBSS members in the ALL ensemble (Fig. 4.7b).

The percentage of time each member is the worst is also calculated (Fig. 4.8). For the SBSS ensemble (Fig. 4.8a), the GRBLK and K2MY members are consistently the worst members at all stations, while the GFS-WRF and K2MRF are rarely the worst member (the ensemble mean is never and can never be the worst member) (Fig. 4.8a). With exception of the K2MRF, the WRF members are less often the worst members when compared with the MM5 members. For the ALL ensemble (Fig. 4.8b), the NOAA-

ET member is the worst member the most amount of time at all stations (30-50%) Members 9a, GFS, K2MY, and NOG are the worst members the most infrequently (< 10%). The NOAA-ET member has a relatively high percentage for the best and worst member ranking. This is because the NOAA-ET is frequently an outlier in storm surge forecasts. When it performs well, it would mean that the other models do not and vice versa.

In order to quantify how well the surge ensemble represents the spread around the observed values, rank (Talagrand) histograms are were calculated for each ensemble with noise added (0.02 m) according to Hamill (2001) (Fig. 4.9). To increase sample size and to see how the models performed across the region, the rank histograms are calculated for all stations. The rank histogram for the SBSS ensemble is U-shaped (Fig. 4.9a), which suggests that the ensemble is under-dispersed. Thus, the observation falls outside the ensemble envelope, either above or below, a majority of the time (~68% at all stations) (Fig. 4.9a). In addition, the observations is outside the envelope on the high end of the ensemble more (~38% of time) than the low end (30 % of time) of the ensemble. These results are consistent with the clustering of the SBSS members deterministic verification above (Figs. 4.2, 4.3). The rank histograms are also calculated using a bias correction for all members. The ensemble is slightly more dispersed as the amount of the time the observation fell outside the ensemble dropped from 68% to ~58% (Fig. 4.9a).

The addition of the SIT and NOAA-ET members helps make the ALL ensemble more disperse (~ 38% more inside ensemble that SBSS) (Fig. 4.9b). In particular, the addition of the NOAA-ET into the ensemble helps to broaden the dispersion of the ensemble, especially for the lower ranks as the NOAA-ET consistently has a more

negative bias than the other members (Fig. 4.2a, Fig. 4.5a). However, the middle ranks are still under-populated, creating a U shaped rank histogram, though not as severe as in the SBSS ensemble. Performing the bias correction on the ALL ensemble helps the ensemble bias very slightly. The BC-ALL changes the sign of the bias of the ensemble as seen by the decrease in the percentage of time the observation falls above the ensemble (ALL – 31%; BC-ALL – 13%) and reverse happening when the observations falls below the ensemble (ALL – 9 %; BC-ALL – 24 %). Unfortunately, the BC-ALL only makes the ensemble slightly more disperse, very slightly increasing the amount of time the observation falls outside the ensemble by ~3 % (Fig. 4.9b).

To determine the probabilistic skill of the ensemble, the Brier Score (BS) and Brier Skill Score (BSS) were calculated for each ensemble for five positive storm surge thresholds determined by the observations and model forecasts (>0, >0.1,>0.2,>0.3,>0.4) for all 5 stations during the 12-48 h period (Fig. 4.10). The BS in the SBSS and ALL ensembles decreases (improves) as the positive surge became larger (Fig. 4.10a). This suggests that the probabilistic skill increases for the larger observed surge events. The BC-ALL has a statistically significantly lower BS for all events than the ALL and SBSS ensembles (Fig. 4.10a). The BS differences between the SBSS and ALL ensembles are statistically significant at the 90% level for all surge thresholds except for > 0.4 m with the ALL ensemble having a lower BS. The improvement of the ALL ensemble relative to the SBSS ensemble is relatively small, since the SBSS accounts for 80% of the ALL ensemble. A multi-model ensemble with equal weight between the 3 modeling systems was also verified, which included only the SBSS control member 9a, the ET and the SIT (hereafter referred to as ENS-3 ensemble). This allows each forecasting system to be

weighted evenly. The three member ensemble (ENS-3) has statistically significant lower (better) BS than the SBSS at all thresholds except > 0.4 m and the ALL for events > 0 m. The limited benefit of the ENS-3 compared to the ALL ensemble is likely related to relatively small membership of the ENS-3. However, these results illustrate that a multi-model ensemble made up of only a few different surge models can yield more skill than using one surge model (ADCIRC) and multiple atmospheric forcings from WRF and MM5. A bias corrected ENS-3 (BC-ENS-3) also performs better than all ensembles except the BC-ALL where it is instead statistically insignificantly different at all thresholds except > 0.4 m (Fig. 4.10a).

The BSS was calculated relative to the best member of the SBSS ensemble (control member 9a). There is positive skill across all thresholds for the SBSS and ALL ensembles (Fig. 4.10b). For the SBSS, the BSS (~0.10) remained constant as the surge threshold increases. The ALL ensemble has statistically significantly higher BSS than the SBSS for all surge thresholds, maintaining a BSS of ~ 0.2 (Fig. 4.10b). Meanwhile, the ENS-3 ensemble outperforms the ALL at all thresholds except > 0.4 m with a BSS of ~ 0.25. The BC-ALL and BC-ENS-3 ensembles shows the benefit of removing biases. The BC-ALL outperforms all other ensembles at 90% confidence level for all surge thresholds. The BC-ENS-3 performs statistically similar to the BC-ALL at all thresholds although for surge thresholds > 0.4 m the BC-ENS-3 performs notably worse than the BC-ALL (Fig. 4.10b). All models show a decreasing BSS for larger storm surge events as all models seem to perform better probabilistically for these events.

The BS was broken down into reliability and resolution (Table 4.2). Since the uncertainty is greater than the BS for all ensembles, it can be seen that the ensembles will

have a positive BSS against climatology. Across positive surge thresholds, the ensemble's BSS versus climatology shows similar results to figure 4.10b (Fig. 4.11). All ensembles except BC-ALL and BC-ENS-3 show only slightly positive BSS for cases > 0 m which exemplifies the ensembles low bias. All ensembles at all thresholds have BSS greater than 0, which shows a positive skill against climatology. The BC-ALL and BC-ENS-3 outperform all ensembles and climatology at all thresholds. However, the BC-ALL has larger BSS compared to the BC-ENS-3 (statistically significant at 90% confidence interval) for surge events > 0.4 m.

Fig.4.12 shows reliability diagrams for surge events greater than 0.3 m (Fig. 4.12a) and 0.4 m (Fig. 4.12b). On each diagram, the ALL and SBSS are plotted along with a horizontal line referring to climatological relative frequency of a > 0.3 m (> 0.4 m) surge event. This line represents an ensemble having no resolution. The diagonal line represents an ensemble having perfect reliability. A perfectly reliable ensemble would forecast its probabilities equal to that of which occurs naturally (i.e., if half the ensemble predicts the event to occur, it should occur 50% of the time). The smaller dash diagonal line represents no skill. If the ensemble lies above this line, the ensemble has skill over climatology. Lastly, histograms of the number of events for each forecast probability for the ALL and BC-ALL are added to the plot.

For surge events greater than 0.3 m, the SBSS and ALL ensembles are above the perfect reliability line for forecast probabilities less than 80%. For events in which only a few ensemble members predict greater than a 0.3 m surge, the observed probability of event occurring is higher than it should be. For instance, when the ensembles are predicting an event to occur 30 % of the time, the event actually occurs 60% and 70% of

the time for the SBSS and ALL ensembles, respectively (Fig. 4.12a). This continues for moderate probability events (0.4 -0.7). For higher probabilities (> 0.8), both ensembles have forecast probabilities roughly equal to the observed probability (Fig. 4.12a). The various ensembles become more reliable as forecasted probabilities increase for both models. In general, the ALL ensemble, however, is slightly more reliable than the SBSS ensemble, although both of their probabilities are unconfident in general. For a slightly larger storm surge event (> 0.4 m), the ensembles perform similarly (Fig. 4.12b). For smaller forecast probabilities, events occur more often than they should for both ensembles making the ensembles unreliable. However, for higher forecast probabilities, the ensembles become more reliable (Fig. 4.12b).

In addition to the SBSS and ALL ensembles, the BC-ALL and ENS-3 ensembles are also plotted (Fig. 4.12). For events larger than 0.3 m, the BC-ALL ensemble performs surprisingly worse, hovering around the no skill line for all probabilities less than 0.8 (Fig. 12a). In addition, the BC-ALL lies below the perfect reliability line for all probabilities. For probabilities forecasts lower than 0.8, the BC-ALL is overconfident. For instance, when the BC-ALL predicts an event to occur 30% of the time, the event occurs 20% of the time. For high probability events, the BC-ALL is still generally reliable. When the BC-ALL predict an event occurs 90% of the time, it occurs 60% of the time. Meanwhile, the ENS-3 ensemble, consisting of the SBSS ensemble mean, NOAA-ET and SIT, is the most reliable (Fig. 4.12a), once again illustrating the benefit of running a multi-surge model ensemble system. For events larger than 0.4 m, the BC-ALL lies far underneath the perfect reliability line and around the no skill line for all forecast probabilities of < 0.9. Although, the BC-ALL does show skill for higher probabilities

(> 0.8). The BC-ALL continues to be overconfident. Meanwhile, the ENS-3 ensemble is closest to the perfect reliability line of any ensemble for its five available forecast probabilities (Fig. 4.12b). Interestingly, while the BC-ALL performed better than the other models deterministically and in BSS, it has a worse reliability. A potential reason for its overconfidence can be seen in the histogram of events. In both surge events, the BC-ALL increases the number of events that have higher forecast probabilities (> 0.8) when compared to the ALL ensemble (Fig. 4.12). By bias correcting the ALL ensemble, events which would have moderate or low forecast probabilities in the ALL ensemble become events that have high forecast probabilities in the BC-ALL. Events that have low or moderate forecast probabilities in the BC-ALL now mainly come from events that had low or 0 forecast probability in the ALL. This leads the BC-ALL to be overconfident as seen in Fig. 4.12. This suggests that the BC-ALL is overcorrecting the forecasts leading to an overpopulation of higher forecast probability events in the overconfident BC-ALL. This overconfidence is seen in the ME error plots as well as the BC-ALL consistently has a ME > 0 m. To test this idea that the bias correction is in fact over correcting the ensemble, a bias correction only using 75 % of the current bias correction is performed. Reliability diagrams for surge events > 0.3 m and > 0.4 m using this 75% BC-ALL show that the ensemble now becomes more reliable at all forecast probabilities and less overconfident compared to the original BC-ALL (Fig. 4.13). By slightly undercorrecting the ensemble forecasts, the 75% BC-ALL is less overconfident and a more reliable model.

e. Discussion

Sections b and c in this chapter showed that the predicted surges in the individual model members were slightly too low on average. This was independent of time of day, forecast period, and the station. The smallest negative errors occurred in the Long Island Sound (Fig. 4.5a), yet the errors were still negative. Fig. 4.14 shows a time series of the error at all stations combined for the 75 available days for the 24-48 hr forecasting period. Overall, there is a slight negative bias (< 0.05 m) in the SBSS ensemble mean (Fig. 4.14); however, there are periods of positive error as well. For example, the SBSS errors are positive from 20-28 January 2008, while there is a negative error in late November 2008. Overall, there are several events with relatively large errors (> 0.2 m), thus suggesting room for improvement in the 1-2 day forecast.

It was hypothesized that there are certain weather patterns that yield larger negative and positive errors. A composite was constructed using the NCEP-NCAR daily reanalysis (Kalnay et al. 2001) showing sea level pressure (SLP) for the 10 most negative and positive daily errors days in the SBSS control member 9a at the Battery (Fig.4.15, Fig.4.16). The composite of the 10 largest negative error days (> -0.18 m) for control member 9a shows the presence of a cyclone located just off Cape Cod, MA (Fig. 4.13c). This is a familiar setup during the cool season for larger surge events for NYC as Nor' Easters track up the coast (Colle et al. 2009). At 48 hours before the time of the largest error (Fig. 4.15a), the SLP pattern of the largest negative surge error events consists of a strong high pressure over the western United States, with a trough of low pressure beginning to dig southward over the western Great Lakes (Fig. 4.15a). At 24 h before the event, a deepening low pressure system has developed off the Mid Atlantic Coast (Fig.

4.15b). The development of the cyclone allows for a couple days of northeasterly flow which helps push water along the southern New England coast and cause high storm surge. It appears that the largest negative errors occur during this type of long fetch storm surge events.

In contrast, the SLP composite of the 10 largest positive errors (> 0.10 m) in 9a shows the presence of a large ~1030 hPa high pressure off the coast of VA and NC (Fig. 4.16c). Two days prior, the area of high pressure (~1025 hPa) was located over the Great Plains, with higher pressures spreading eastward towards the northeast (Fig. 4.16a). At 24 hours prior to the surge error, the high pressure (~1025 hPa) has moved further east, situated over the Southeast, while low pressure is departing the Northeast (Fig. 4.16b). Due to the position of the high and low pressure at the surface, a pressure gradient exists over the northeast to increase northwest winds. By the day of the event, when the high has moved offshore to the south of the region, winds have been blowing offshore for a couple days over the northeast helping storm surges become negative. These results suggest that the model has a difficult time capturing the movement of water away from the coast during these events.

The largest negative surge error day involved the strongest coastal storm during the time period in question. From 2 November to 4 November, 2007, Hurricane Noel was becoming extratropical and moving quickly north parallel to the eastern seaboard. At 0000 UTC 3 November, Noel was located east of North Carolina with 10-20 kt northeasterly winds across the Northeast (Fig. 4.17a). At 12 hours later, Noel was located east of Maryland, while northeasterly winds increased to 20-30 kts across Cape Cod (Fig. 4.17b). At 0000 UTC 4 November, the low has deepened to 970 hPa and is due east of

Cape Cod. The 10-20 kt winds over New York and Long Island veers to the northwest (Fig. 4.17c). By 1200 UTC 4 November, the extratropical cyclone moves northward into the Canadian Maritimes, leaving behind 5-10 westerly winds to the south (Fig. 4.17d). While the largest daily error occurred on the 4th, the storm really affected the area more so on the 3rd. However, since the SBSS ensemble is "hot started" whatever error was left over on the 3rd was carried over into the 4th, which helped to contribute to the large negative error seen on that day.

One of the possible reasons for the negative surge on storm days is the impact of waves on storm surge, which is neglected in the models. Westerink et al (2008) mentioned that waves positively enhance storm surges by 0.6 - 1.2 m during hurricane Katrina across Louisiana. To illustrate the potential impact of waves on the verification surge results, significant wave heights at Buoy 44017 (~40 km southwest of Montauk; 8 on Fig. 1.1), are averaged daily over the 75 days (Table 4.1) in question and compared with the corresponding day two (24-48 h) surge error of the SBSS ensemble mean at Montauk Point (Fig. 4.18). When there are larger wave heights, the surge errors tend to be more negative. This occurs more so when waves are larger (> 2.5 m). While the negative surge errors that occur during large wave days could also be caused by the storm that caused the waves, this trend still does exist for smaller waves between 1-2 m. Though it is not a large component, waves do appear to influence the surge over the course of the cool season and the exclusion of them in the models can cause a negative storm surge bias.

Lastly, the impact of the ocean model used for storm surge predictions cannot be neglected. The SBSS model uses ADCIRC, a 2D model, which parameterizes bottom

stress based on a depth averaged velocity within the water column. SIT uses a 3D ocean model (POM) that includes a bottom layer velocity which is used to calculate bottom stress. Weisberg et al. (2008) found that 2D models over estimate bottom stress, compared to 3D models, which leads to an underestimation of surge heights in 2D ocean models.

Chapter V: Conclusions

a. Hurricane Gloria

The Weather Research and Forecasting model (WRF) at 4-km grid spacing was shown to realistically simulate the evolution of Gloria from 0000 UTC 26 September until landfall 42-h later. A comparison of WRF with P-3 aircraft observations showed that the control (CTL) Gloria simulation produces similar wind speeds (within 5 m s^{-1}) at flight level two hours before landfall (1400 UTC) 80-180 km away from the storm center. The CTL simulated Gloria made landfall within one hour of the observed, only ~20 km west of the observed landfall. By changing certain parameterized schemes (PBL and microphysics), a suite of runs similar in development of Gloria yet slightly different in strength, track and timing is created. The use of the MYJ PBL scheme causes a significantly deeper hurricane Gloria simulation (10 hPa deeper than observations at lowest central pressure) due to the different handling of latent heat fluxes within the PBL scheme. These various runs allow for a test of the ocean model's (ADCIRC) performance and sensitivity to small changes in track, timing and intensity of hurricane Gloria with regards to storm surge forecasts for Southern New England.

Overall, there is a large sensitivity in surge and water levels due to small changes in track and intensity of a fast moving hurricane impacting southern New England. The CTL simulation, which tracks hurricane Gloria 20 km west of the observed, overpredicts surge by 0.1 m at the Battery and underpredicts the surge slightly (by 0.12 m) at Sandy Hook. However, surge heights differ by up to 0.7 m (0.8 m) in other simulations at the

Battery (Sandy Hook) by changing the PBL and microphysics schemes, which track the cyclone up to 40 km east of the observed and change the storms intensity. In comparison, all simulations have difficulty resolving storm surge along the south shore of Long Island at Freeport, which is underpredicted by over 1.0 m. While surge predictions might be correct or within 0.1 m of the observed, there still might be a large difference in peak water level. Even an hour shift in landfall timing can have a large impact on the water levels. Differences between the one hour early CTL run and the CTL are up to 0.3 m different in water level. In a worst case scenario if hurricane Gloria made landfall 5 hours earlier during high tide, water levels would have peaked at the Battery 0.8 m higher than the flood level of 2.4 m MLLW.

In all of these simulations, the storm surge and water level forecasts tended to underpredict the storm surge. One factor that has been left out as an influence on storm surge is the impact of wave forcing. By including wave forcing in the ADCIRC simulations, the storm surge levels landfall increased by 0.15-0.2 m at the Battery in the 4 hours preceding landfall.

Future work would include using the latest version of WRF (V 3) and its simple ocean model so that the effect of wake-induced cooling of SSTs on storm intensity is included. In addition, water level simulations should be undertaken using a 3D model as opposed to the 2D ADCIRC to test the influence of parameterizing bottom stress in 2D models.

b. Validation of three storm surge models and ensembles

During the November 2007 – March 2008 and October 2008 – December 2008, there were 75 days during which there was a full 8 member SBSS ensemble. The SBSS, SIT and NOAA-ET models are compared for this time period. For 12 hour bins, the SBSS ensemble shows a clustering of members within 0.03 m in ME and RMSE. This suggests that the addition of more atmospheric members into the ensemble would not be beneficial.

An inter-model comparison at all stations shows that the NOAA-ET has the largest negative error (-0.10 m) while the SIT has the lowest RMSE (~ 0.13 m). The SBSS ensemble mean and the SBSS control member 9a have show a decreasing trend in RMSE with forecast hour, which could be due to hot starting the model using the previous days forecast. The creation of a new multi-ocean model ensemble including all members of the SBSS ensemble, the SIT and NOAA-ET (ALL ensemble) lowers the RMSE compared to the SBSS by ~0.02 m, statistically significant at 90% confidence interval for all forecast hours. A bias correction is also applied to the ALL ensemble. However, since the bias changes depending on whether there is a positive or negative surge, a bias correction that is simply an average of the previous 7 forecasts would not be beneficial. Therefore, a bias correction using a regression line calculated from storm surge observations and storm surge errors is calculated for each member. The BC-ALL reduces the RMSE to ~0.11 m while slightly overcorrecting each member as seen by a slight positive ME (0.01 m) over all forecast hours.

Probabilistically, the SBSS and ALL ensemble are underdisperse as seen in their rank histograms. The SBSS is severely underdisperse with storm surge observations

falling outside the ensemble 68% of the time. Bias correction helps to improve the dispersion slightly in the ALL ensemble (by 3%). Brier score and Brier skill scores calculated for surge thresholds > 0, > 0.1, > 0.2, > 0.3, and > 0.4 m shows the benefit of including ensemble members that use a different ocean model. The ALL ensemble improves upon the SBSS in brier score and brier skill score even though 80% of the members of the ALL consist of the SBSS. If a three member ensemble is created (ENS-3), consisting of the SBSS ensemble mean, SIT, and NOAA-ET, the brier score is lowered and brier skill score increased for all surge thresholds compared to the ALL ensemble. The BC-ALL and BC-ENS-3 perform better (statistically significant at 90% confidence interval) in brier score and brier skill score than all other ensembles. Reliability diagrams for surge events > 0.3 m and > 0.4 m shows that the SBSS and ALL ensembles are underconfident for forecast probabilities less than 0.8. The ENS-3 increases the reliability even though the ensemble is only three members. Since the BC-ALL is overcorrected, the ensemble is overconfident for forecast probabilities less than 0.8 while less overconfident for higher forecast probability events. The BC-ALL overpopulates higher forecast probabilities which lead to this overconfidence. If only 75% of the bias correction is used so that the ensemble is slightly undercorrected, the ensemble becomes much more reliable.

Overall, ADCIRC seems to have trouble moving water into and out of the region. This is seen through a composite of SLP for the ten largest positive and negative error days for SBSS control member 9a at the Battery. ADCIRC tends to underpredict cyclone events while overpredicting offshore (negative surge) events. . This is not from a low MM5 or WRF bias in surface wind speed or the surface drag formation (SIT applies a larger drag), but it

may be the result of running ADCIRC in two dimensions. Another potential impact neglected in all models is the influence of waves. It is shown that larger wave heights at an offshore buoy tend to correspond to larger negative error days at the coast.

Overall, these results have shown the benefits of using a multi-model surge prediction system. Forecasters would benefit if surge predictions from different operational ocean models could be combined, rather than using just one surge model and an atmospheric ensemble.

Future work is needed to determine the day to day influence waves have on real-time surge forecasting systems as well as determining the benefit of a multi-ocean model ensemble approach over the course of an entire cool season.

Literature Cited

Atallah, E. H., and L. F. Bosart, June 2003: The Extratropical Transition and Precipitation Distribution of Hurricane Floyd (1999). *Mon. Wea. Rev.*, **131**, 1063-1081.

Baxter, P. J., June, 2005: The east coast Big Flood, 31 January – 1 February 1953: a summary of the human disaster. *Phil. Trans. R. Soc. A.*, **363**, 1293-1312. doi:10.1098/rsta.2005.1569

Bender, M. A., and I. Ginis, Apr. 2000: Real-Case Simulations of Hurricane-Ocean Interaction Using a High-Resolution Coupled Model: Effects on Hurricane Intensity. *Mon. Wea. Rev.*, **128**, 917-94.

_____,_____, I. Ginis, R. Tuleya, B. Thomas and T. Marchok, Dec. 2007: The Operational GFDL Coupled Hurricane-Ocean Prediction System and a Summary of Its Performance. *Mon. Wea. Rev.*, **135**, 3965-3989.

Betts, A. K., and M. J. Miller, 1993: The Betts-Miller scheme: The representation of cumulus convection in numerical models. *K. A. Emanuel and D. J Raymond, Eds., Amer. Meteor. Soc.*, 246 pp.

Blier, W., Keefe, S., Shaffer, W., and S. Kim, Dec. 1997: Storm Surge in the Region of Western Alaska. *Mon. Wea. Rev.*, **125**, 3094-3208.

Blumberg, A. F., L. A. Khan, and J, P. St. John, Aug. 1999: Three-Dimensional Hydrodynamic Model of New York Harbor Region. *J. Hydr. Engrg.*, **125**, 799-816.

Booij, N., R. C. Ris, and L. H. Holthuijsen, Apr. 1999: A third-generation wave model for coastal regions 1. Model description and validation. *J. Geophys. Res.*, **104**, 7649-7666.

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 Report. Marine Sciences Research Center Technical Report. Stony Brook University*, 28 pp.

Braun S. A. and W.K. Tao, Dec. 2000: Sensitivity of High-Resolution Simulations of Hurricane Bob (1991) to Planetary Boundary Layer Parameterizations. *Mon. Wea. Rev.*, **128**, 3941-3961.

Brink, K. H., July, 1989: Observations of the Response of Thermocline Currents to a Hurricane. *J. Phys. Ocean.*, **19**, 1017-1022.

Brooks, C. F., Jan. 1939: Hurricanes into New England: Meteorology of the Storm of September 21, 1938. *Geo. Rev.*, **29**, 119-127.

Bruno, M. S., A. F. Blumber, and T. O. Herrington, 2006: The urban ocean observatory – coastal ocean observations and forecasting in the New York Bight. *J. Mar. Sci. Environ.*, **No. C4**, 31-39.

Case, R., July 1986. Annual Summary: Atlantic Hurricane Season of 1985. *Mon. Wea. Rev*., **114**, 1390-1405.

Chen, S. H., W. Y., Sun, 2002: A one-dimensional time-dependent cloud model. *J. Meteor. Soc. Japan*, **80**, 99-118.

Chen, S. S., J. A. Knaff, and F. D. Marks Jr., Nov. 2006: Effects of Vertical Wind Shear and Storm Motion on Tropical Cyclone Rainfall Asymmetries Deduced from TRMM. *Mon. Wea. Rev.*, **134**, 3190-3208.

Chou, K.H. and C.C. Wu, Mar. 2008: Typhoon Initialization in a Mesoscale Model - Combination of the Bogused Vortex and the Dropwindsonde Data in DOTSTAR. *Mon. Wea. Rev.*, **136**, 865-879.

Colle, B. A., Dec. 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.

_____,_____, 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. Buonaiuto, 2009: New York City Storm Surges: Climatology and an Analysis of the wind and cyclone evolution. *J, Appl. Meteor. Climatol.* (In press)

Davis, C., W. Wang, S. S. Chen, Y. Chen, K. Corbosiero, M. DeMaria, J. Dudhia, G. Holland, J. Klemp, J. Michalakes, H. Reeves, R. Rotunno, C. Snyder and Q. Xiao, June, 2008: Prediction of Landfalling Hurricanes with the Advanced Hurricane WRF Model. *Mon. Wea. Rev.*, **136**, 1990-2005.

DeMaria, M., M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan, Aug. 2005: Further Improvements to the Statistical Hurricane Intensity Prediction Scheme (SHIPS). *Wea. Forecasting*, **20**, 531-543.

Dickey, T., D. Frye, J. McNeil, D. Manov, N. Nelson, D. Sigurdson, H. Jannasch, D. Siefel, T. Michaels, and R. Johnson, May 1998: Upper-Ocean Temperature Response to Hurricane Felix as Measured by the Bermuda Testbed Mooring. *Mon. Wea. Rev.*, **126**, 1195-1201.

Donelan, M. A., B. K. Haus, N. Reul, W. J. Plant, M. Stiassnie, H. C. Graber, O. B. Brown and E. S. Saltzman, 2004: On the limiting aerodynamic roughness of the ocean in very strong winds. *Geophys. Res. Letters*, **31**, L18306, doi:10.1029/2004GL019460.

Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077-3107.

Eckel, F. A, and C. F. Mass, June, 2005: Aspects of Effective Mesoscale, Short-Range Ensemble Forecasting. *Wea. Forecasting*, **20**, 328-350.

Farfán, L., and J. Zehnder. Aug. 2001: An Analysis of the Landfall of Hurricane Nora (1997). *Mon. Wea. Rev*., **129**, 2073-2088.

Ferrier, B., Jan. 1994: A Double-Moment Multiple-Phase Four-Class Bulk Ice Scheme. Part I: Description. *J. Atmos. Sci.*, **51**, 249-280.

_____,_____, Y. Jin, Y. Lin, T. Black, E. Rogers, and G. DiMego, 2002: Implementation of a new grid-scale cloud and precipitation scheme in the NCEP Eta model. *19th Conf. on weather Analysis and Forecasting/15th Con on Numerical Weather Prediction*.

Fovell, R. G., and H. Su, 2007: Impact of cloud microphysics on hurricane track forecasts. *Geophys. Res. Letters*, **34**, L24810, doi:10.1029/2007GL031723.

Frank, N. and S.A. Husain, June 1971: The Deadliest Tropical Cyclone in History?. *Bull. Amer. Meteor. Soc.,* **52**, 438-444.

Franklin, J. L., S. J. Lord, and F. D. Marks, Jr., 1988: Dropwindsonde and radar observations of the eye of Hurricane Gloria. *Mon. Wea. Rev.*, **116**, 1237-1244.

_____,_____, S. J. Lord, S. E. Feuer and F. D. Marks, Jr., Sept. 1993: The Kinematic Structure of Hurricane Gloria (1985) Determined from Nested Analyses of Dropwindsonde and Doppler Radar Data. *Mon. Wea. Rev.*, **121**, 2433-2451.

_____,_____, Feb. 2008: 2007 National Hurricane Center Forecast Verification Report. *NHC*, 54 pp.

Garratt, J. R., July, 1977: Review of Drag Coeffcients over Oceans and Continents. *Mon. Wea. Rev.*, **105**, 915-929.

Gerritsen, H., June 2005: What happened in 1953? The Big Flood in the Netherlands in retrospect. *Phil. Trans. R. Soc. A.,* **363**, 1271-1291, doi:10.1098/rsta.2005.1568

Glickman, T. S., 2000: Glossary of Meteorology. 2nd ed. *Amer. Meteor. Soc.*, 855 pp.

Grell, G. A., 1993: Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev.*, **121**, 764-787.

_____,_____, J. Dudhia, and D. R. Stauffer, June 1995: A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5). *NCAR Tech. Note NCAR/TN-398+STR*, 122 pp.

Hack, J. J., B. A. Boville, B. P. Briegleb, J. T. Kiehl, P. J. Rasch and D. L. Williamson, 1993: Description of the NCAR Community Climate Model (CCM2). *NCAR Tech. Note, NCAR/TN-382+STR,* National Center for Atmospheric Research, Boulder, CO., 108 pp.

Hamill, T. M., Mar. 2001: Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Mon. Wea. Rev.*, **129**, 550-560.

Harris, D. L., 1963: Characteristics of the Hurricane Storm Surge. *U.S. Department of Commerce, Weather Bureau, Technical Paper NO. 48.*, 140 pp.

Hasselmann, S., K. Hasselmann, E. Buer, P. A. E. M. Janssen, G. J. Komen, L. Bertotti, P. Lionello, A. Guillaume, V. C. Cardone, J. A. Greenwood, M. Reistad, L. Zambresky, and J. A. Ewing, 1988: The WAM model – A third generation ocean wave prediction model. *J. Phys Oceanogr.*, **18**, 1775-1810.

Hill, K. A. and G. M. Lackmann, Feb 2009: Analysis of Idealized Tropical Cyclone Simulations Using the Weather Research and Forecasting Model: Sensitivity to Turbulence Parameterization and Grid Spacing. *Mon. Wea. Rev.*, **137**, 745-765.

Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.*, **124**, 2322-2339.

_____,_____, J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice-microphysical processes for the bulk parameterization of cloud and precipitation. *Mon. Wea. Rev.*, **132**, 103-120.

_____,_____, Y. Noh, and J. Dudhia, Sept. 2006: A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes. *Mon. Wea. Rev*., **134**, 2318-2341.

Janjic, Z. I., 2002: Nonsingular Implementation of the Mellor-Yamada Level 2.5 Scheme in the NCEP Meso model. *NCEP Office Note*, **No. 437**, 61 pp.

_____,_____, T. Black, M. Pyle, B. Ferrier, H.-Y. Chuang, D. Jovic, R. Rozumalski, J. Michalakes, D. Gill, J. Dudhia, M. Duda, M. Demirtas, L. Nance, J. Wolff, L. Bernardet, P. McCaslin, and M. Stoelinga, 2008: WRF-NMM V3: User's Guide.

NOAA/NCEP and Developmental Testbed Center (DTC), Aug. 5th 2009 <http://www.dtcenter.org/wrf-nmm/users/docs/overview.php> .

Jarvinen, B., and J. Gebert, 1986: Comparison of Observed versus SLOSH Model Computed Storm Surge Hydrographs along the Delaware and New Jersey Shorelines for Hurricane Gloria, September 1985. *NOAA Technical Memorandum, NWS NHC 32*.

_____,_____, and _____,_____, 1987: Observed versus SLOSH Model Storm Surge for Connecticut, New York and Upper New Jersey in Hurricane Gloria, September 1985. *NOAA Technical Memorandum*, *NWS NHC 36.*

_____,_____, and A. McDuffie, 1987: Observed versus SLOSH Model Storm Surge for North Carolina in Hurricane Gloria September 1985. *NOAA Technical Memorandum, NWS NHC 37*.

Jelesnianski, C., Chen, J., and Shaffer, W., 1992: SLOSH: Sea, Lake, and Overland Surges from Hurricanes. *NOAA Technical Report*, NWS 48.

Jones, M., B. A. Colle, and J. Tongue, 2007: Evaluation of a short-range ensemble forecast system over the Northeast U.S.. *Wea. Forecasting*, **22**, 36-55.

Kain, J.S., 2002: The Kain-Fritsch convective parameterization: An update. *J. Appl. Meteor.*, **23**, 170-181.

Kistler, R., E. Kalnay, W. Collins, S. Saha, G. White, J. Woollen, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H. van den Dool, R. Jenne, and M. Fiorino, Feb. 2001: The NCEP-NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247-267.

Klein, P. M., P. A. Harr and R. L. Elsberry, Aug. 2000: Extratropical Transition of Western North Pacific Tropical Cyclones: An Overview and Conceptual Model of the Transformation Stage. *Wea. Forecasting*, **15**, 373-395.

Knabb, R, Rhome, R., and D. Brown. Aug. 2006: Tropical Cyclone Report: Hurricane Katrina 23-30 August 2005. *National Hurricane Center.*

Koch, S. E., B. Ferrier, M. T. Stoelinga, E. Szoke, S. J. Weiss, and J. S. Kain, 2005: The use of simulated radar reflectivity fields in the diagnosis of mesoscale phenomena from high-resolution WRF model forecasts. Preprints, *32nd Conf. on Radar Meteorology,* Albuquerque, NM, Amer. Meteor. Soc., J4J.7.

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

Kurihara, Y., M. A. Bender, R. E. Tuleya, and R. J. Ross, Oct. 1990: Prediction Experiments of Hurricane Gloria (1985) Using a Multiply Nested Movable Mesh Model. *Mon. Wea. Rev.*, **118**, 2185-2198.

_____,_____, R. E. Tuleya and M. A. Bender, May 1998: The GFDL Hurricane Prediction System and Its Performance in the 1995 Hurricane Season. *Mon. Wea. Rev.*, **126**, 1306-1322.

Large, W. G., and S. Pond, May 1982: Sensible and Latent Heat Flux Measurements over the Ocean. *J. Phys. Oceanogr.*, **12**, 464-482.

Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate. Appl. Meteor.*, **22**, 1065-1092.

Liu, Y., Dec. 1997: A Multiscale Numerical Study of Hurricane Andrew (1992). Part I: Explicit Simulation and Verification. *Mon. Wea. Rev.*, **125**, 3073-3093.

Low-Nam, S., and C. Davis, 2001: Development of a tropical cyclone bogusing scheme for the MM5 system. *Preprint, 11th PSN/NCAR Mesoscale Model Users' Workshop*, Boulder, CO, 130-134.

Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. Space Phys.*, **20**, 851-875.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jovic, J. Woollen, E. Rogers, E. H. Berbery, M. B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Li, G. Manikin, D. Parrish, and W. Shi, Mar. 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343-360.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validate correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16,663-16,682.

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

Nolan, D. S., D. P. Stern, and J. A. Zhang, 2009: Validation and Comparison of Planetary Boundary Layer Parameterizations in Tropical Cyclones by Comparison of in-situ Observations and High-Resolution Simulations of Hurricane Isabel (2003). Part II: Inner-core Boundary Layer and Eyewall Structure. In Press.

Pawlowicz, R., B. Beardsley, and S. Lentz, 2002: Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE. *Comp. Geosci.*, **28**, 929-937.

Reisner, J., R. M. Rasmussen, R. T. Bruintjes, 1998: Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. *Quart. J. Roy. Meteor. Soc.*, **124**, 1071-1107.

Resio, D. T., and J. J. Westerink, 2008: Hurricanes and the Physics of Surges. *Physics Today*, **61**, 33-38.

Ris, R. C., L. H. Holthuijsen and N. Booji, Apr. 1999: A third generation wave model for coastal regions 2. Verification. *J. Geophys. Res.*, **104**, 7667-7681.

Rutledge, S. A., and P. V. Hobbs, 1984: The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. XII: A diagnostic modeling study of precipitation development in narrow cloud-frontal rainbands. *J. Atmos. Sci.*, **20**, 2949-2972.

Signorini, J. S., J. S. Wei, and C. D. Miller, Feb. 1992: Hurricane-Induced Surge and Currents on the Texas-Louisiana Shelf. *J. Geophys. Res.*, **97**, 2229-2242.

Schade, L. R., and K. A. Emanuel, Feb. 1999: The Ocean's Effect on the Intensity of Tropical Cyclones: Results from a Simple Coupled Atmosphere-Ocean Model. *J. Atmos. Sci.*, **56**, 642-651.

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

Shaffer, W. A., J. P. Dallavalle, and L. D. Burroughs, 1997: East Coast Extratropical Storm Surge and Beach Erosion Guidance. *Technical Procedures Bulletin* No. 436.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang and J. G. Powers, 2005: A Description of the Advanced Research WRF Version 2. *NCAR Tech. note, NCAR/TN-468+STR*.

Shapiro, L. J. and J. L. Franklin, May, 1995: Potential Vorticity in Hurricane Gloria. *Mon. Wea. Rev.*, **123**, 1465-1475.

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.

Tao, W.-K., J. Simpson and M. McCumber, 1989: An ice-water saturation adjustment. *Mon. Wea. Rev.*, **117**, 231-235.

Tenerelli, J.E., and S.S. Chen, 2001: High-Resolution Simulation of Hurricane Floyd(1999) using MM5 with a vortex-following mesh refinement. *Preprints*, 14th Conference on Numerical Weather Prediction, 30 July- 2 August 2001. Ft. Lauderdale, FL. AMS.

Tilburg C.E., and R.W. Garvine, Jun. 2004: A Simple Model for Coastal Sea Level Prediction. *Wea. Forecasting*, **19**. 511-519.

Troen, I. and L. Mahrt, 1986: A simple model of the atmospheric boundary layer: Sensitivity to surface evaporation. *Boundary-Layer Meteor.*, **37**, 129-148.

Wang, Y., 1998: On the Bogusing of Tropical Cyclones in Numerical Models: The Influence of Vertical Structure. *Meteor. Atmos. Phys.*, **65**, 153-170.

Weisberg, R. H., and L. Zheng. Dec. 2006: Hurricane Storm Surge Simulations for Tampa Bay. *Est. and Coasts*, **29**, 899-913.

_____,_____, and _____,_____, Dec. 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, Jr. and N.W. Scheffner, 1993: ADCIRC: an advanced three-dimensional circulation model for shelves coasts and estuaries, report 3: development of a tidal constituent data base for the Western North Atlantic and Gulf of Mexico. *Dredging Research Program Technical Report DRP-92-6*, U.S. Army Engineers Waterways Experiment Station, Vicksburg, MS, 154p.

_____,_____, R. A. Luettich, J. C. Feyen, F. H. Atkinson, C. Dawson, H. J. Roberts, M. D. Powell, J. P. Dunion, E. J. Kubatko and H. Pourtaheri, Mar. 2008: A basin to channel scale unstructured grid hurricane storm surge model as implemented for southern Louisiana. *Mon. Wea. Rev.*, **136,** 833-864.

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

Wornom, S. D., D. J. S. Welsh and K. W. Bedford, Sep. 2001: On coupling the SWAN and WAM wave models for accurate nearshore wave predictions. *Coast. Engrg.*, **43**, 161-201.

Wu, C.C., and T.H. Yen. Oct. 2002: Rainfall Associated with Typhoon Herb (1996) near Taiwan. Part I: The Topographic Effect. *Wea. Forecasting,* **17**, 1001-1015.

Wu, J., 1982: Wind-stress coefficients over sea surface from breeze to hurricane. *J. Geophys. Res.*, **87**, C12, 9704-9706.

Xiao, Q., X. Zou and B. Wang, July 2000: Initialization and Simulation of a Landfalling Hurricane Using a Variational Bogus Data Assimilation Scheme. *Mon. Wea. Rev.*, **128**, 2252-2269.

Xu F., B. Toulany, and P.C. Smith. 2007: Wind-generated waves in Hurricane Juan. *Ocean. Mod.* **16**, 188-205.

Yau, M. K., Y, Liu, D.-L. Zhang and Y. Chen, Jun. 2004: A Multiscale Numerical Study of Hurricane Andrew (1992). Part VI: Small-Scale Inner-Core Structures and Wind Streaks. *Mon. Wea. Rev.*, **132**, 1410-1433.

Yu, T.-W., and V. M. Gerald. 2004: Evaluation of NCEP Operational Model Forecasts of Surface Wind and Pressure Fields over the Oceans. *Preprint, 20th Conference on Weather Analysis and Foreacasting/16th Conference on Numerical Weather Prediction*., Seattle WA., 7 pp.

Zhang, D., and R. A. Anthes, Nov. 1982: A High-Resolution Model of the Planetary Boundary Layer – Sensitivity Tests and Comparisons with SESAME-79 Data. *J. Appl. Meteor.*, **21**, 1594-1609.

Zhu, T., and D.-L., Zhang, Jan. 2006: Numerical Simulation of Hurricane Bonnie (1998). Part II: Sensitivity to Varying Cloud Microphysical Processes. *J. Atmos. Sci.*, **63**, 109-126.

Zilkoski, D. B., J. H. Richards, and G. M. Young, 1992: Results of the General Adjustment of the North American Vertical Datum of 1988. *Surveying and Land Information Systems*, **52**, 133-149.

Zwiers, F. W., 1990: The effect of serial correlation on statistical inferences made with resampling procedures. *J. Clim.*, **3**, 1452-1461.

Appendix

Tables

	Storm Surge Forecasting Systems		
Institution	**Atmospheric Forcing**	**Ocean Model**	**Start Time**
Stony Brook	5 MM5 / 3 WRF members	ADCIRC	0000 UTC
Stevens Institute of Technology	NCEP - NAM model	Princeton Ocean Model	12:00 AM
NOAA	NCEP - GFS model	NOAA Extratropical Storm Surge model	0000 UTC

Table 2.1. Description of three storm surge forecasting systems for the NYC-LI region.

Stony Brook Storm Surge Model Ensemble Members						
Members	**Model**	**Microphysics**	**PBL Scheme**	**Radiation**	**Initial Condition**	**Cumulus**
9a	MM5	Simple Ice	MRF	Cloud Radiation	WRF-NMM	Grell
BMMY	MM5	Simple Ice	MY	CCM2	GFS	Betts Miller
GRBLK	MM5	Simple Ice	Blackadar	CCM2	NOGAPS	Grell
K2MRF	MM5	Reisner	MRF	Cloud Radiation	GFS	Kain Fritsch
K2MY	MM5	Simple Ice	MY	CCM2	Canadian Model	Kain Fritsch
221	WRF	Ferrier	YSU	RRTM	WRF-NMM	Kain Fritsch
GFS	WRF	Ferrier	YSU	RRTM	GFS model	Grell
NOGAPS	WRF	WSM3	YSU	RRTM	NOGAPS	Betts Miller

Table 2.2. SBSS ensemble model members description including model used, microphysical schemes, PBL schemes, radiation schemes, initial conditions and cumulus schemes.

		Hurricane Gloria Simulations			
Members	**Model**	**Microphysics**	**PBL Scheme**	**Initial Condition**	**Cumulus**
CTL	**WRF - 4 km**	**Lin**	**GFS**	**NARR**	**None**
CTL-12	**WRF - 12 km**	**Lin**	**GFS**	**NARR**	**Kain Fritsch**
GFS-FERR	**WRF - 4 km**	**Ferrier**	**GFS**	**NARR**	**None**
YSU-LIN	**WRF - 4 km**	**Lin**	**YSU**	**NARR**	**None**
YSU-FERR	**WRF - 4 km**	**Ferrier**	**YSU**	**NARR**	**None**
MYJ-LIN	**WRF - 4 km**	**Lin**	**MYJ**	**NARR**	**None**

Table 3.1. List of the hurricane Gloria simulations and the different microphysics and PBL schemes used by each simulation.

Days With Full 8 member SBSS Ensemble		
Year	**Month**	**Day**
2007	11	1
2007	11	2
2007	11	3
2007	11	4
2007	11	6
2007	11	14
2007	11	15
2007	12	17
2007	12	18
2007	12	19
2007	12	21
2007	12	27
2007	12	28
2007	12	29
2008	1	1
2008	1	5
2008	1	6
2008	1	12
2008	1	13
2008	1	14
2008	1	20
2008	1	21
2008	1	22
2008	1	23
2008	1	24
2008	1	25
2008	1	26
2008	1	27
2008	1	29
2008	10	4
2008	10	5
2008	10	6
2008	10	22
2008	10	23
2008	10	24
2008	10	25
2008	10	28
2008	10	29
2008	10	30
2008	11	1
2008	11	3
2008	11	4
2008	11	5

2008	11	16
2008	11	17
2008	11	20
2008	11	21
2008	11	22
2008	11	23
2008	11	24
2008	11	25
2008	11	26
2008	11	27
2008	11	28
2008	11	29
2008	11	30
2008	12	2
2008	12	3
2008	12	4
2008	12	5
2008	12	7
2008	12	9
2008	12	14
2008	12	16
2008	12	17
2008	12	18
2008	12	20
2008	12	21
2008	12	22
2008	12	23
2008	12	24
2008	12	25
2008	12	27
2008	12	29
2008	12	30

Table 4.1. List of forecast days that fulfilled the requirement of having a full 8 member SBSS ensemble.

Brier Score Components					
		REL	RES	UNC	BS
> 0.3 m	SBSS	0.0053	0.018	0.069	0.056
	ALL	0.0099	0.029	0.069	0.05
	BC-ALL	0.0075	0.039	0.069	0.038
	ENS-3	0.0027	0.027	0.069	0.045
	BC-ENS-3	0.0063	0.039	0.069	0.036
> 0.4 m	SBSS	0.0031	0.0098	0.035	0.028
	ALL	0.0044	0.013	0.035	0.026
	BC-ALL	0.0048	0.02	0.035	0.02
	ENS-3	0.00066	0.01	0.035	0.026
	BC-ENS-3	0.0052	0.017	0.035	0.023

Table 4.2. Brier score components: reliability (REL), resolution (REL), uncertainty (UNC) and brier score (BS) for four models: SBSS, ALL, BC-ALL, ENS-3 and BC-ENS-3 ensemble for two surge thresholds > 0.3 m and > 0.4 m surge.

Figures

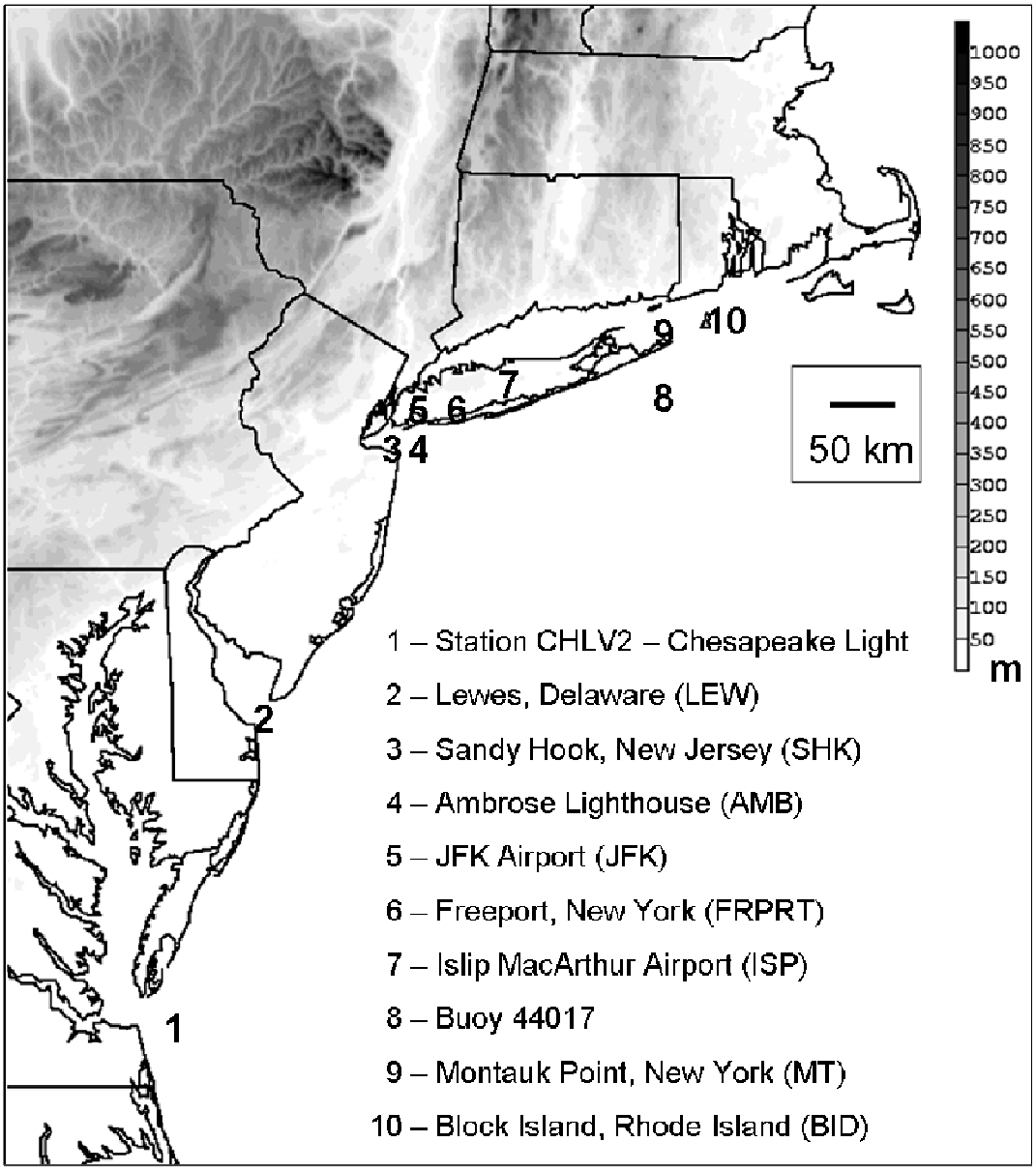


Figure 1.1 Map of the area of interest showing the locations of the observational data points and terrain (shaded, m).

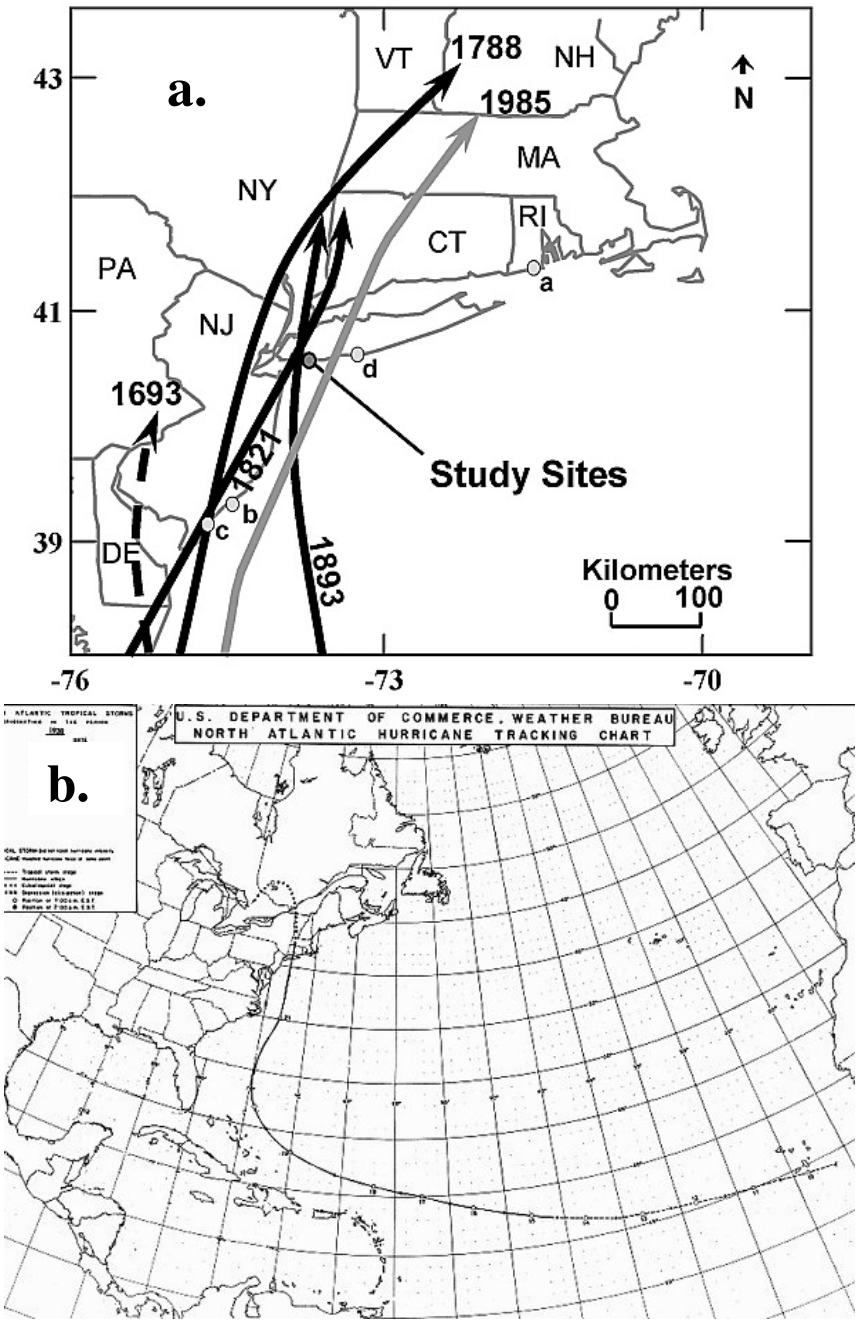


Figure 1.2. (a) Hurricane tracks that impacted western Long Island (Scileppi and Donnelly 2007). The a, b, and c correspond to Succotash Marsh (RI), Brigantine Marsh (NJ), and Whale Beach Marsh (NJ), respectively. Site d is the location of the Great South Bay sea level data (see Scileppi and Donnelly 2007 for more details). (b) National Hurricane Center track of the Long Island Express hurricane of 1938 that made landfall across the east end of Long Island at 2030 UTC 21 September 1938.

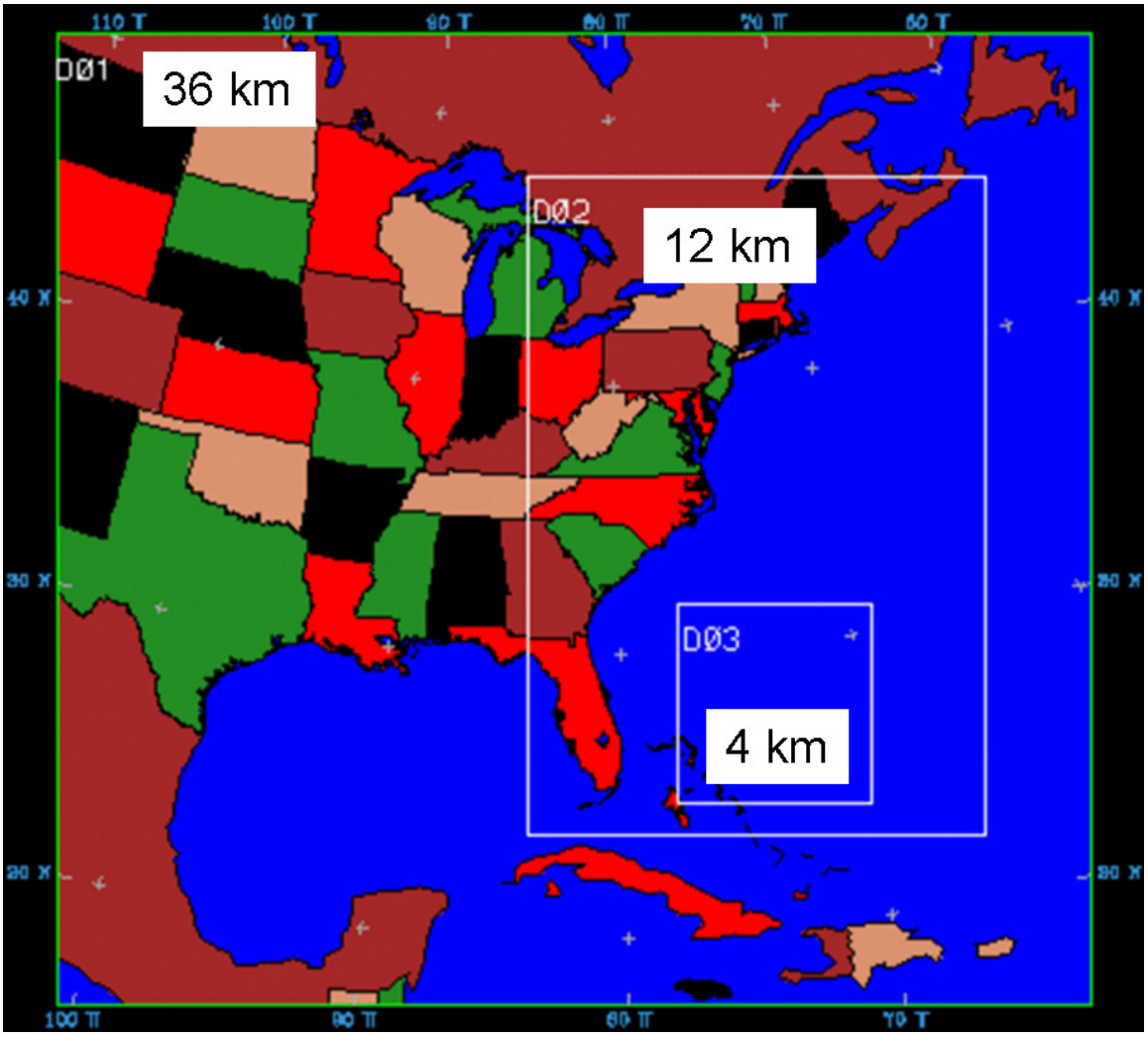
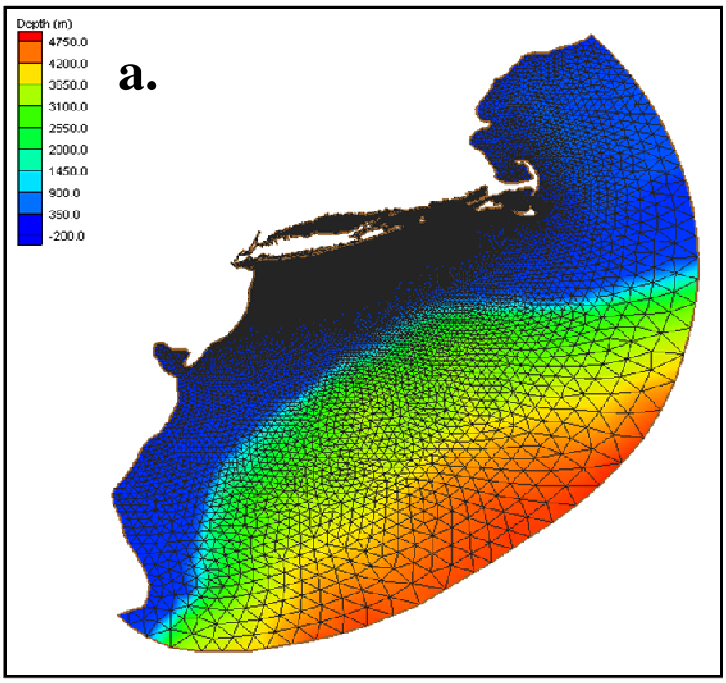


Figure 2.1. Domains used in the atmospheric simulations of hurricane Gloria. Domain 1 has 36 km grid spacing, Domain 2 has 12 km grid spacing. Domain 3 has 4 km grid spacing and is locked onto the vortex and follows it along the eastern seaboard.



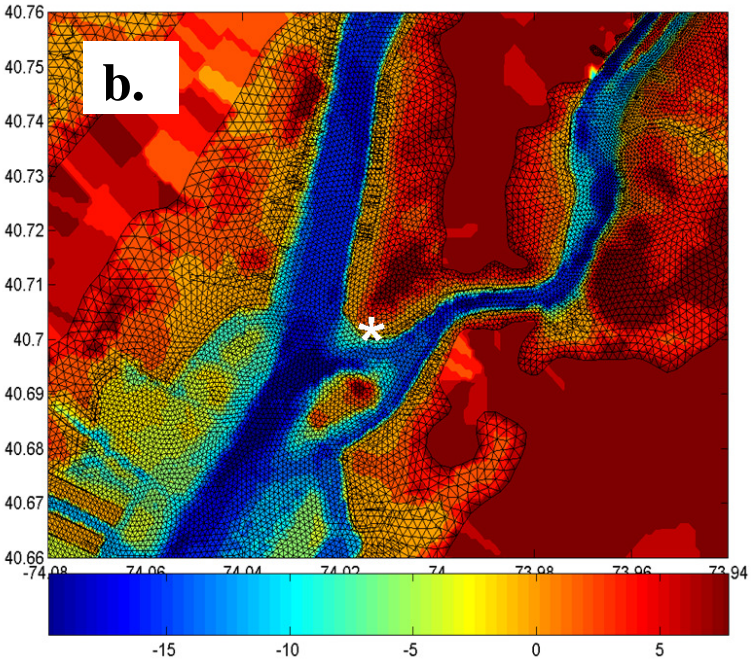


Figure 2.2. (a) ADCIRC domain for hurricane Gloria simulations and SBSS real-time forecasting ensemble. The grid has 108096 nodes ranging in size from 70 km over the open ocean to 5 m in the inner bays around New York. (b) Grid resolution inside New York Harbor in the ADCIRC domain. The Battery is starred. Grid spacing is on the order of meters within the narrow channels in the Harbor.

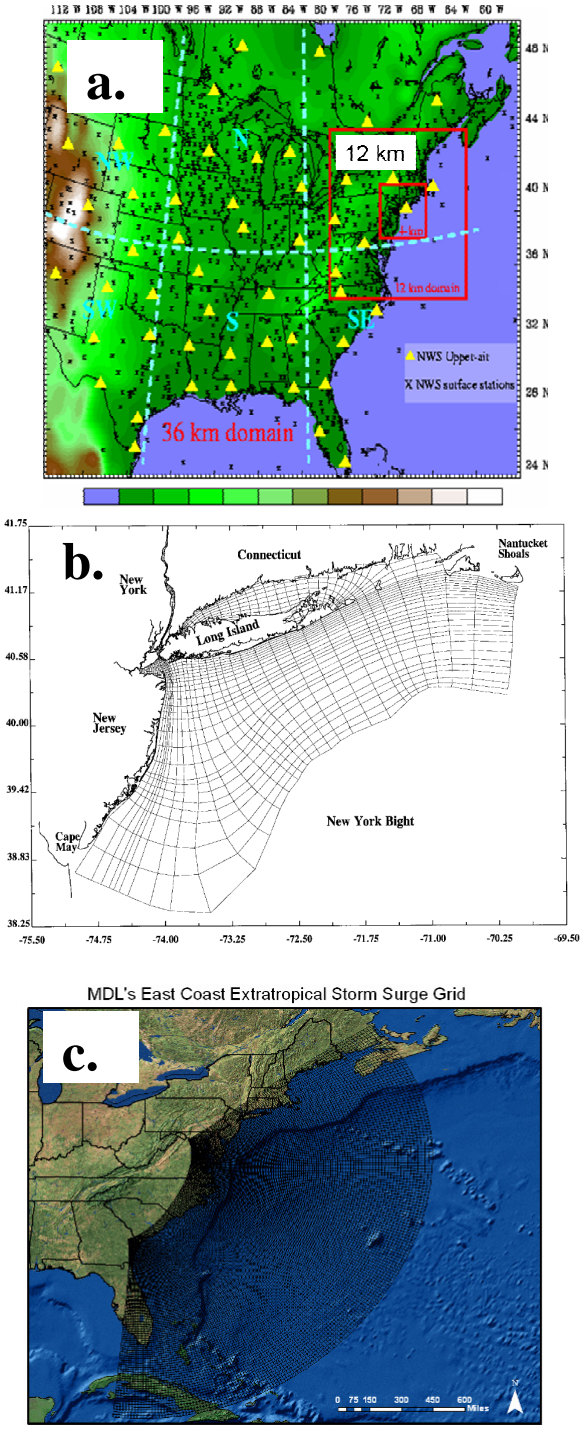


Figure 2.3. (a) Atmospheric domains used in the SBSS real-time storm surge forecasting ensemble. The 10 m winds and SLP from domain 2, 12 km grid spacing is used in the SBSS system. (b) The ocean grid used by Stevens Institute of Technology taken from Blumberg et al. (1999) (c) The NOAA Extratropical Storm Surge model east coast grid

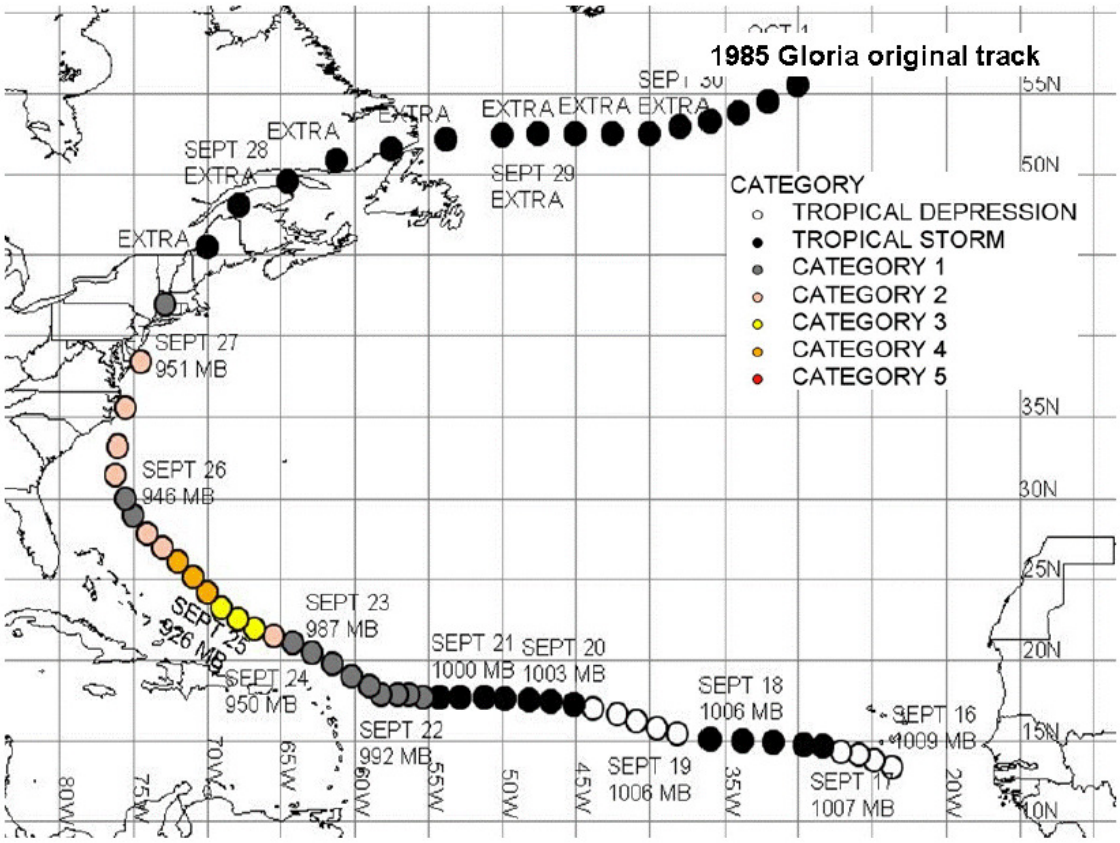


Figure 3.1. Hurricane Gloria track and intensity. Taken from Landsea et al. Tropical Cyclone Reanalysis Project.

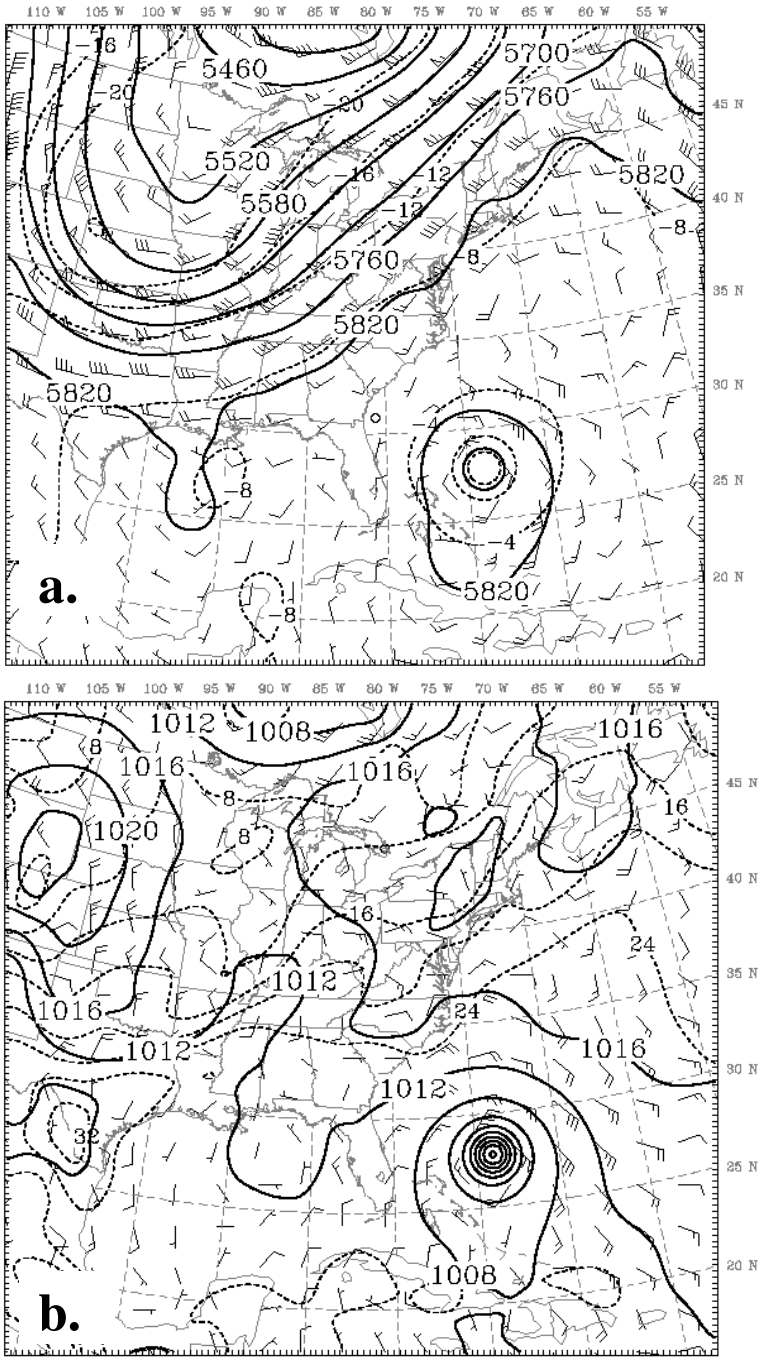


Figure 3.2. (a) 0000 UTC 26 September 1985 500 hPa geopotential heights, m, (contoured every 60 m) and temperature, °C, (dashed every 4 °C) for the CTL simulation 36 km grid. (b) 0000 UTC 26 September 1985 surface map with SLP, hPa, (contoured every 4 hPa) and temperature, °C, (dashed every 4 °C).

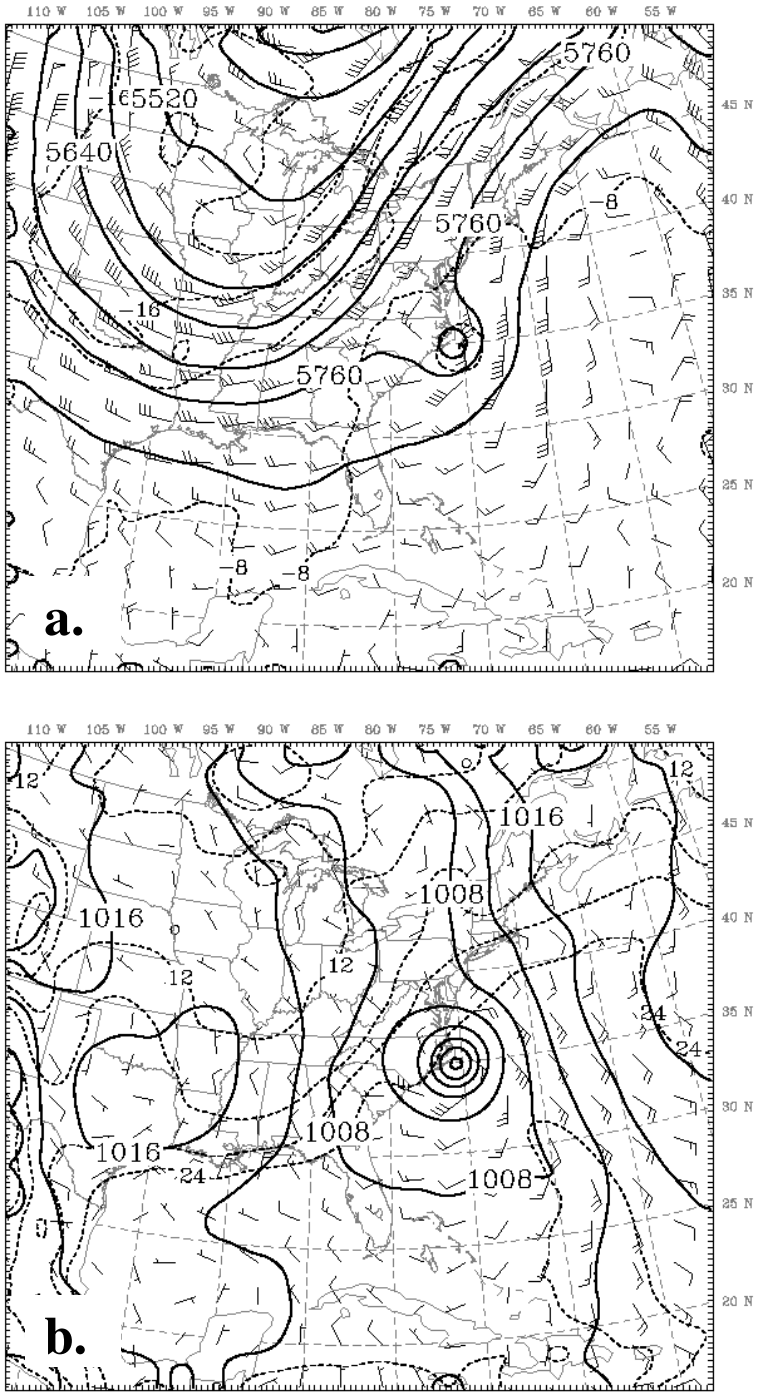


Figure 3.3. (a) 0600 UTC 27 September 500 hPa geopotential heights, m, (contoured every 60 m) with temperature, °C, (dashed every 4 °C) for the CTL 36 km grid. (b) 0600 UTC 27 September surface map with SLP, hPa, (contoured every 4 hPa) and temperature, °C, (dashed every 4 °C).

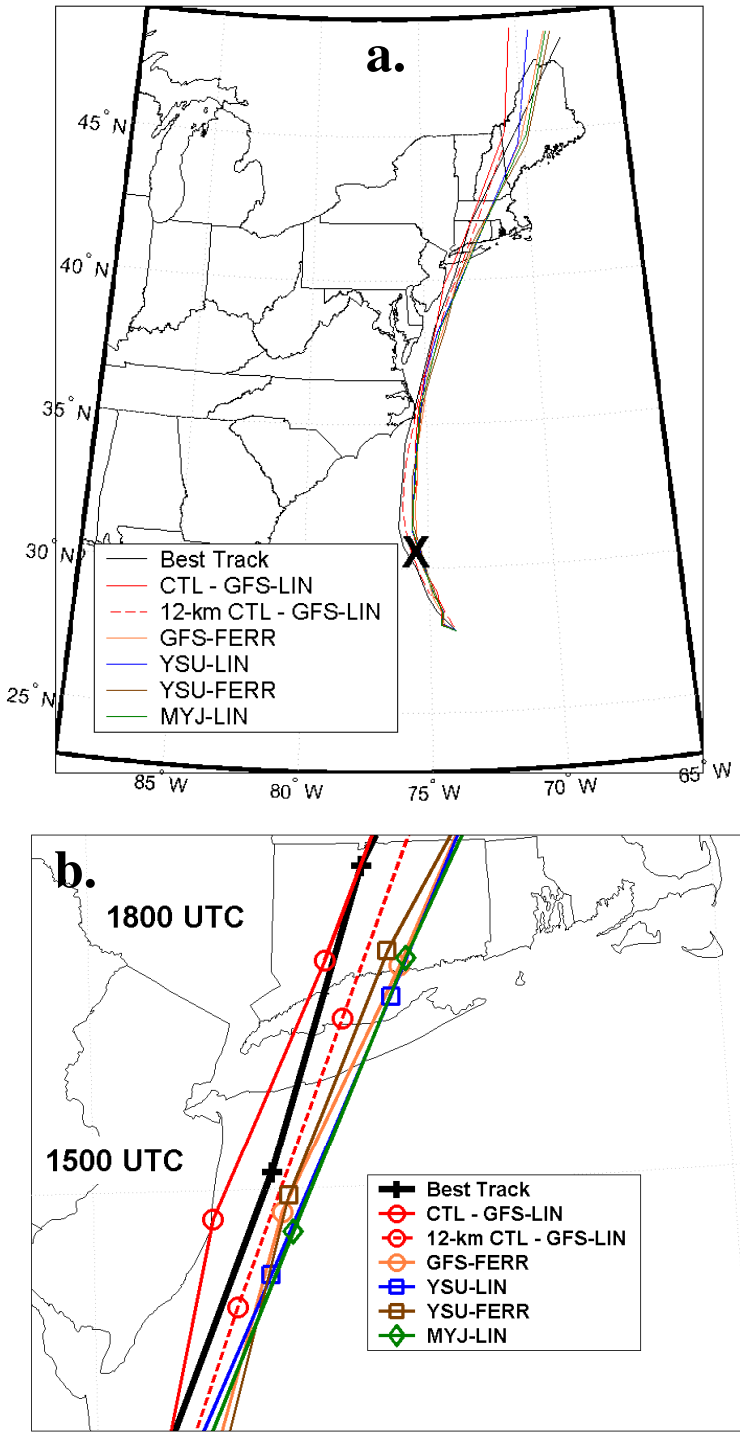


Figure 3.4. (a) Hurricane Gloria simulated tracks starting at initialization (0000 UTC 26 September 1985) through 0600 UTC 28 September 1985. The X marks 0900 UTC 26 September 1985. (b) Hurricane Gloria simulated tracks before landfall (1500 – 1800 UTC 27 September 1985).

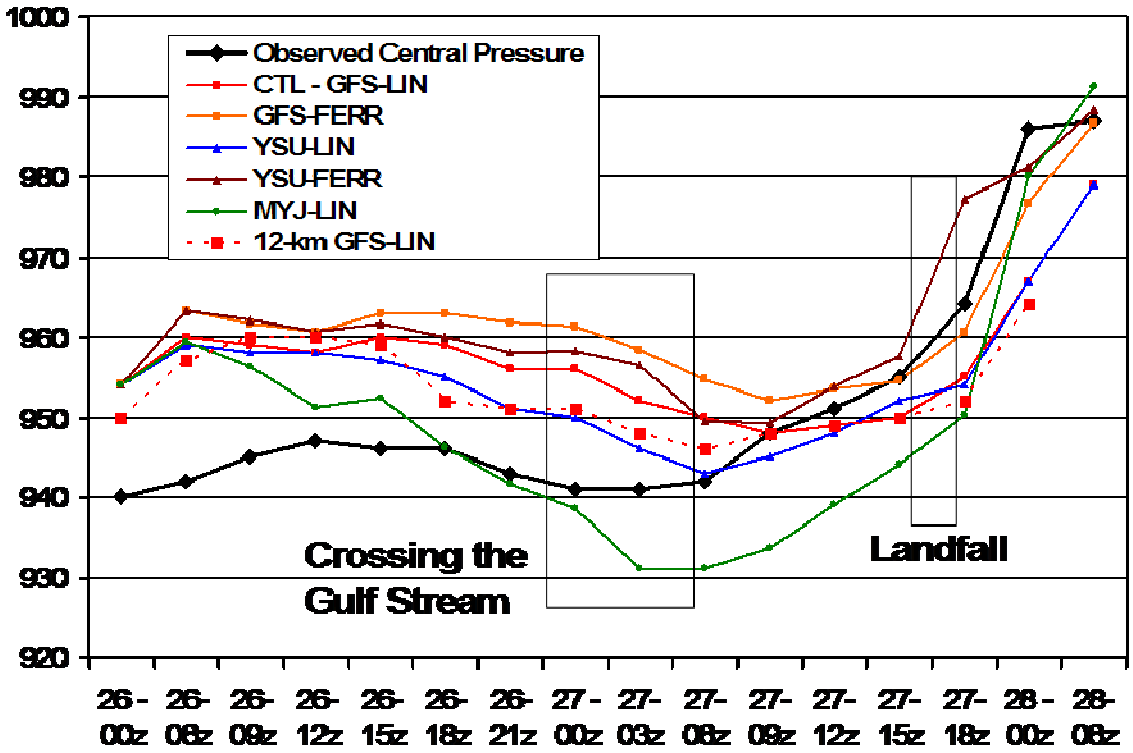


Figure 3.5. Hurricane Gloria simulations central pressure (hPa) time evolution. Specific time frames when the simulations crossed the Gulf Stream and made landfall across Long Island, NY are noted by the boxes.

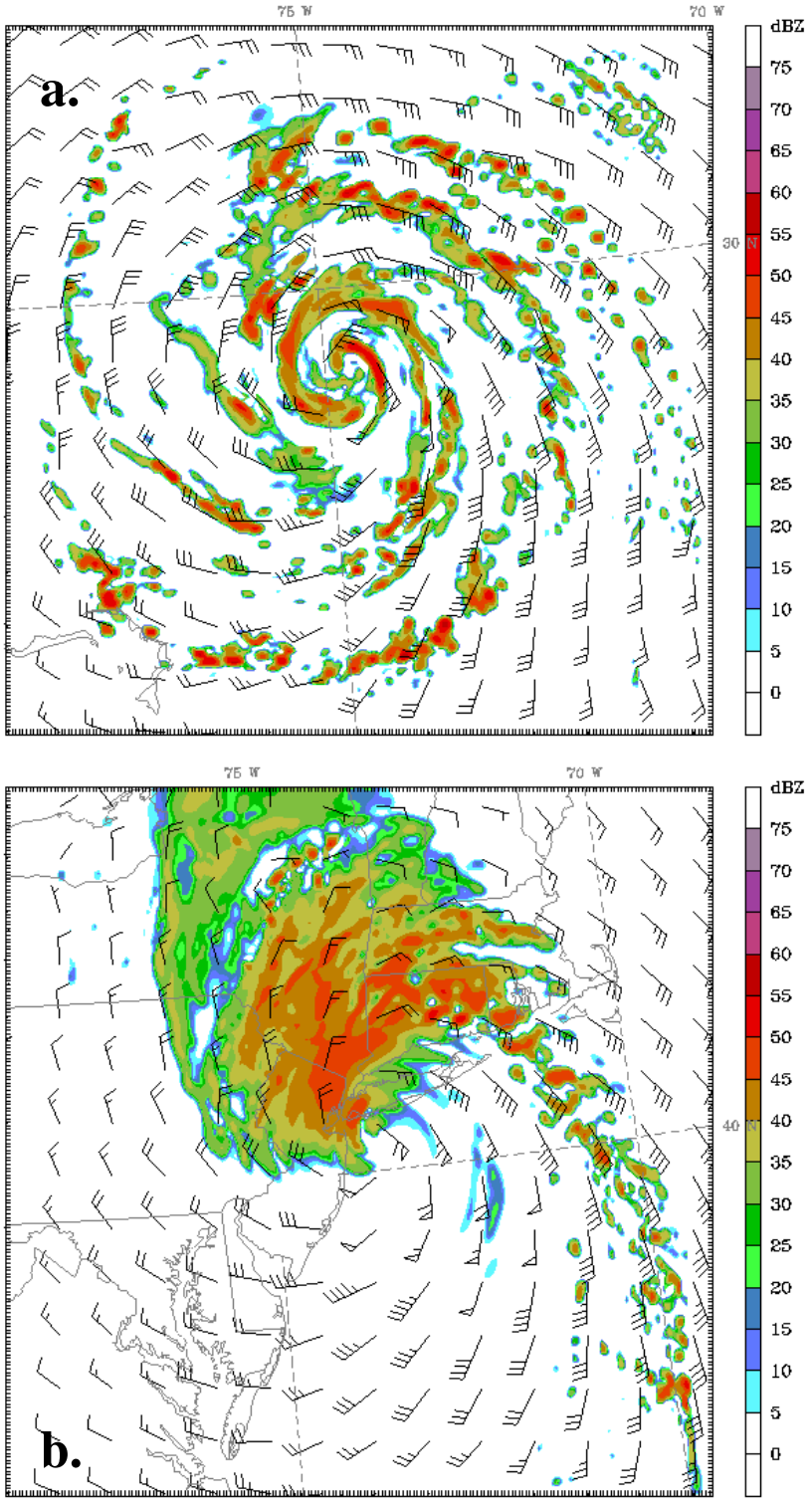


Figure 3.6. (a) Modeled derived reflectivities (see Koch et al. 2005 for procedure) and 10 m wind barbs (full barb = 10 kts) for the CTL simulation at 0900 UTC 26 September 1985. (b) Same as (a) except at 1600 UTC 27 September 1985.

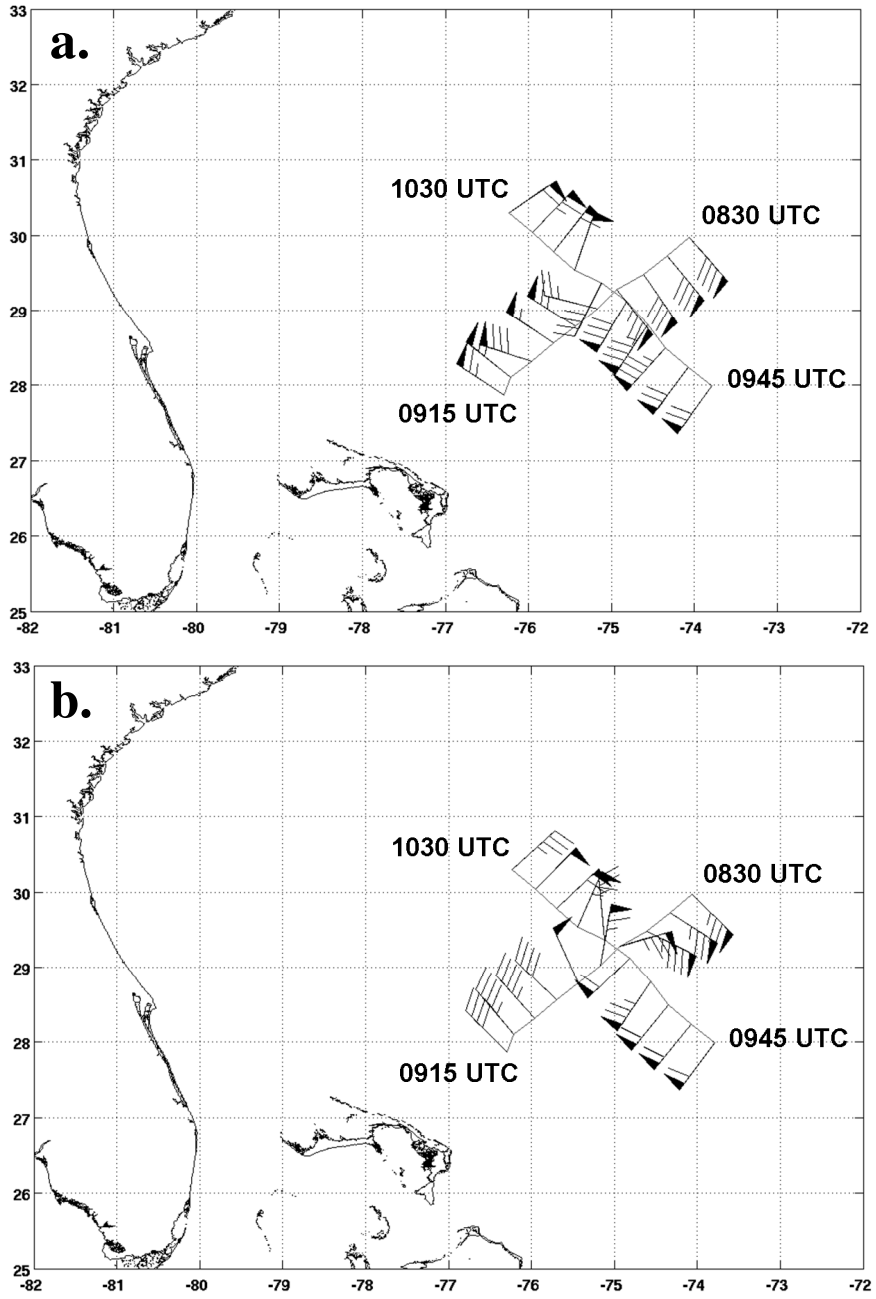


Figure 3.7. (a) Observed NOAA P-3 aircraft flight level winds (kts) at 2.8 km ASL for two transects through the eye of hurricane Gloria, 0830 – 0915 UTC and 0945 – 1030 UTC 26 September 1985, with winds (full barb = 10 kts) reported ~ 5 minutes. (b) Same as (a) except for the CTL simulation. CTL was outputted every 15 minutes and times closest to the 5 minute interval used in (a) were used for comparison, e.g. 0845 UTC used for times 0840, 0845 and 0850 UTC.

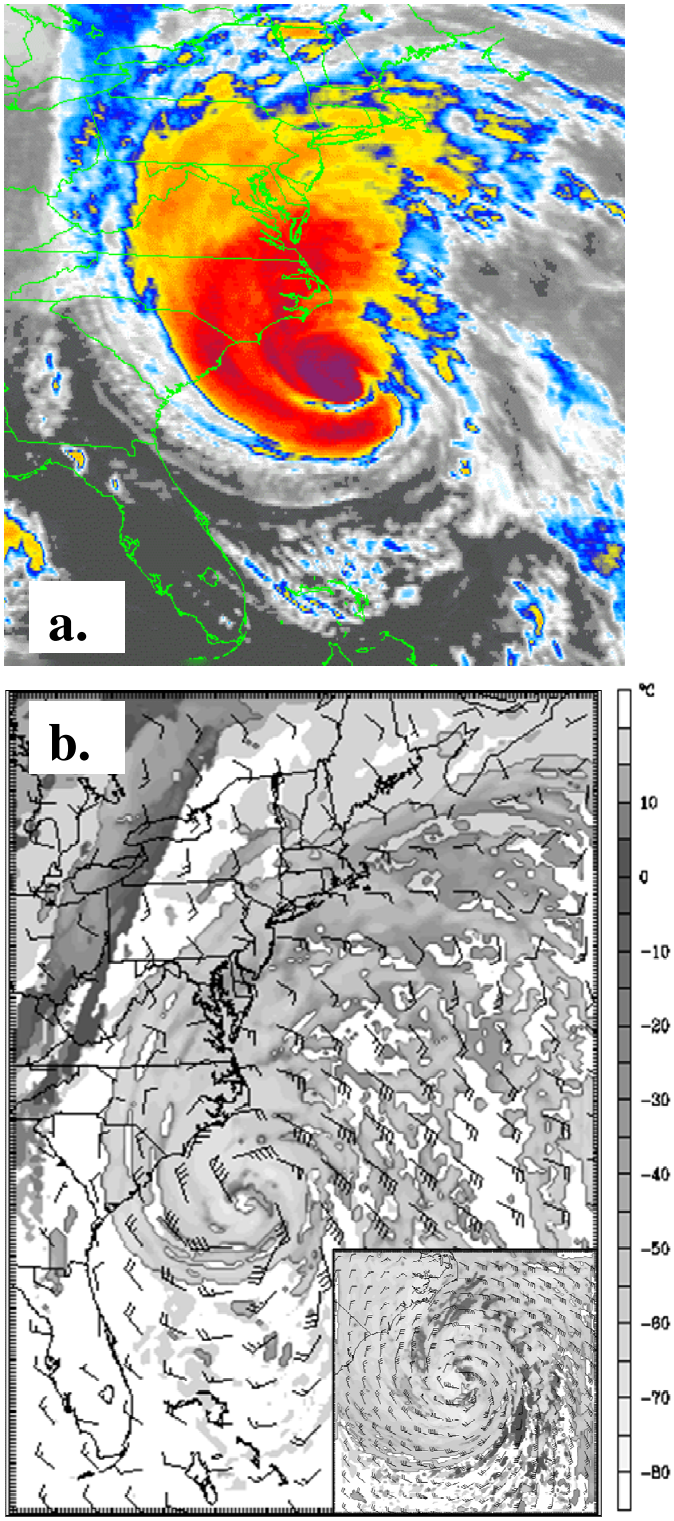


Figure 3.8. (a) GOES-6 infrared satellite imagery on 2200 UTC 26 September 1985 (red = coldest cloud tops). (b) CTL simulation cloud top temperatures, °C, and 10 m winds (full barb = 10 kts) at 2200 UTC 26 September 1985 for the 12 km domain and 4 km domain (inset).

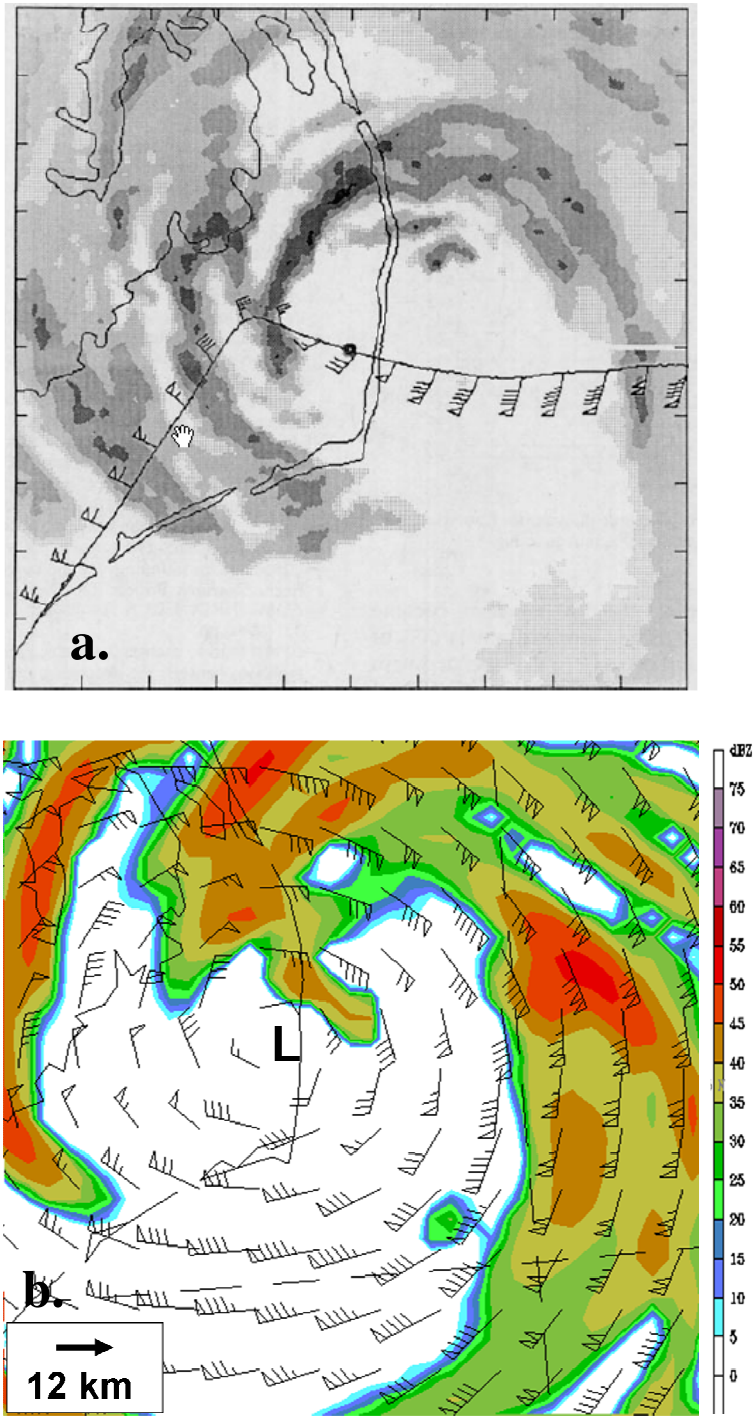


Figure 3.9. (a) Radar reflectivity taken from a NOAA P-3 Aircraft at 2.7 km at 0538 UTC 27 September 1985. Reflectivity contours of 21, 23, 27, 30, 34, and 38 DBZ are shown by increasing shades of gray. Flight level winds (full barb = 5 m s^{-1}) at 1 min intervals are plotted and tick marks indicate a distance of 12 km (as taken from Franklin, Lords and Marks Jr. 1988). (b) CTL model derived radar reflectivity (see Koch et al. 2005) at 0530 UTC 27 September 1985. Winds at 2.7 km are plotted (full barb = 5 m s^{-1}).

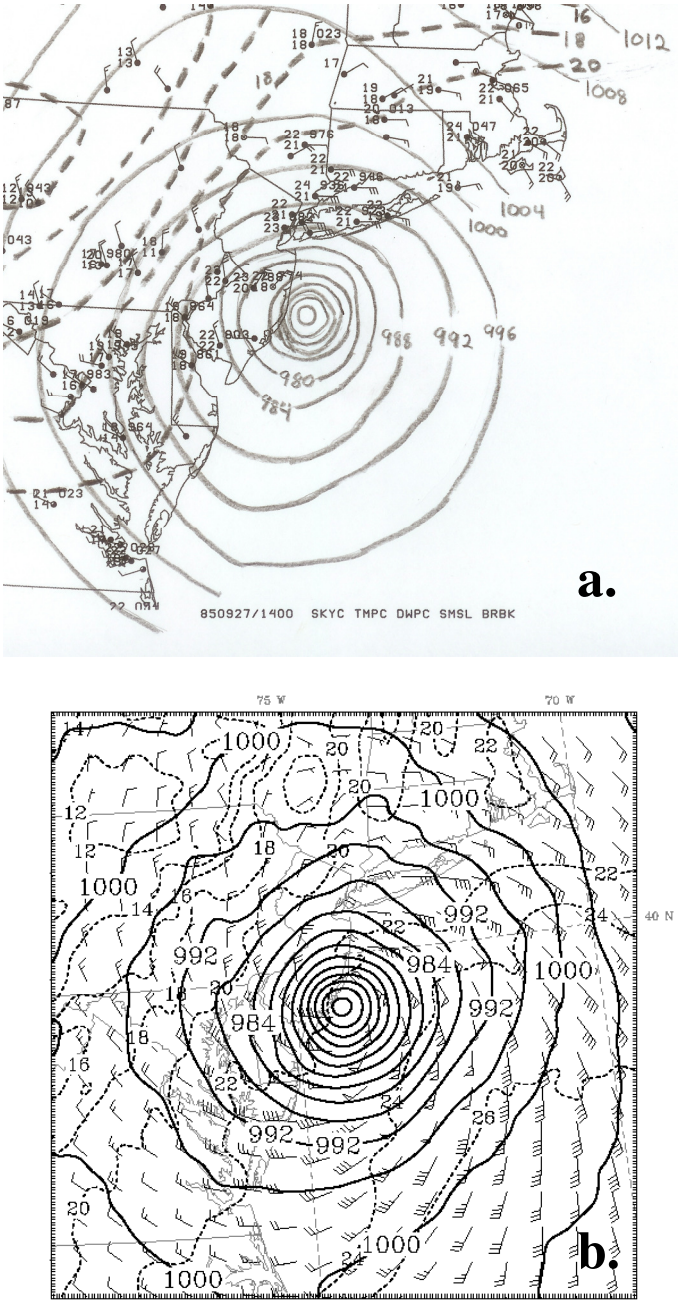


Figure 3.10. (a) Surface observations analysis at 1400 UTC 27 September 1985 using archived station plot data. SLP (black contours) and temperature, °C, (black dashed) are contoured at intervals of 4 hPa and 4 °C respectively. SLP over the ocean are determined from NOAA P-3 aircraft around the time of 1400 UTC as the plane penetrated the eye of hurricane Gloria (central pressure from plane: 959 hPa). (b) CTL simulation surface analysis at 1400 UTC 27 September 1985 with SLP (black contours) plotted every 4 hPa, temperature (dashed) plotted every 2°C and 10 m surface winds (full barb = 10 kts) plotted.

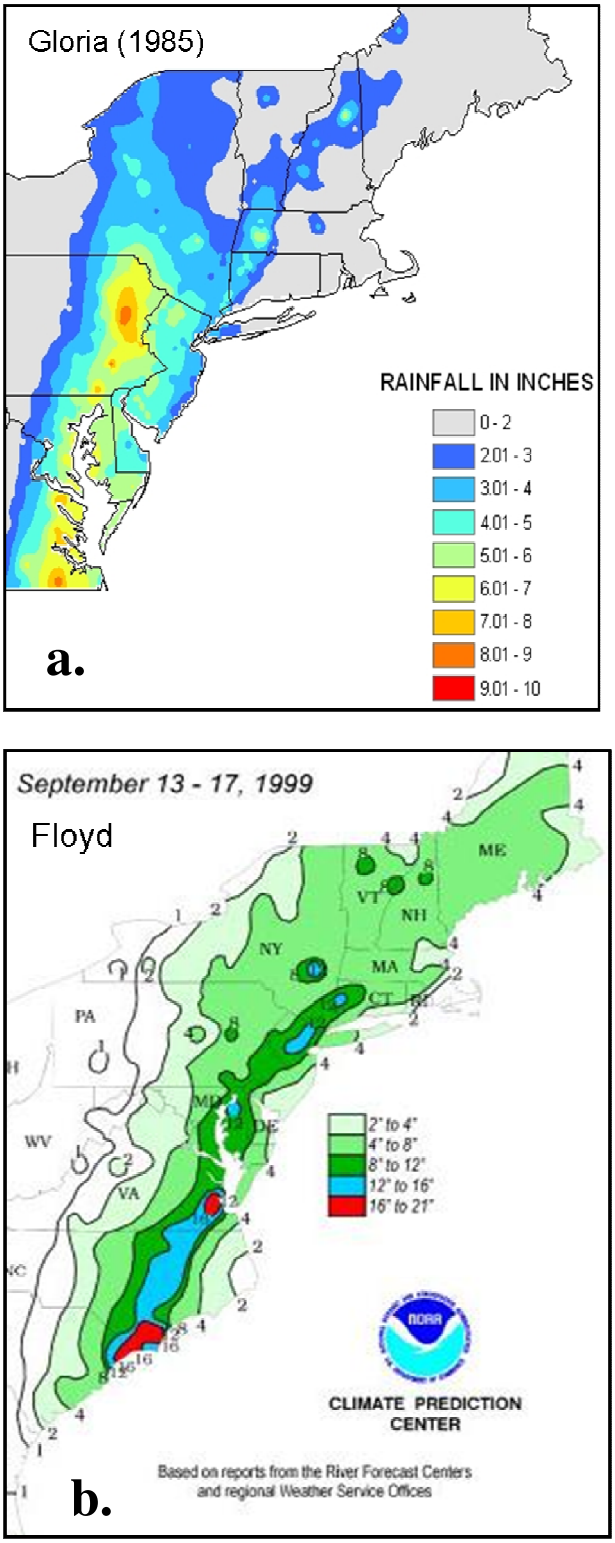


Figure 3.11. (a) Hurricane Gloria storm total rainfall in inches. (b) Hurricane Floyd storm total rainfall in inches.

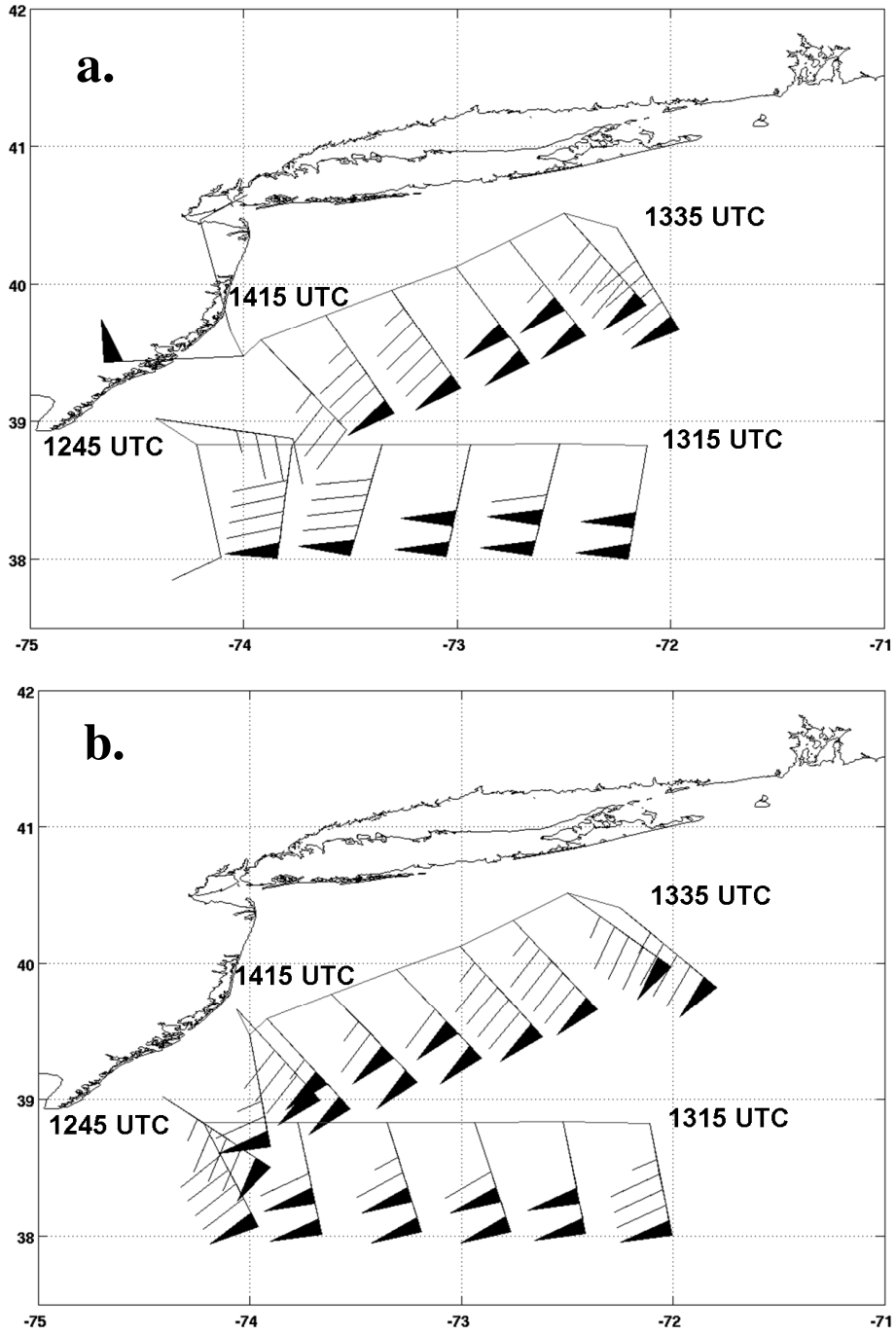


Figure 3.12. (a) Observed NOAA P-3 aircraft flight level winds (kts) at 1.5 km ASL for two transects through the eye of hurricane Gloria, 1245 – 1315 UTC and 1335 – 1415 UTC 27 September 1985, with winds (full barb = 10 kts) reported ~ 5 minutes. (b) Same as (a) except for the CTL simulation. CTL was outputted every 15 minutes and times closest to the 5 minute interval used in (a) were used for comparison, e.g. 1300 UTC used for times 1255, 1300 and 1305 UTC.

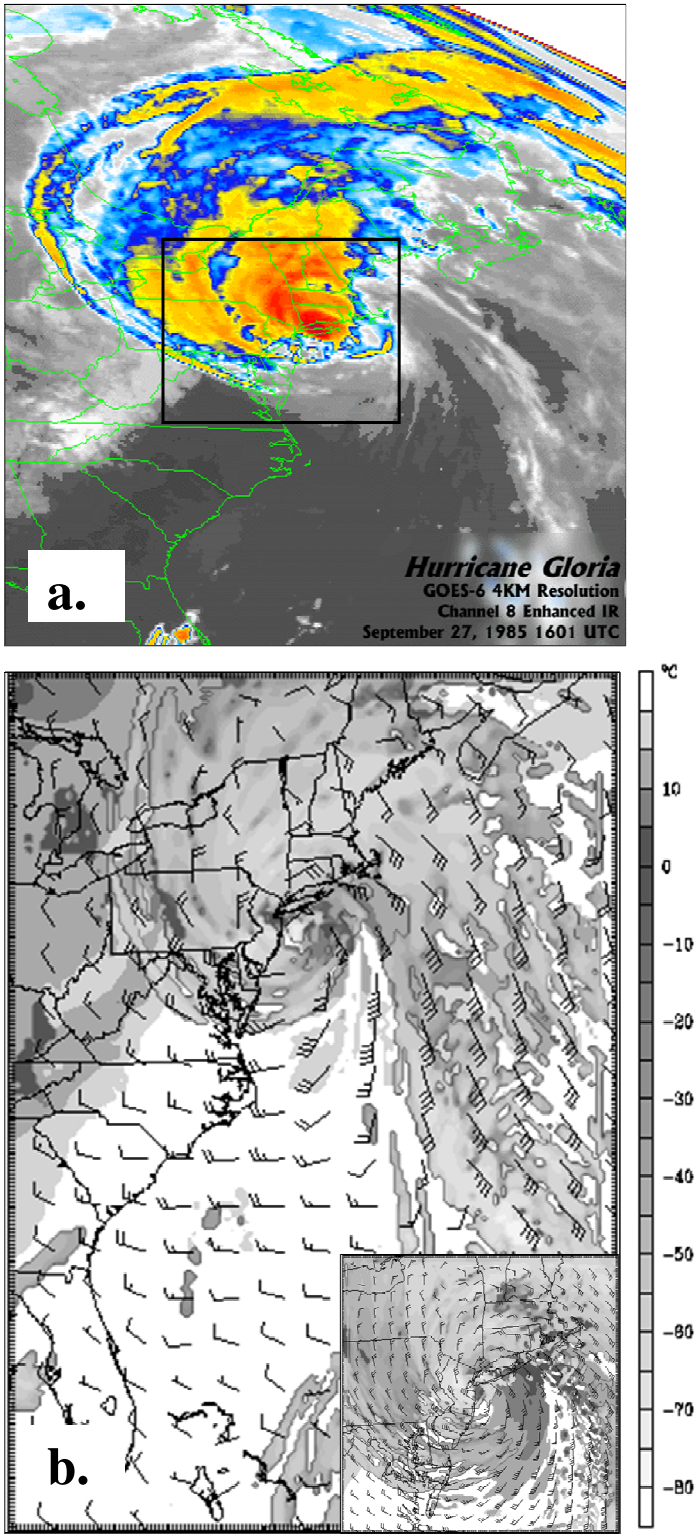


Figure 3.13. (a) GOES-6 4 km resolution infrared satellite imagery at 1601 UTC 27 September 1985. Box indicates the location of the CTL 4 km domain at 1600 UTC. (b) CTL simulation cloud top temperatures, °C, and 10 m winds (full barb = 10 kts) at 1600 UTC 27 September 1985 for the 12 km domain and 4 km domain (inset).

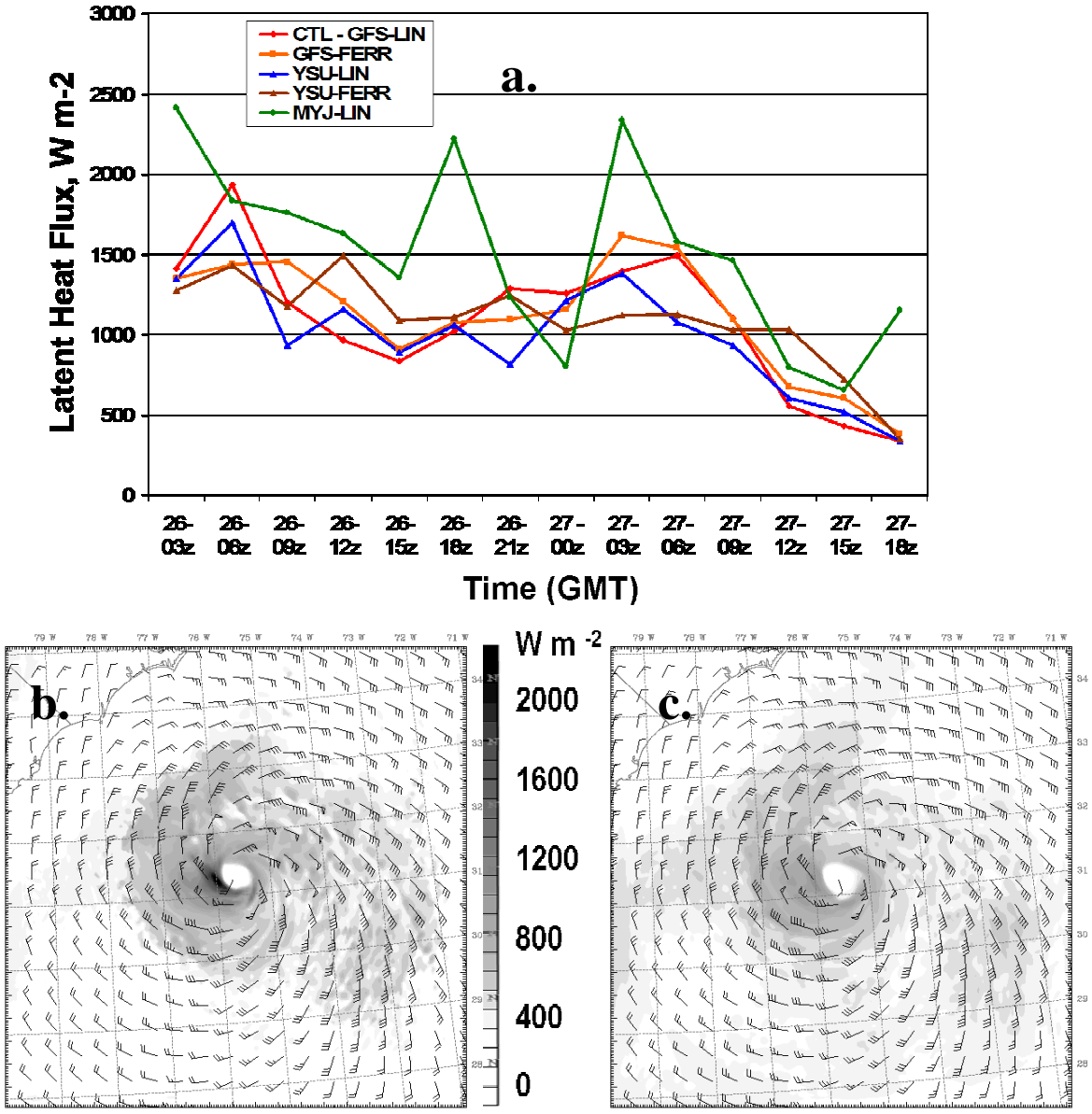


Figure 3.14. (a) Time series of latent heat flux (W m^{-2}) for all simulations every three hours from 0300 UTC 26 September 1985 – 1800 UTC 27 September 1985. (b) Latent heat flux (W m^{-2}) at 1800 UTC 26 September 1985 for the MYJ-LIN simulation. (c) Same as (b) except for the YSU-LIN simulation.

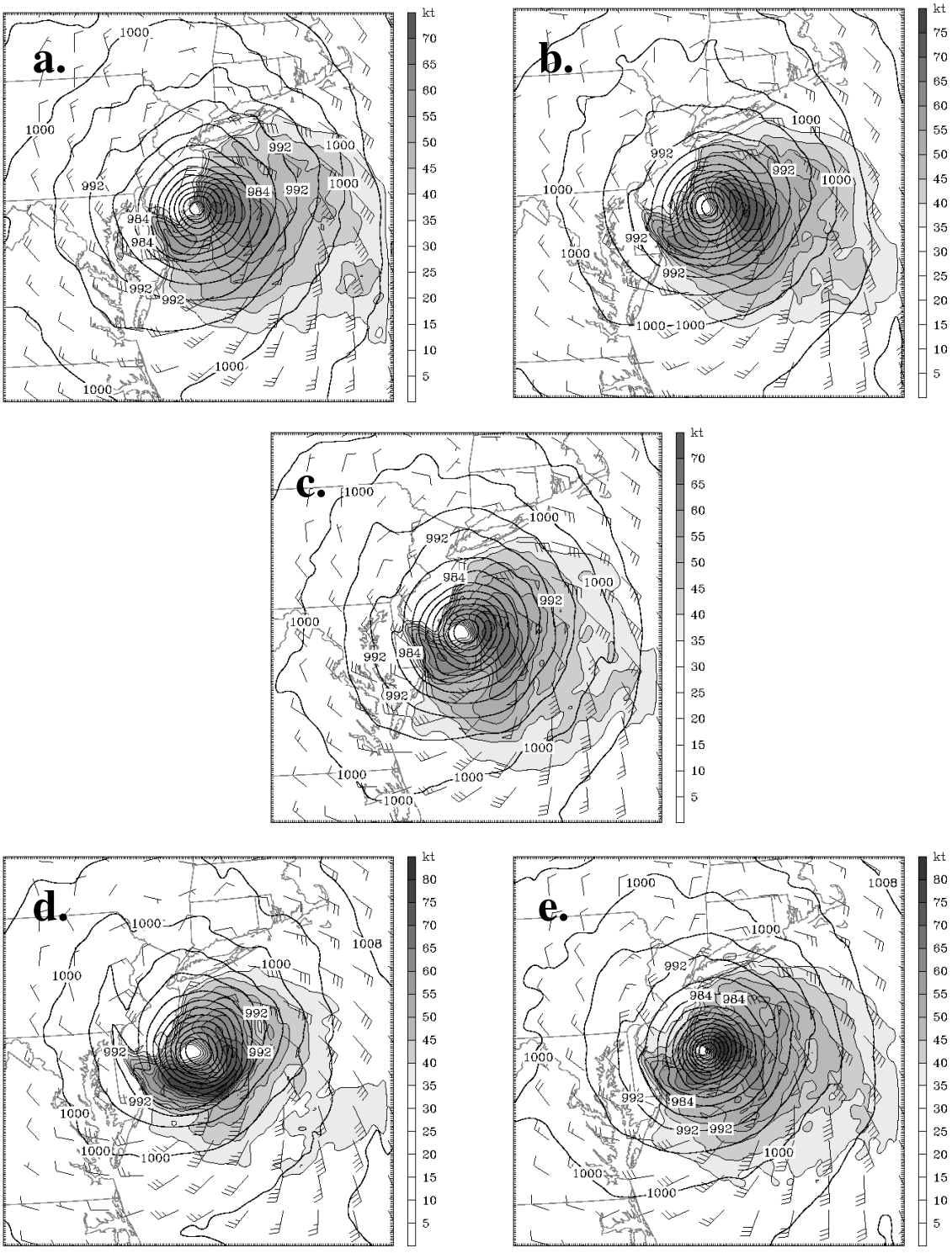


Figure 3.15. (a) 10-m wind speed (shaded) and wind barbs (full barb = 10 kts) at 1400 UTC 27 September 1985 for the CTL simulation. (b) Same as (a) except for the GFS-FERR simulation. (c) Same as (a) except for the YSU-LIN simulation. (d) Same as (a) except for the YSU-FERR. (e) Same as (a) except for the MYJ-LIN simulation.

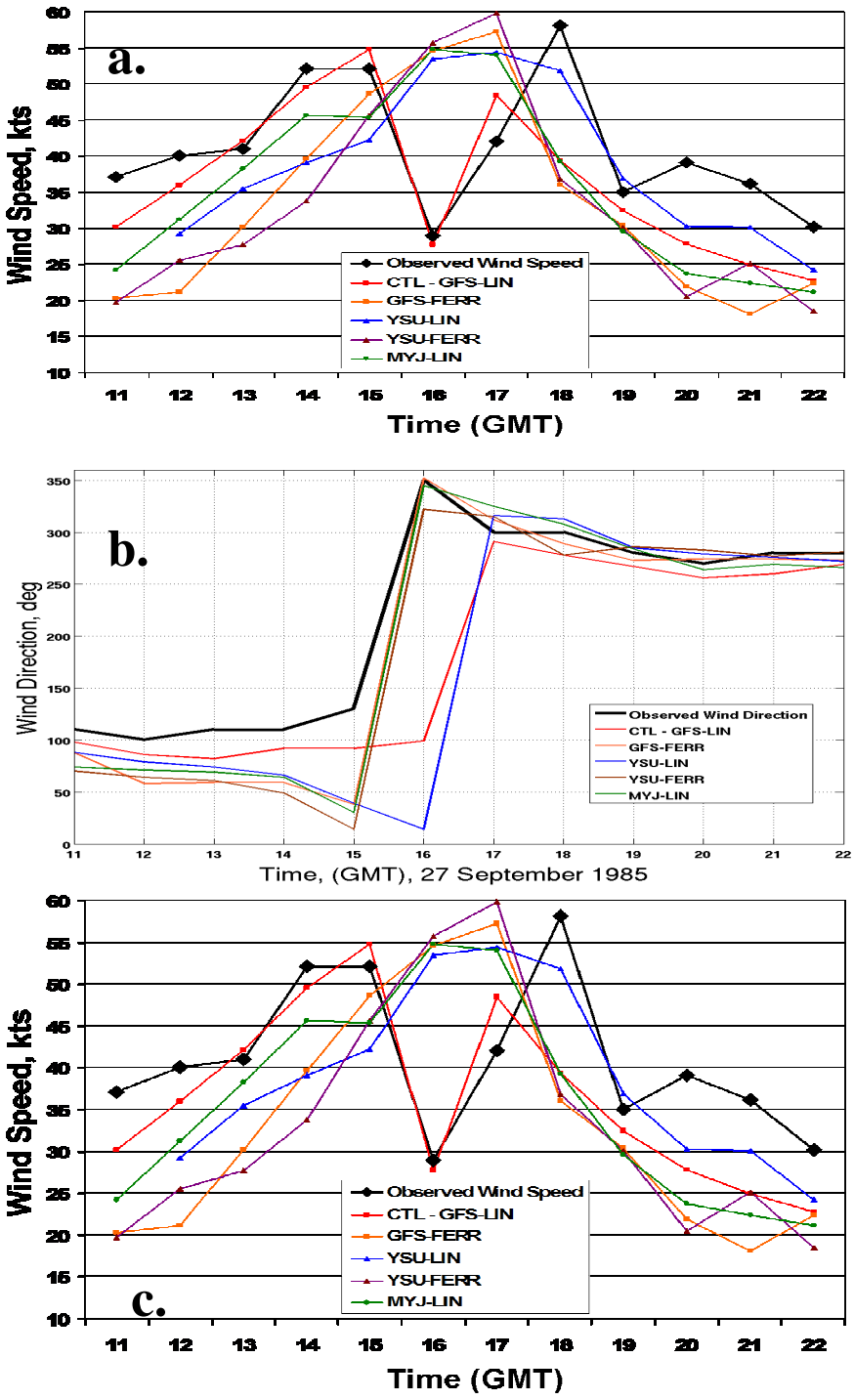


Figure 3.16. (a) 30-m wind speed (kts) at Ambrose Lighthouse for 1100 UTC – 2200 UTC 27 September 1985 for all simulations and the observed. (b) Wind direction (deg) evolution for all simulations and the observed at Ambrose Lighthouse from 1100 UTC – 2200 UTC 27 September 1985. (c) SLP (hPa) evolution at Ambrose Lighthouse for all simulations and the observed for 1100 UTC – 2200 UTC 27 September 1985.

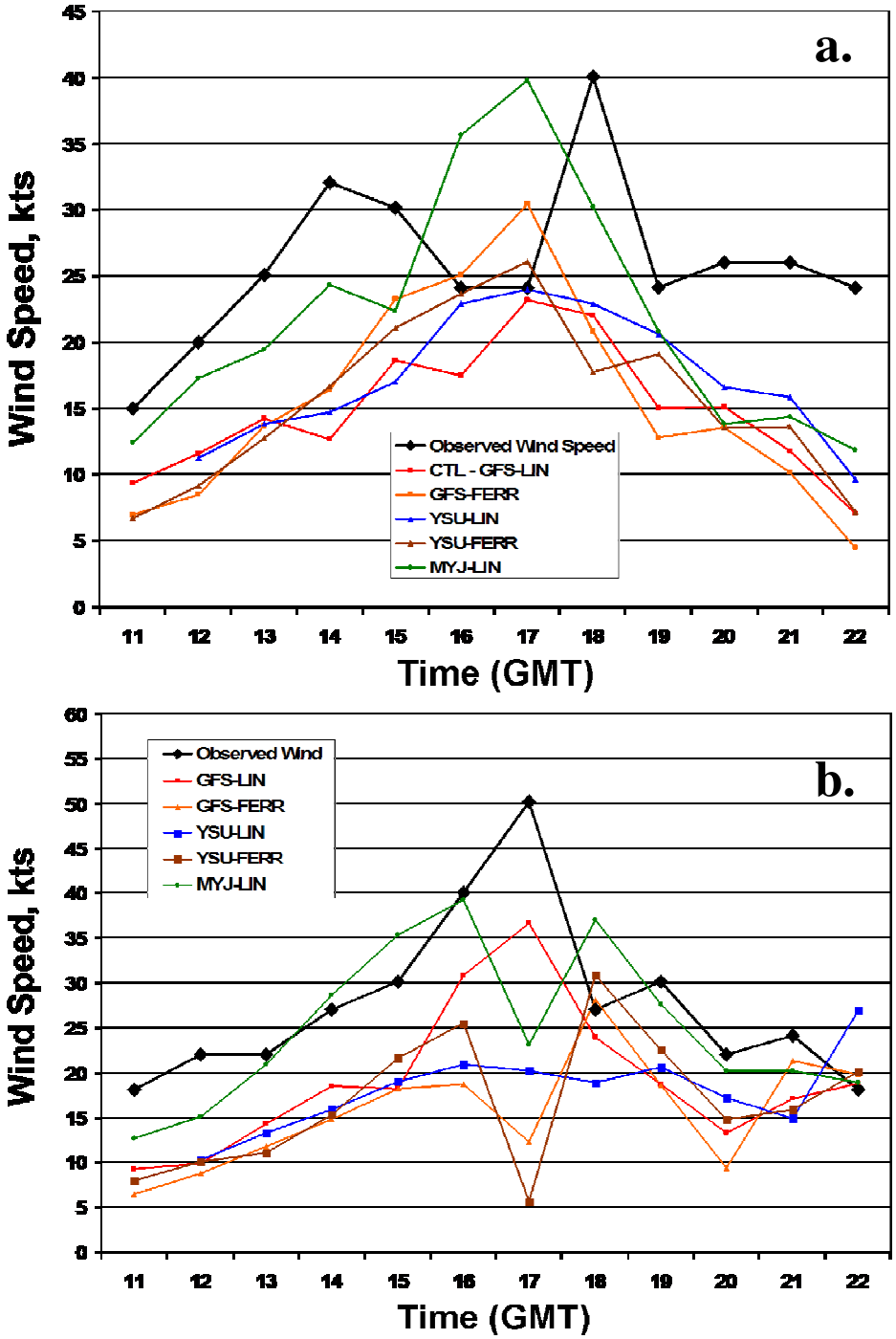


Figure 3.17. (a) 10 m wind speed (in kts) time series at JFK airport for 1100 UTC – 2200 UTC 27 September 1985. (b) Same as (a) except for Islip, NY. The location of the stations is shown in figure 1.1.

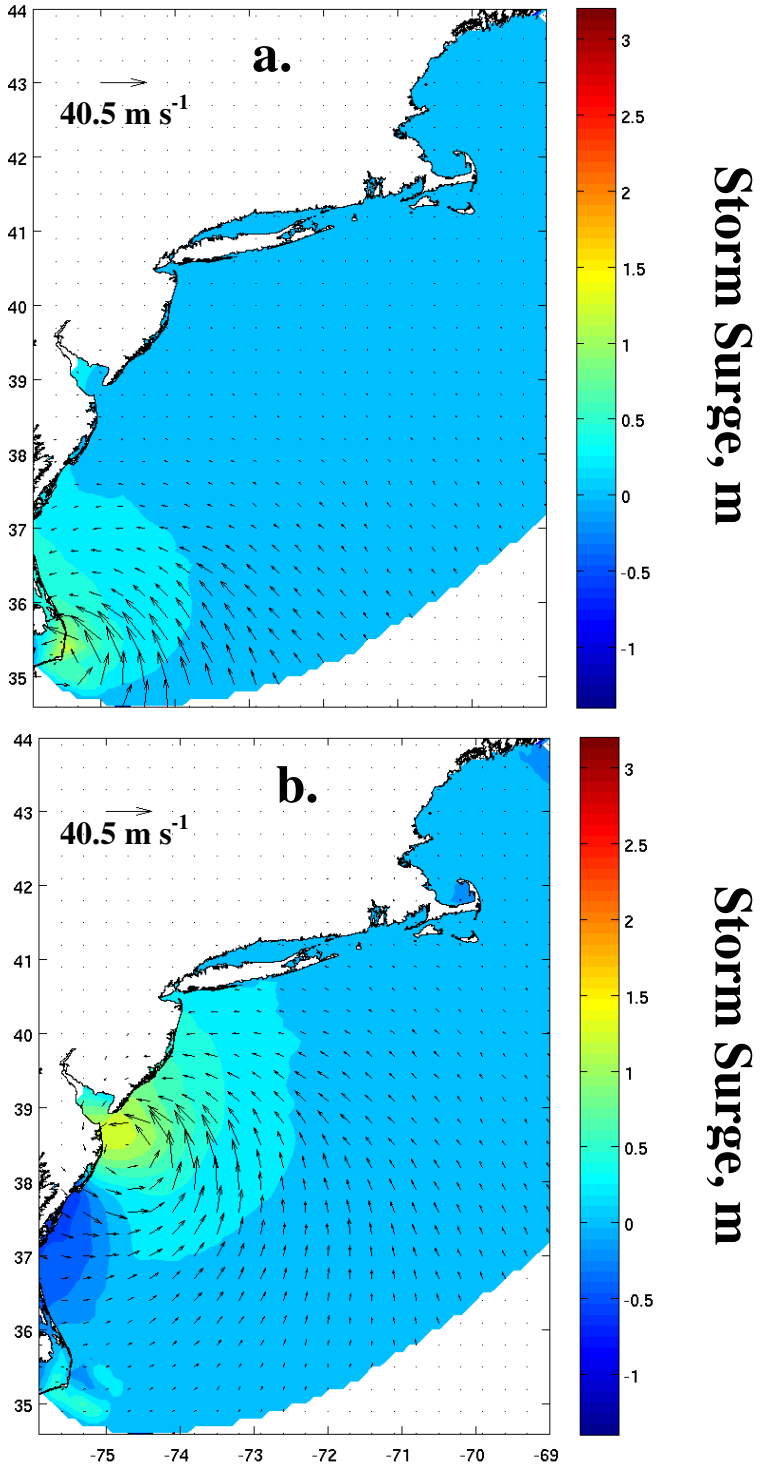


Figure 3.18. (a) Storm surge (m) and wind speed vectors at 0500 UTC 27 September 1985 when hurricane Gloria just enters the ADCIRC domain. (b) Same as (a) except at 1200 UTC 27 September 1985.

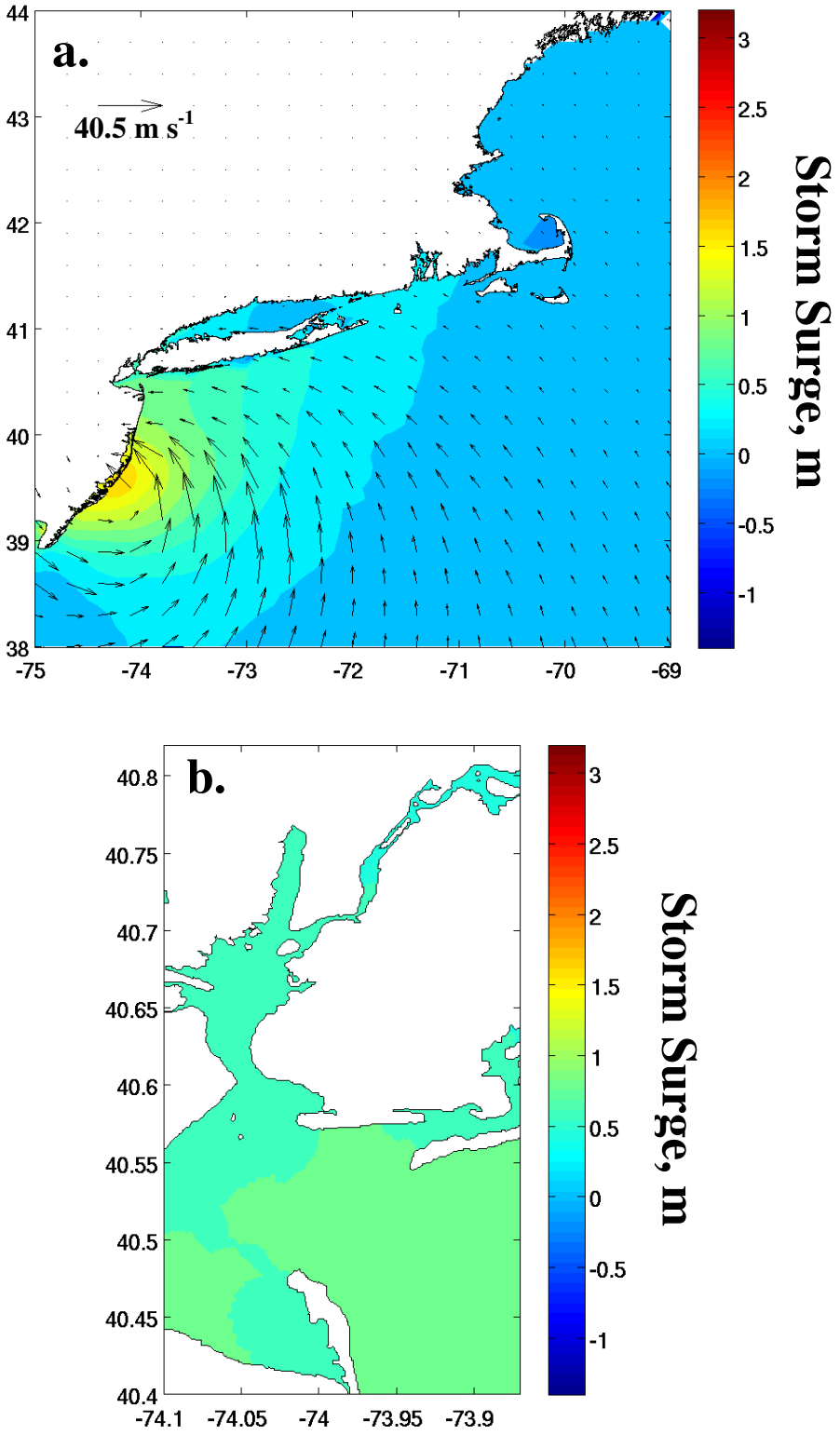


Figure 3.19. (a) CTL simulation storm surge (m) and wind vectors at 1400 UTC 27 September 1985. (b) Same as (a) except without wind vectors and zoomed into New York Harbor.

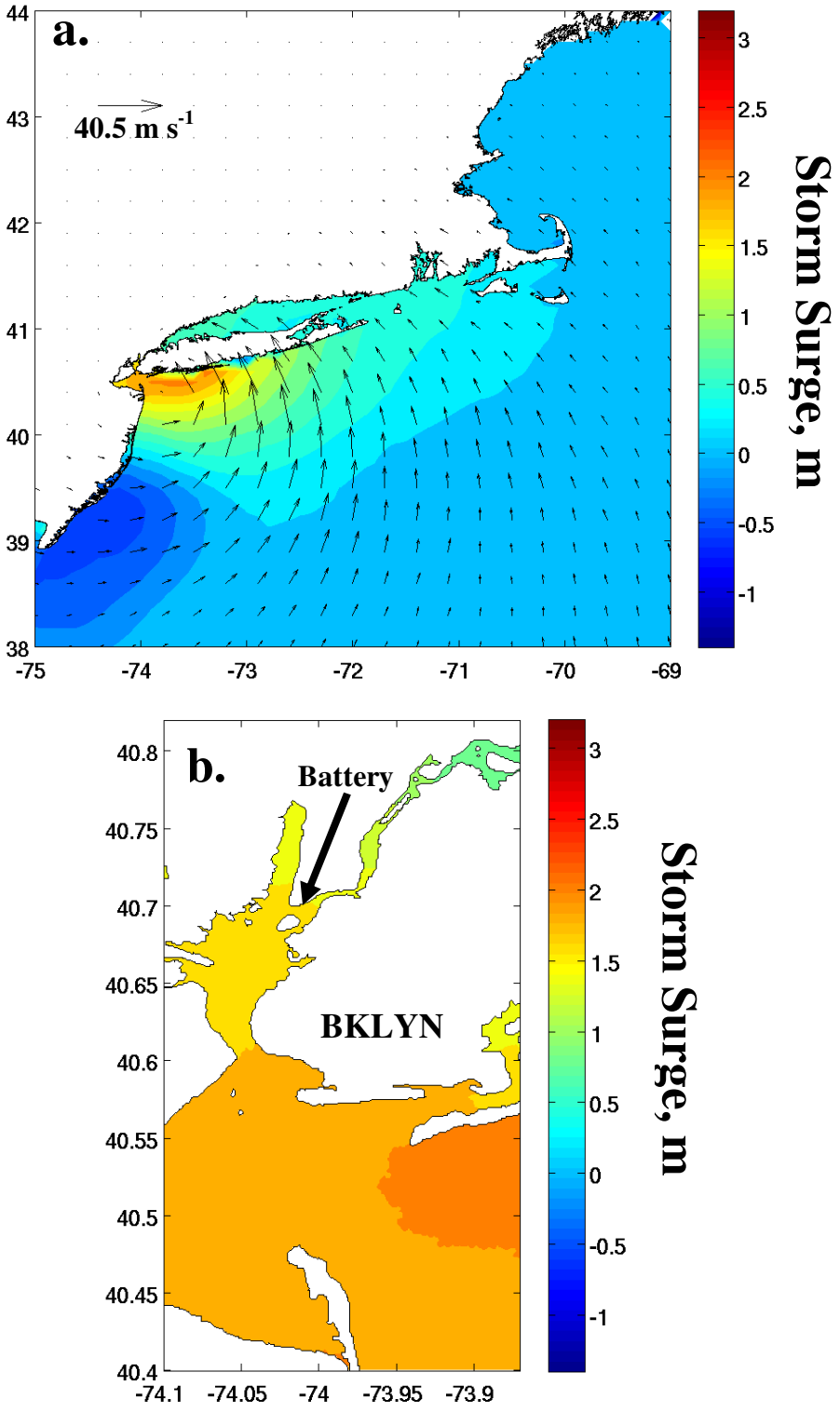


Figure 3.20. (a) CTL simulation storm surge (m) and wind vectors at 1600 UTC 27 September 1985. (b) Same as (a) except without wind vectors and zoomed into New York Harbor.

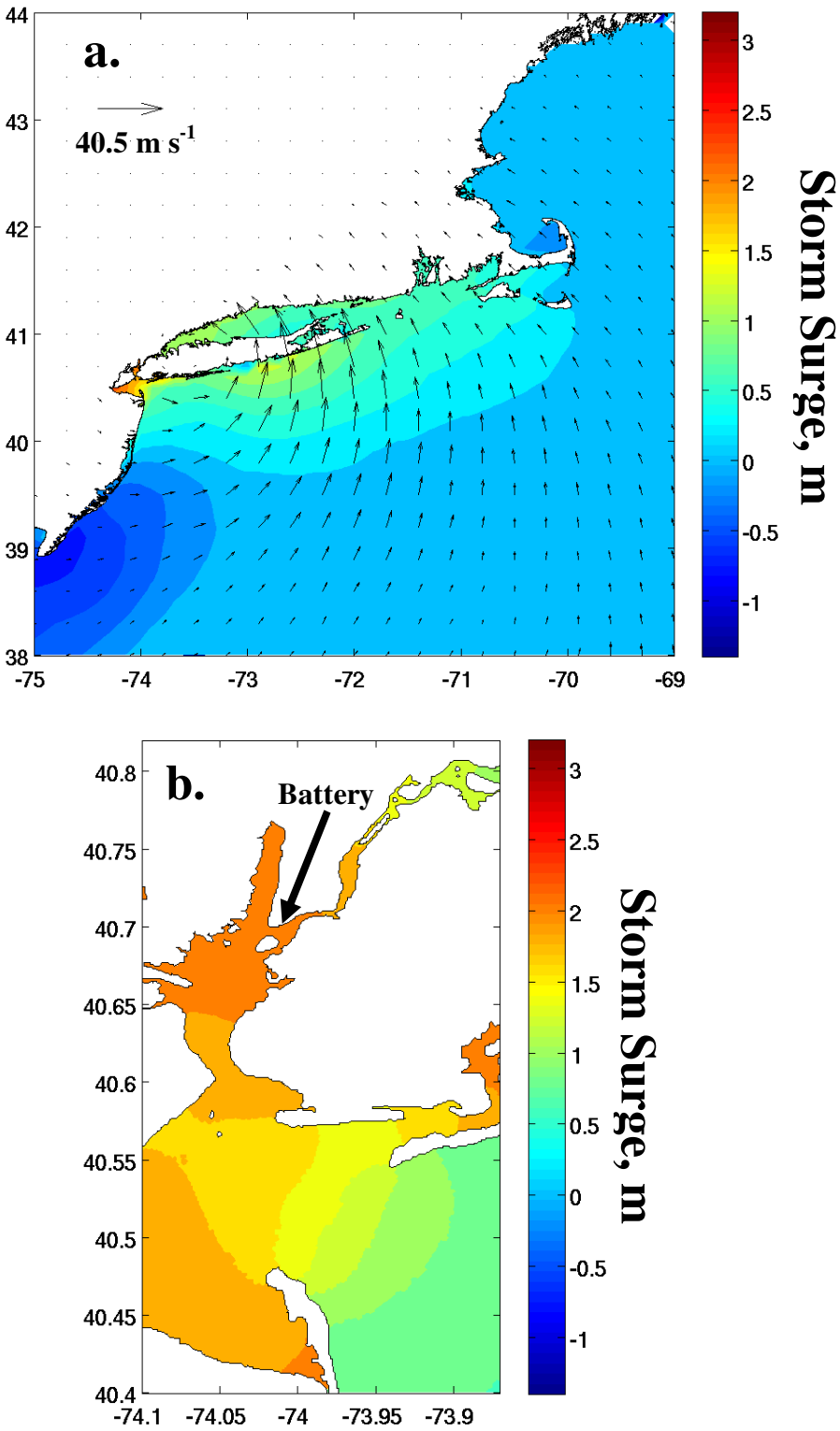


Figure 3.21. (a) CTL simulation storm surge (m) and wind vectors at 1700 UTC 27 September 1985. (b) Same as (a) except without wind vectors and zoomed into New York Harbor.

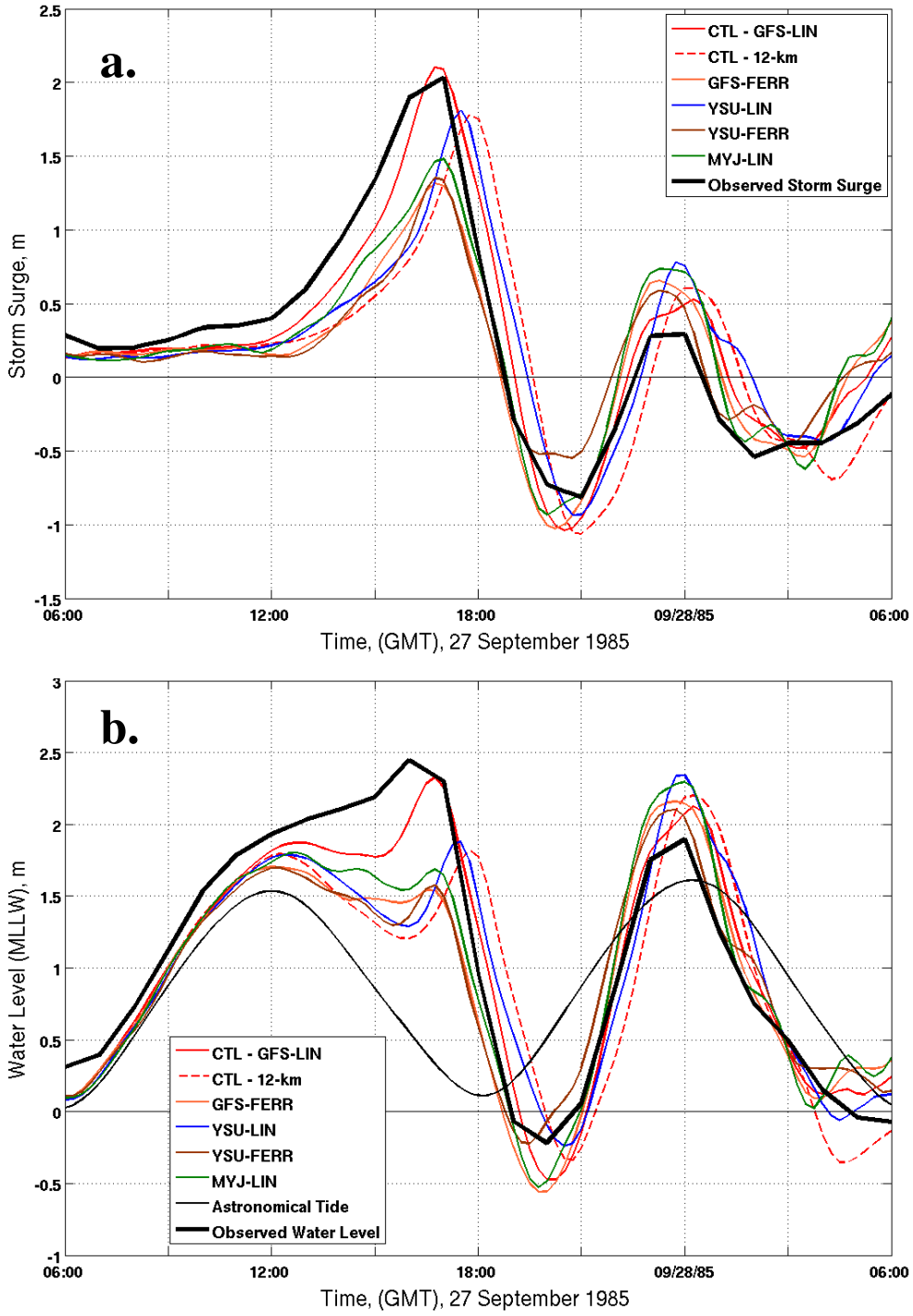


Figure 3.22. (a) Storm surge (m) evolution at the Battery, NY, for all simulations and the observed for 0600 UTC 27 September 1985 – 0600 UTC 28 September 1985. (b) Same as (a) except total water level above MLLW (in m).

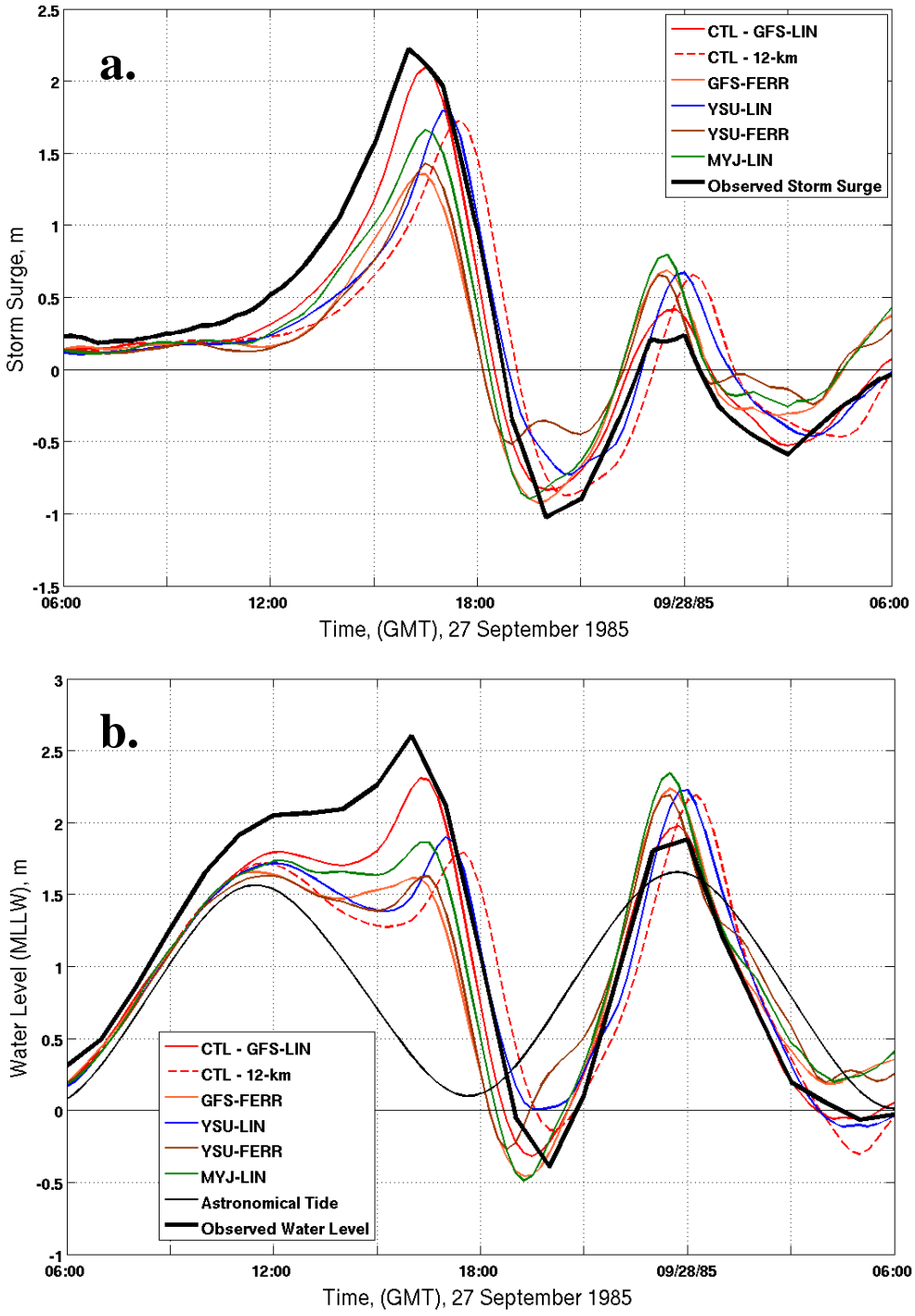


Figure 3.23. (a) Storm surge (m) evolution at the Sandy Hook, NJ, for all simulations and the observed for 0600 UTC 27 September 1985 – 0600 UTC 28 September 1985. (b) Same as (a) except total water level above MLLW (in m).

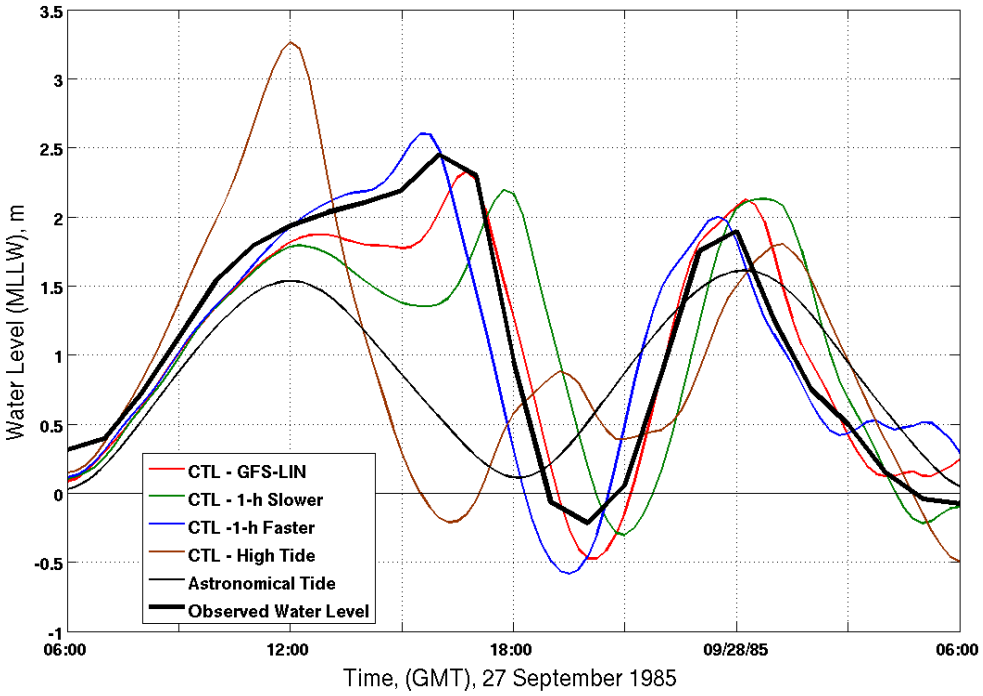


Figure 3.24. Time series of water level above MLLW (in m) at the Battery for experiments looking at the sensitivity of ADCIRC to changes in hurricane landfall timing. Four simulations are plotted using the CTL wind forcing: CTL, CTL started one hour early, CTL started one hour later, CTL started so that Gloria landfall coincided with high tide at the Battery.

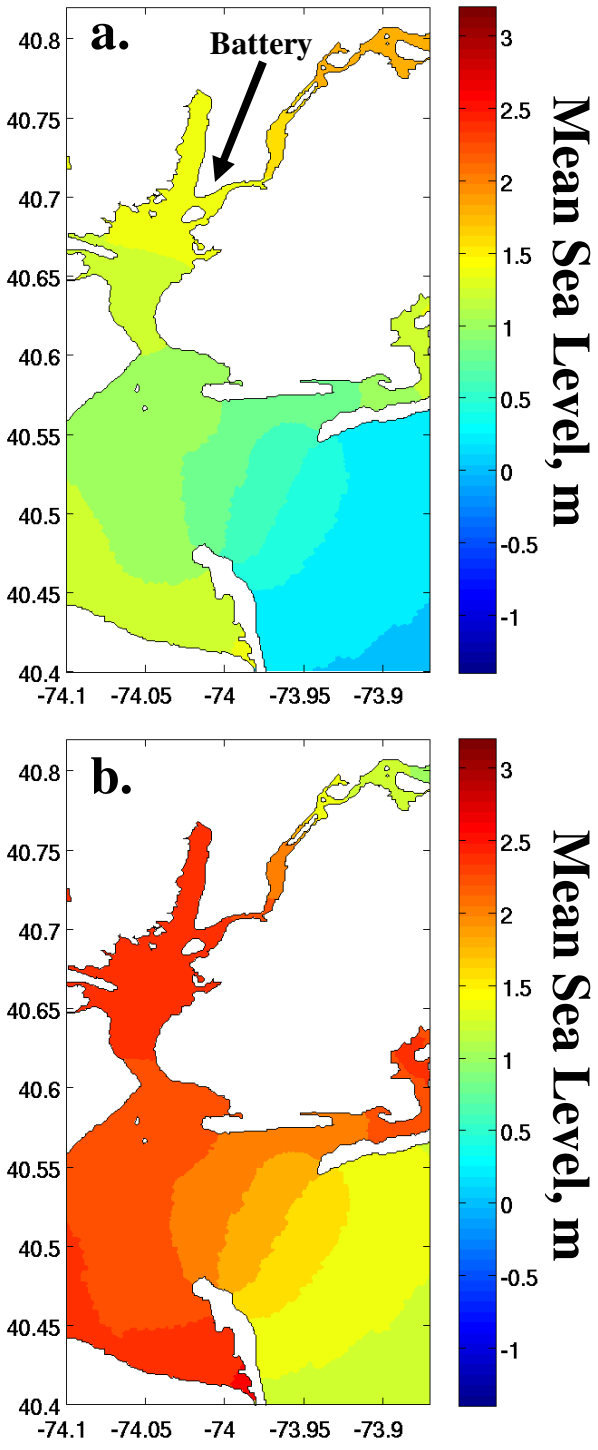


Figure 3.25. (a) CTL simulation water level (MSL), m, at 1700 UTC 27 September 1985, or peak water level, zoomed in on New York Harbor. (b) Same as (a) except for CTL-high tide simulation at 1200 UTC 27 September 1985 or peak water level in the CTL-high tide simulations.

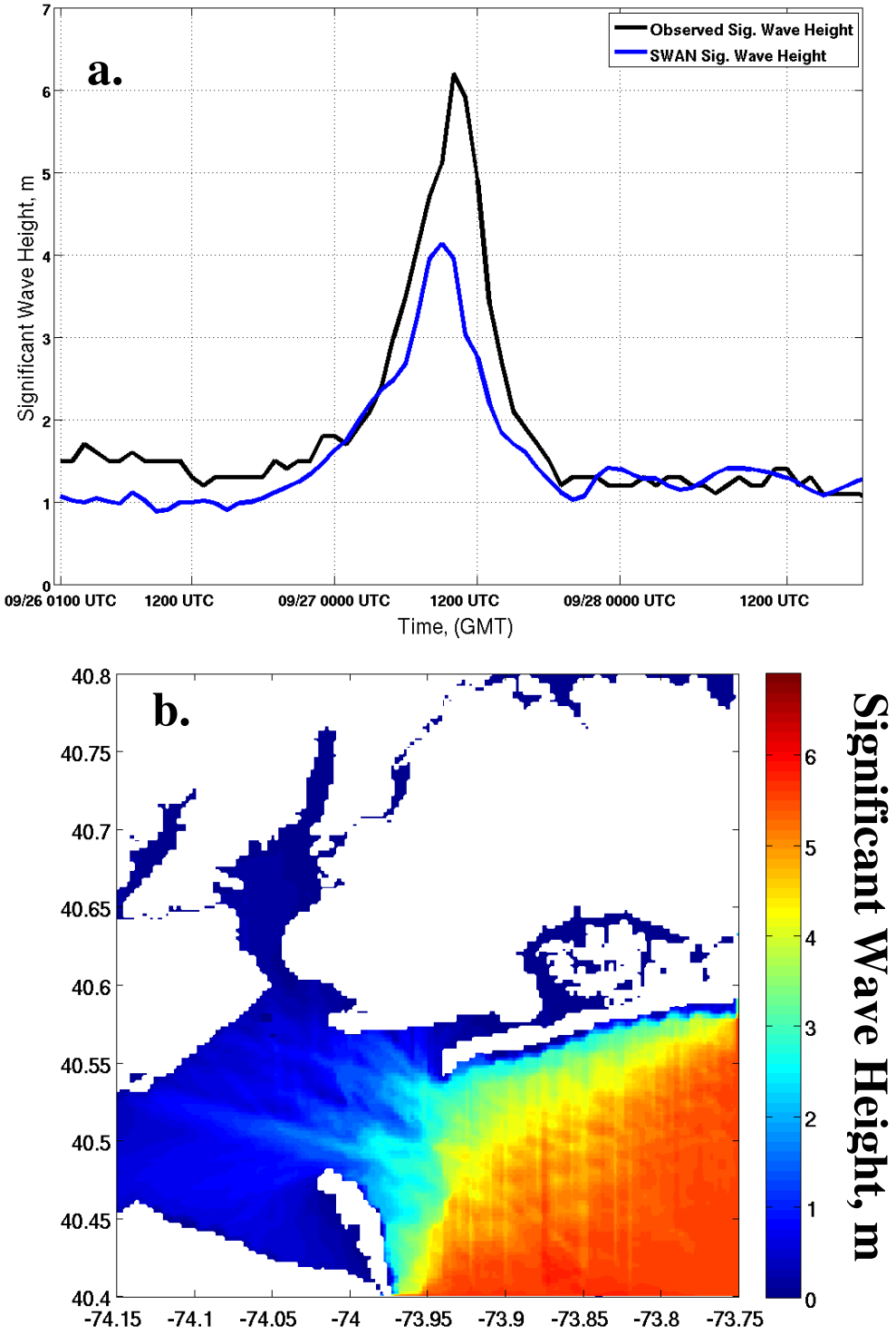


Figure 3.26. (a) SWAN significant wave height, m, comparison with the observed at station CHLV2 (Chesapeake Light), off the coast of VA during hurricane Gloria. (b) SWAN significant wave height (m) for the fine grid (grid spacing = 150 m) at 1700 UTC 27 September 1985.

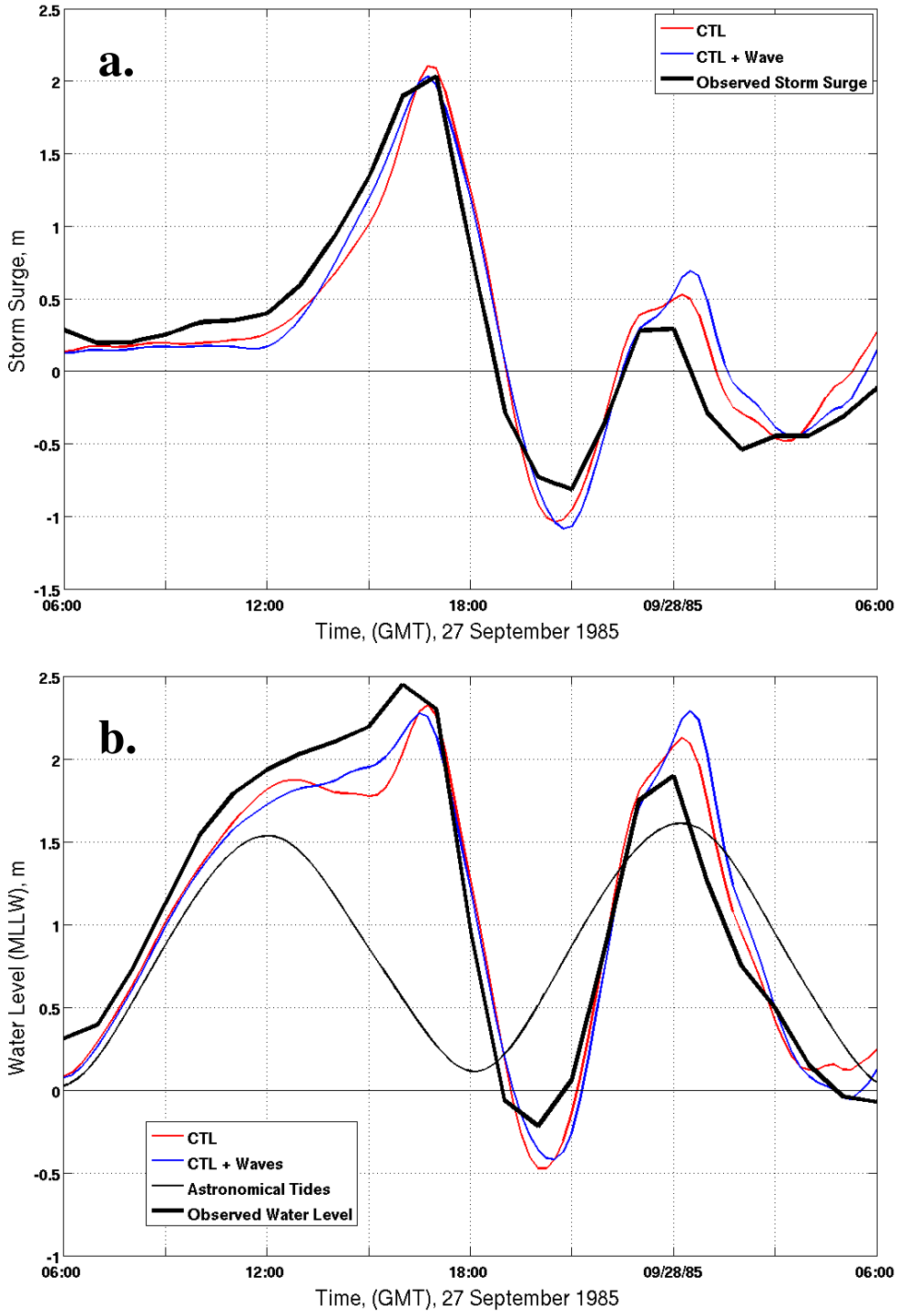


Figure 3.27. (a) Time series of storm surge (in m) comparing the CTL simulation with the CTL simulation including wave forcing. (b) Same as (a) except total water level above MLLW (in m).

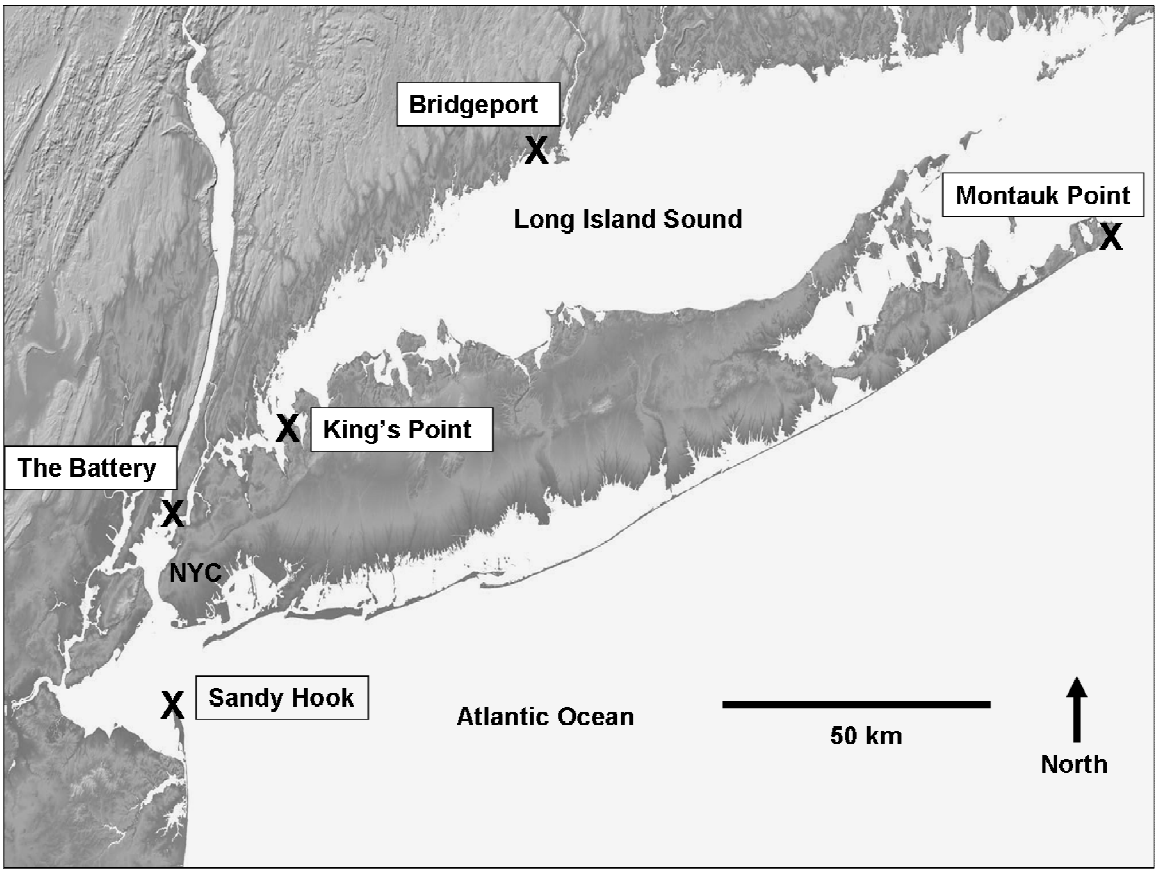


Figure 4.1: Spatial map showing the locations of the 5 water level sites around Long Island used for storm surge verification.

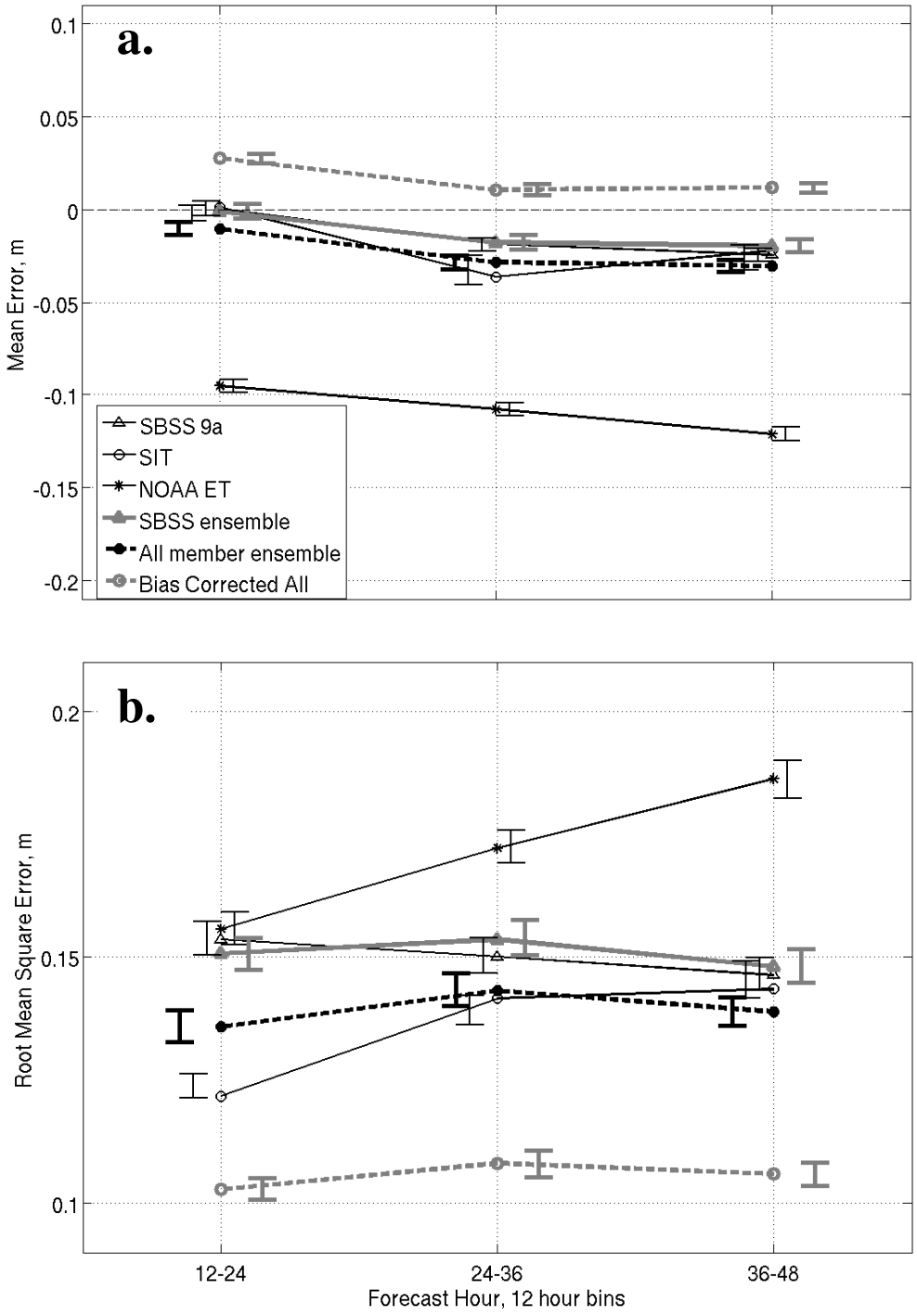


Figure 4.2 (a) Mean error (in m) in predicted storm surge versus forecast hour averaged over 12 hour periods and the 5 stations throughout southern New England for 7 different model forecasts. (b) Same as (a) except for root mean square error (in m).

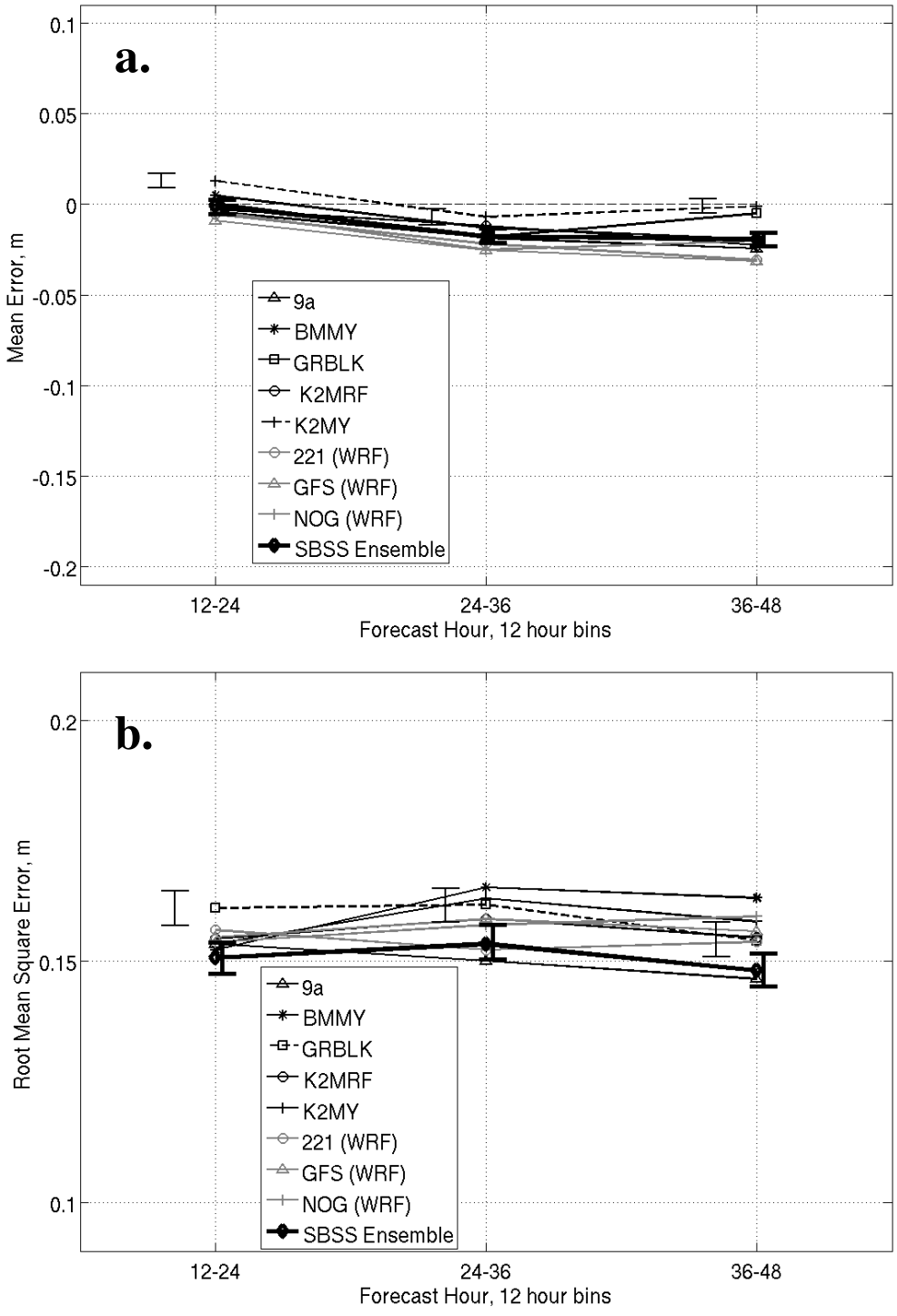


Figure 4.3. (a) Same as Figure 4.2 except for each member of the Stony Brook Storm Surge (SBSS) ensemble. (b) Same as (a) except with root mean square errors.

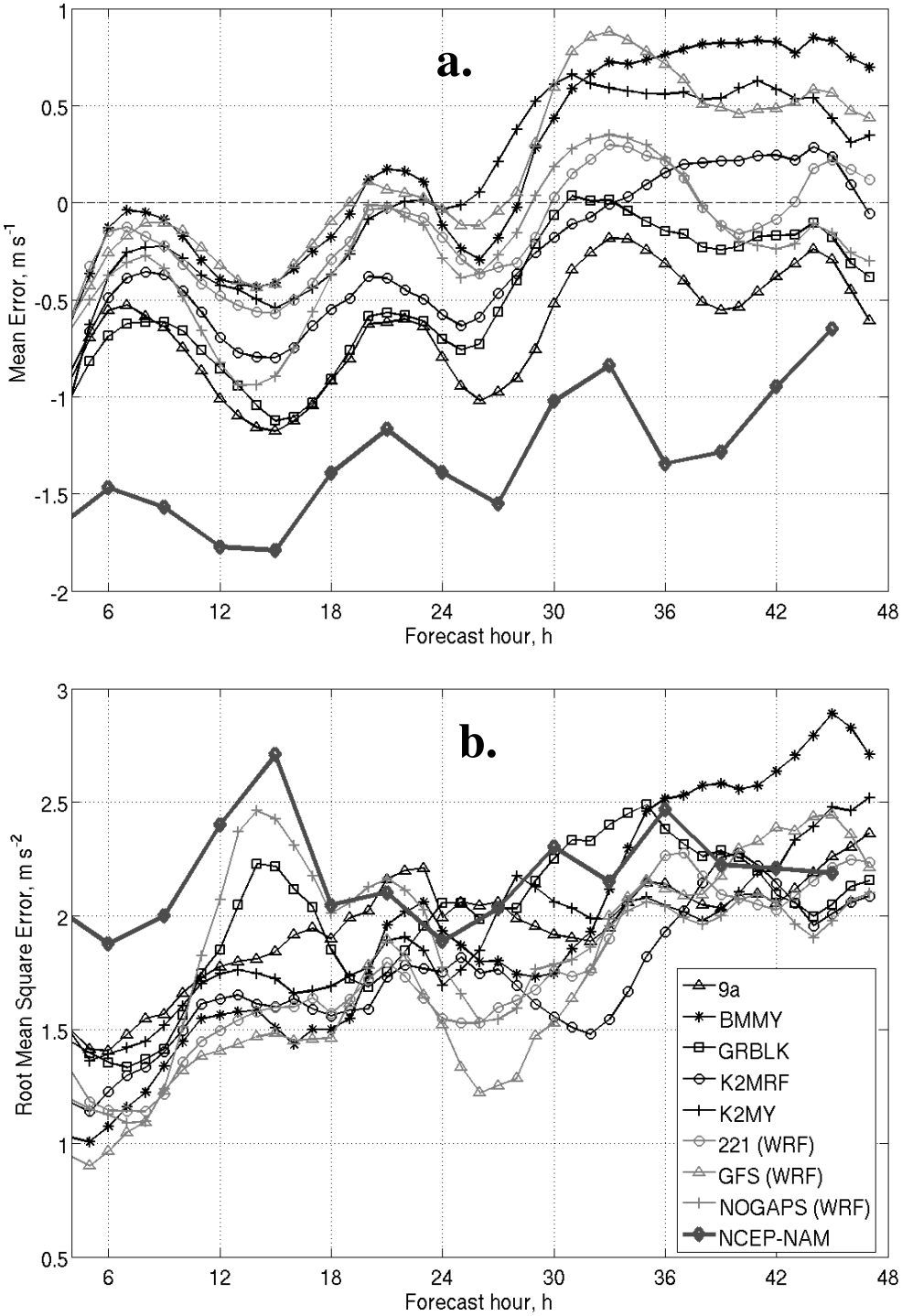
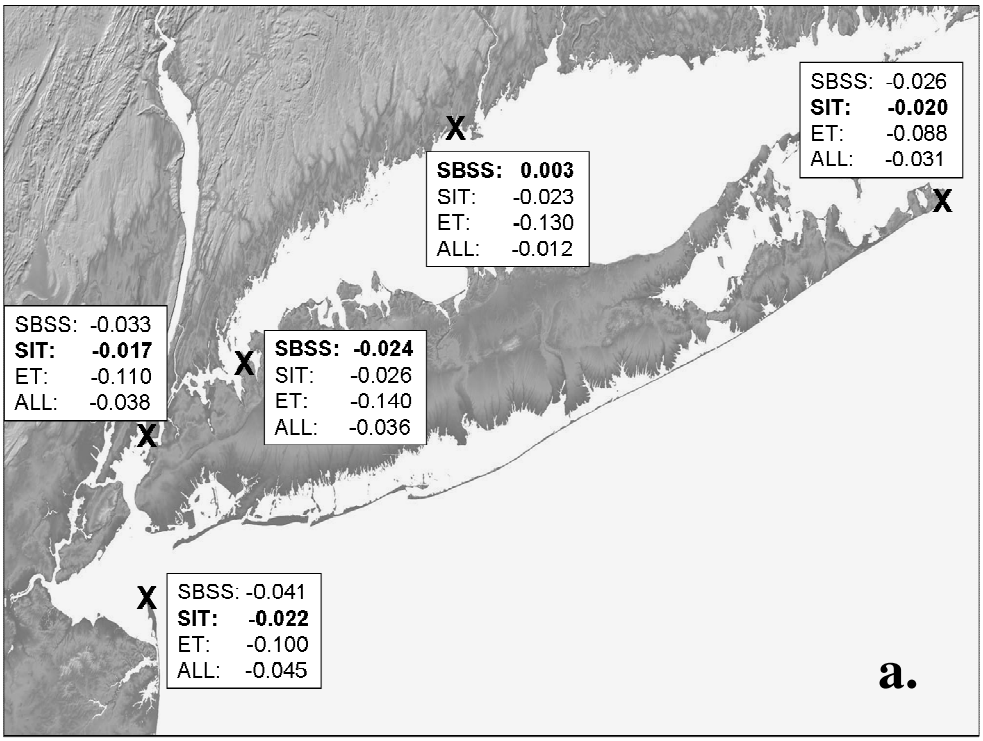


Figure 4.4. (a) Wind speed mean error (m) for all 8 members of the SBSS ensemble and the NCEP-NAM averaged over 48 hours of the forecast. (b) Same as (a) except for root mean square error (m).



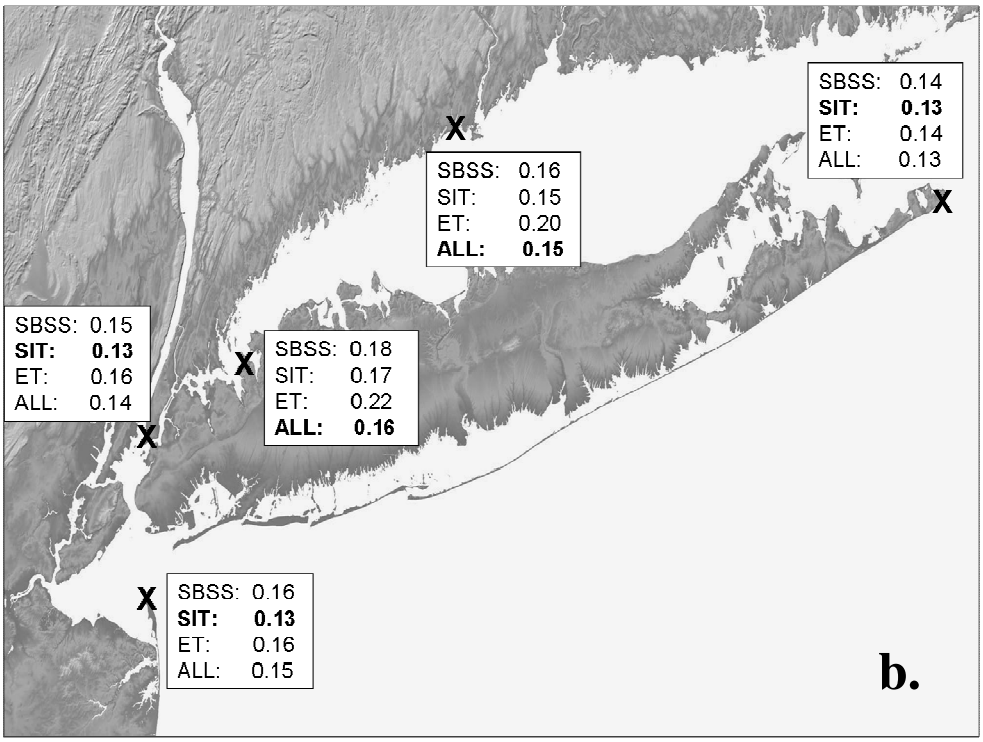


Figure 4.5 (a) Storm surge mean errors averaged over the 75 available days for hours 12-48 at the 5 stations (the Battery, Sandy Hook, King's Point, Bridgeport, and Montauk Point). (b) Same as (a) except with root mean square error.

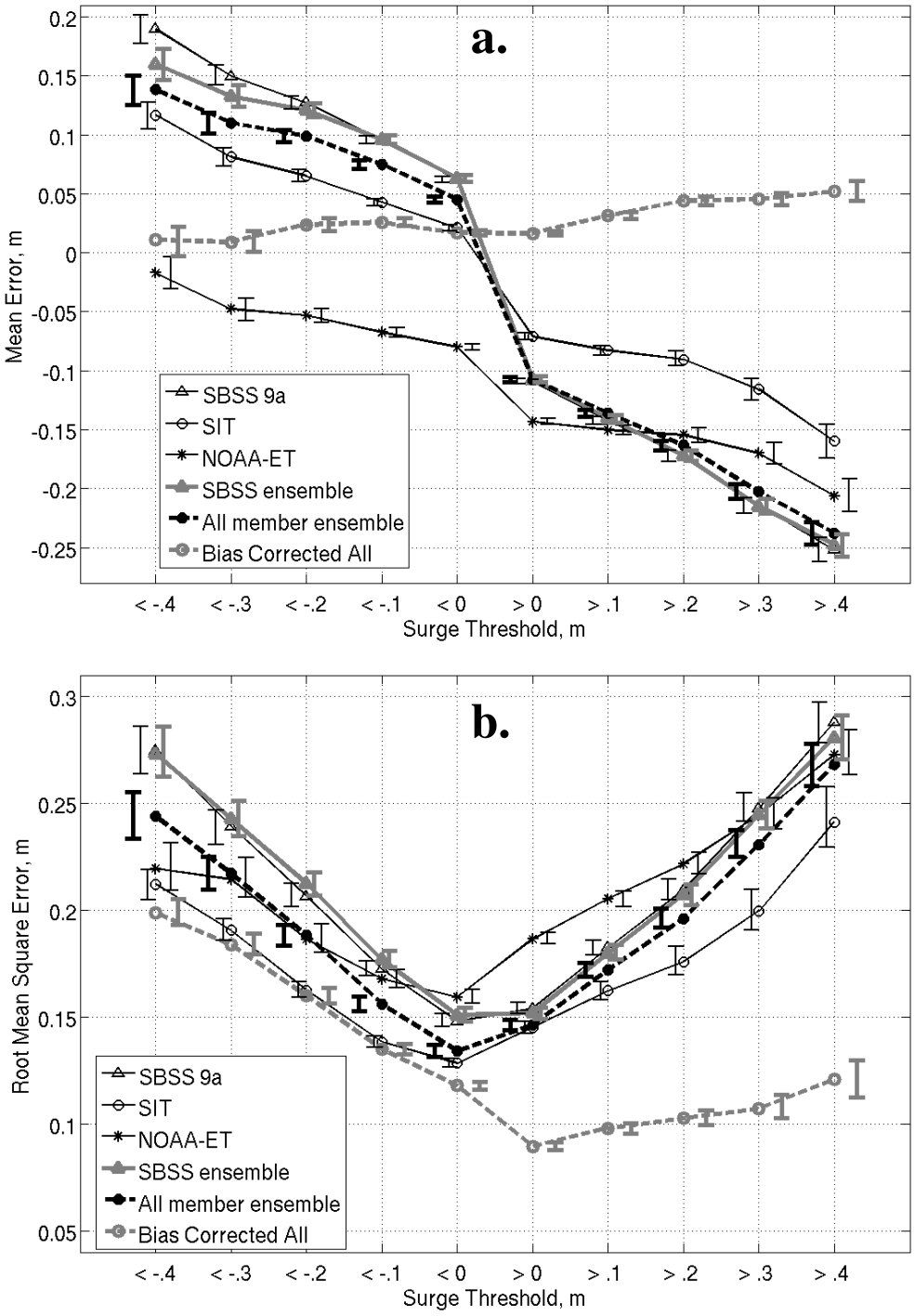


Figure 4.6. (a) Storm surge mean error across 5 stations for each of 10 different storm surge thresholds. For each ensemble member data was binned depending on whether the observation or the model satisfied the threshold. (b) Same as (a) except for RMSE error and the data only binned in each threshold when the observation satisfies the criteria.

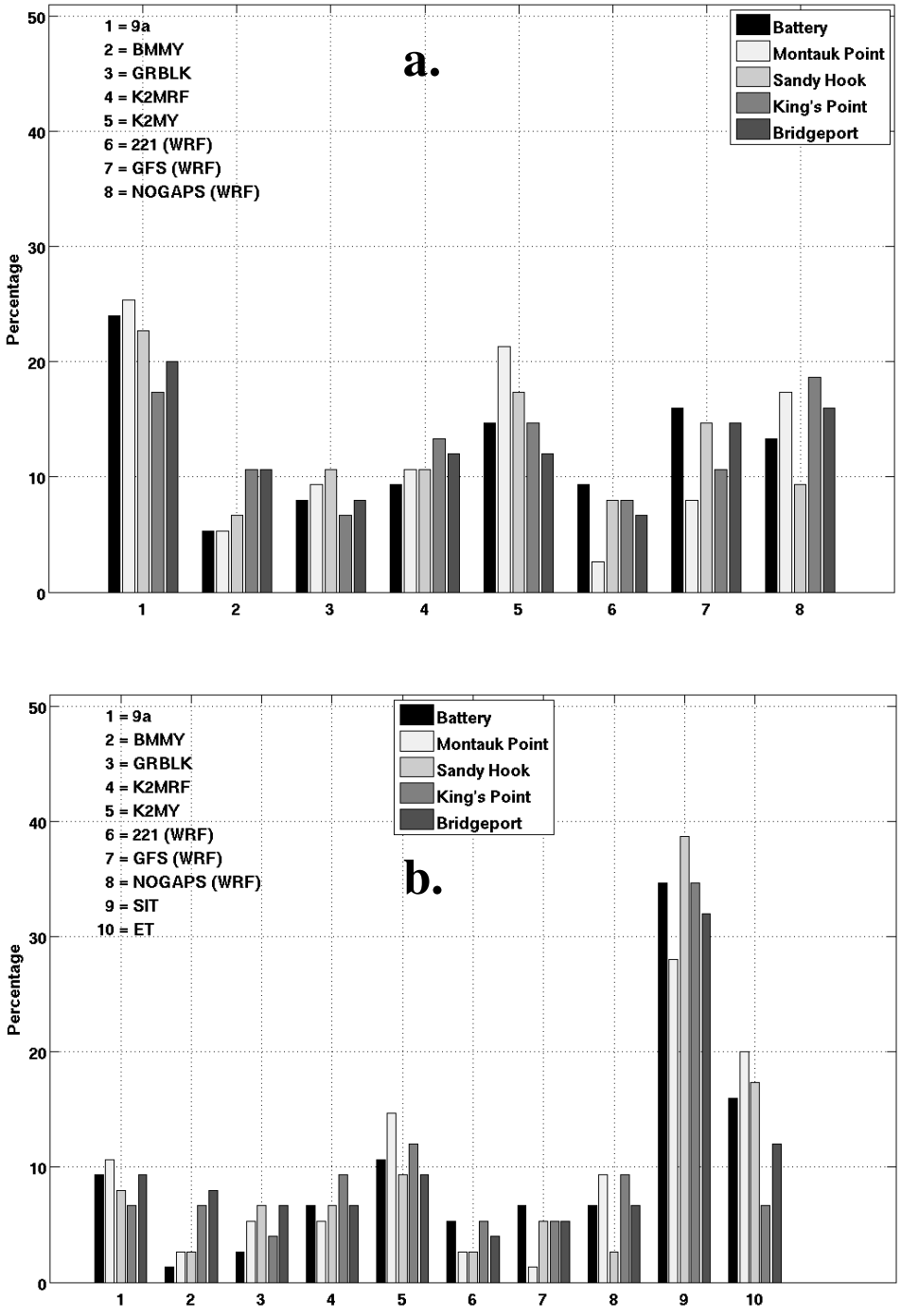


Figure 4.7. (a) Percentage of time each model member in the SBSS ensemble performed the best (lowest 12-48 hr averaged ME) in storm surge forecast at each station. (b) Same as (a) except for the ALL ensemble.

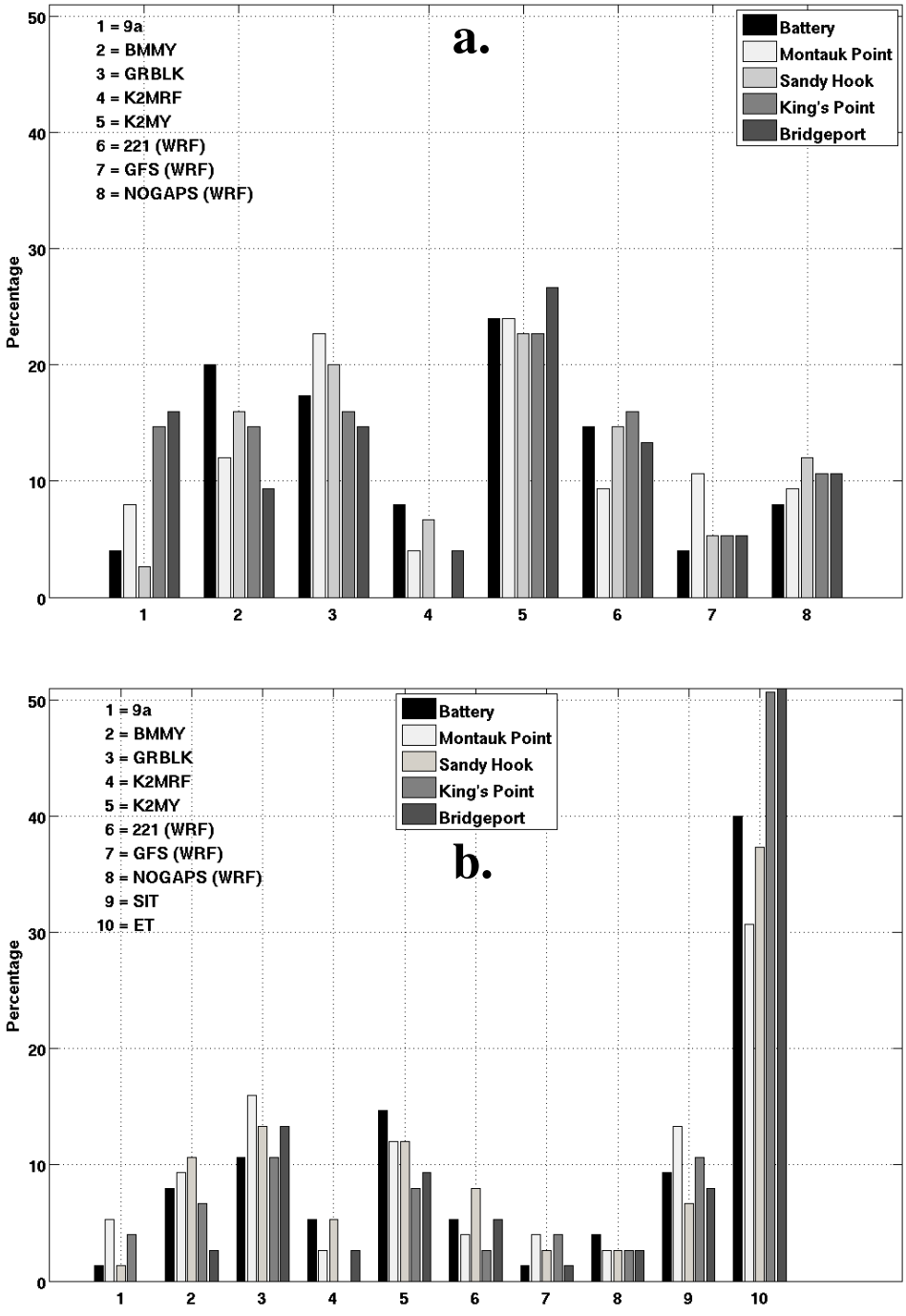


Figure 4.8. (a) Percentage of time each model member in the SBSS ensemble performed the worst (highest 12-48 hr averaged ME) in storm surge forecast at each station. (b) Same as (a) except for the ALL ensemble.

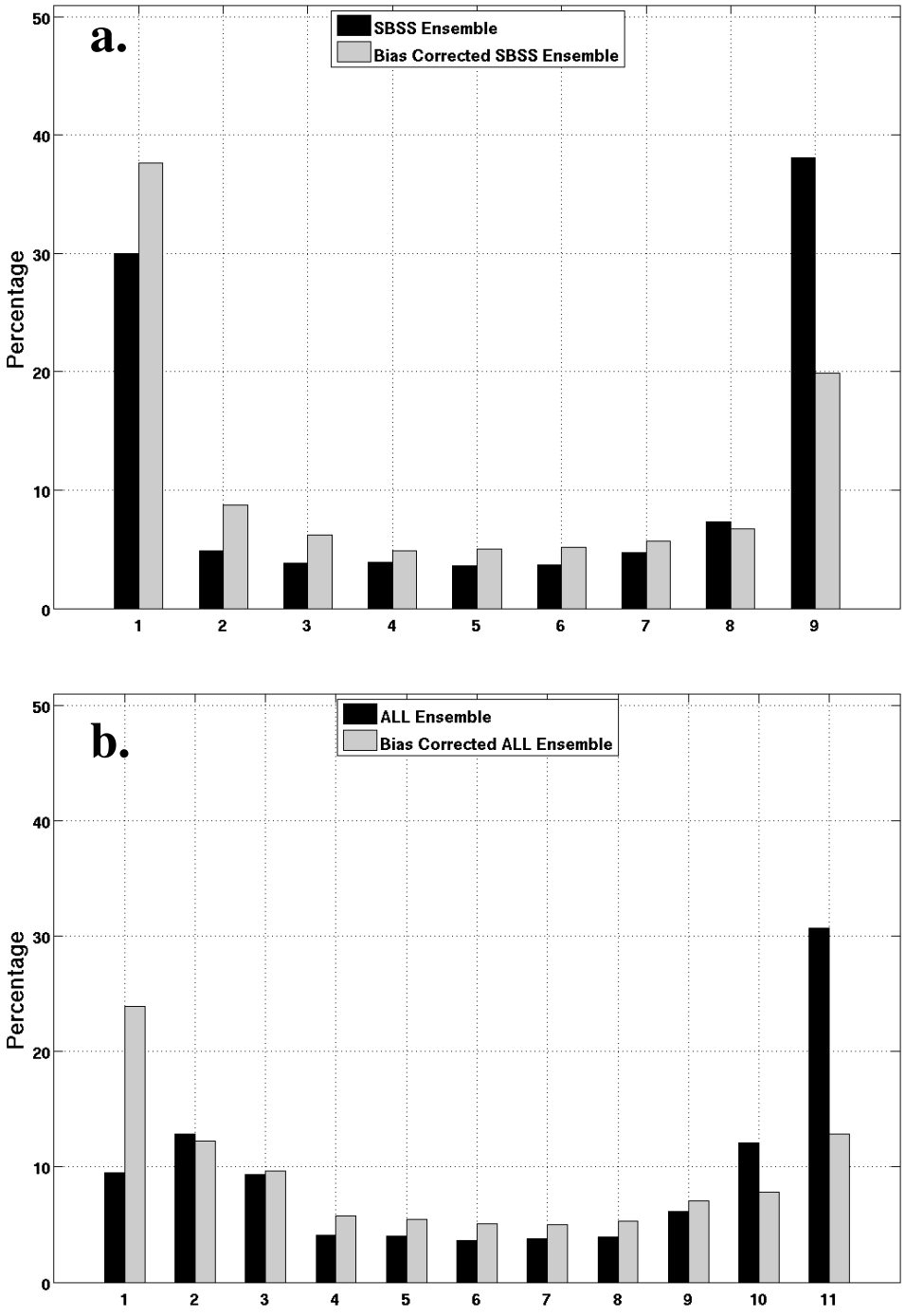


Figure 4.9. (a) Rank (Talagrand) histogram for all five stations using the SBSS ensemble (black) as well as the bias corrected SBSS ensemble. (b) Same as (a) except for the ALL ensemble (black).

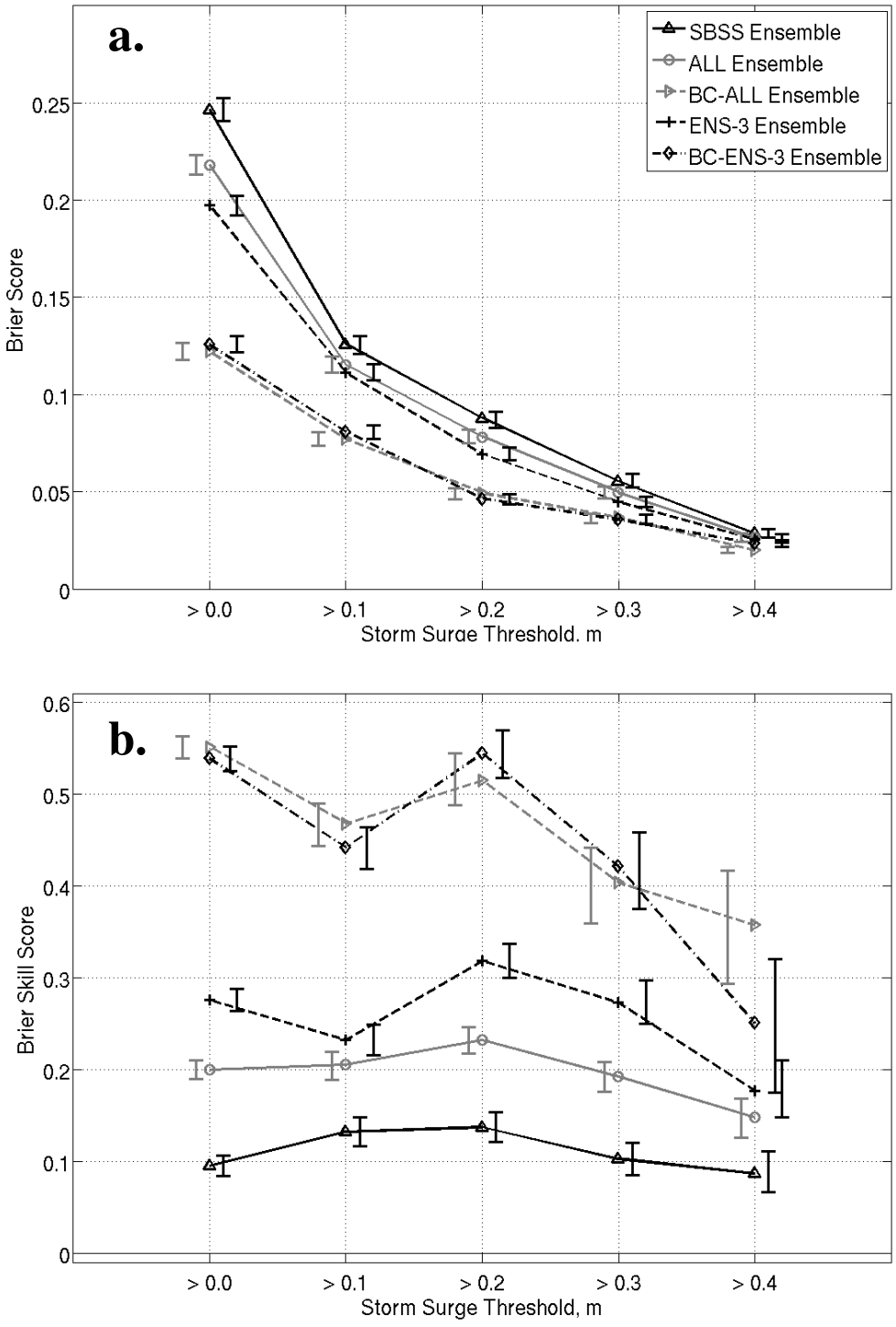


Figure 4.10. (a) Brier score calculated for the SBSS (black), ALL (gray), BC-ALL (gray dashed), ENS-3 (black dashed) and BC-ENS-3 (black dashed-dot) ensembles for 5 positive surge thresholds. (b) Same as (a) except brier skill score versus SBSS control member 9a.

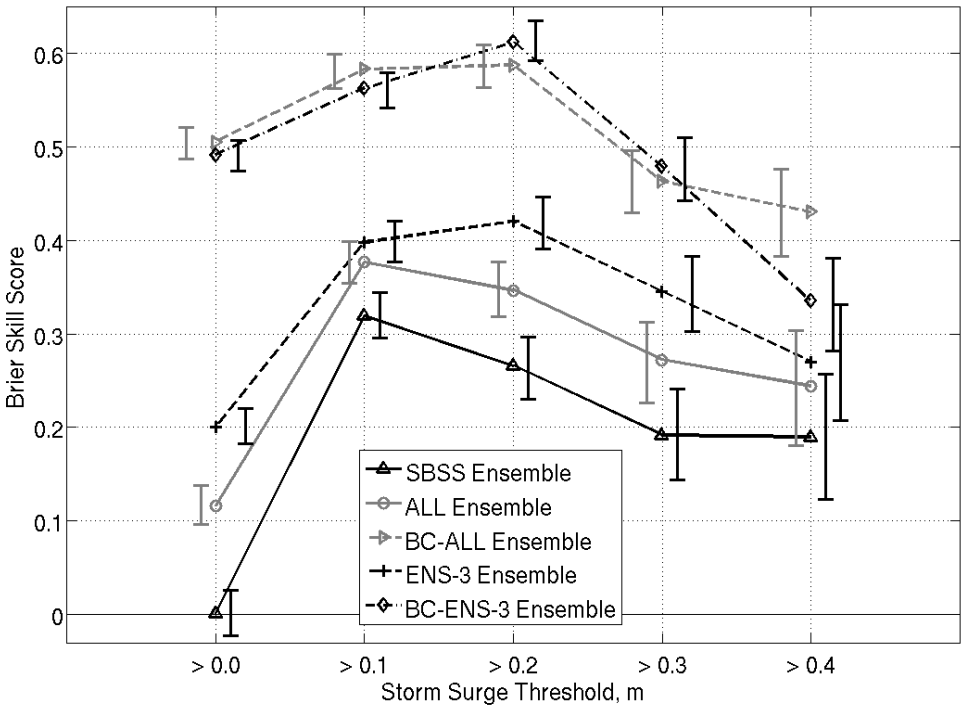


Figure 4.11. Brier skill score calculated against climatology for the SBSS (black), ALL (gray), BC-ALL (gray dashed), ENS-3 (black dashed) and BC-ENS-3 (black dashed dot) ensembles for 5 positive storm surge thresholds.

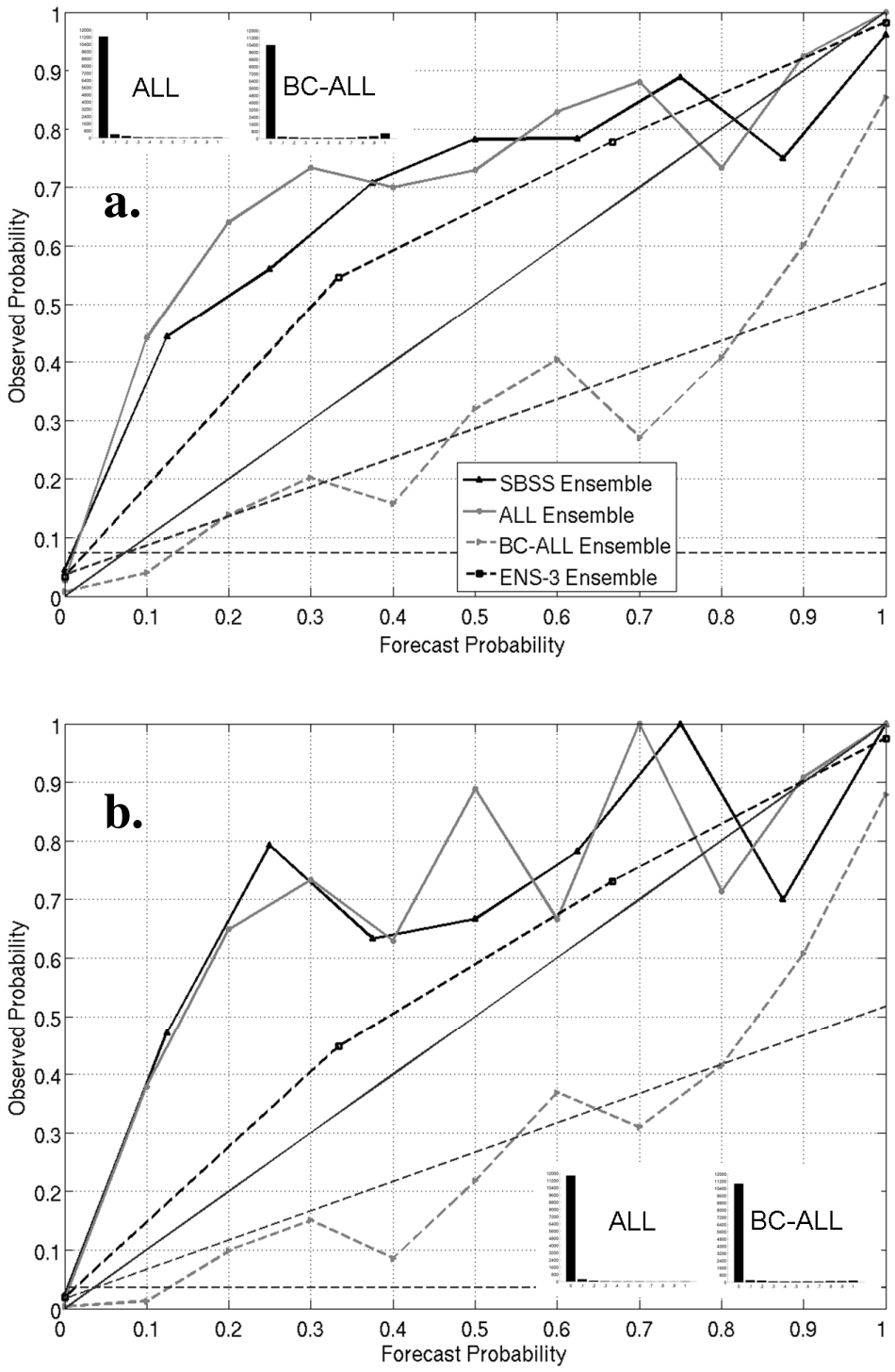


Figure 4.12. (a) Reliability diagram for the SBSS (black), ALL (gray), BC-ALL (gray dashed) and ENS-3 ensemble (black dashed) for surge observations exceeding 0.3 m. (b) Same as (a) except for surge observations exceeding 0.4 m.

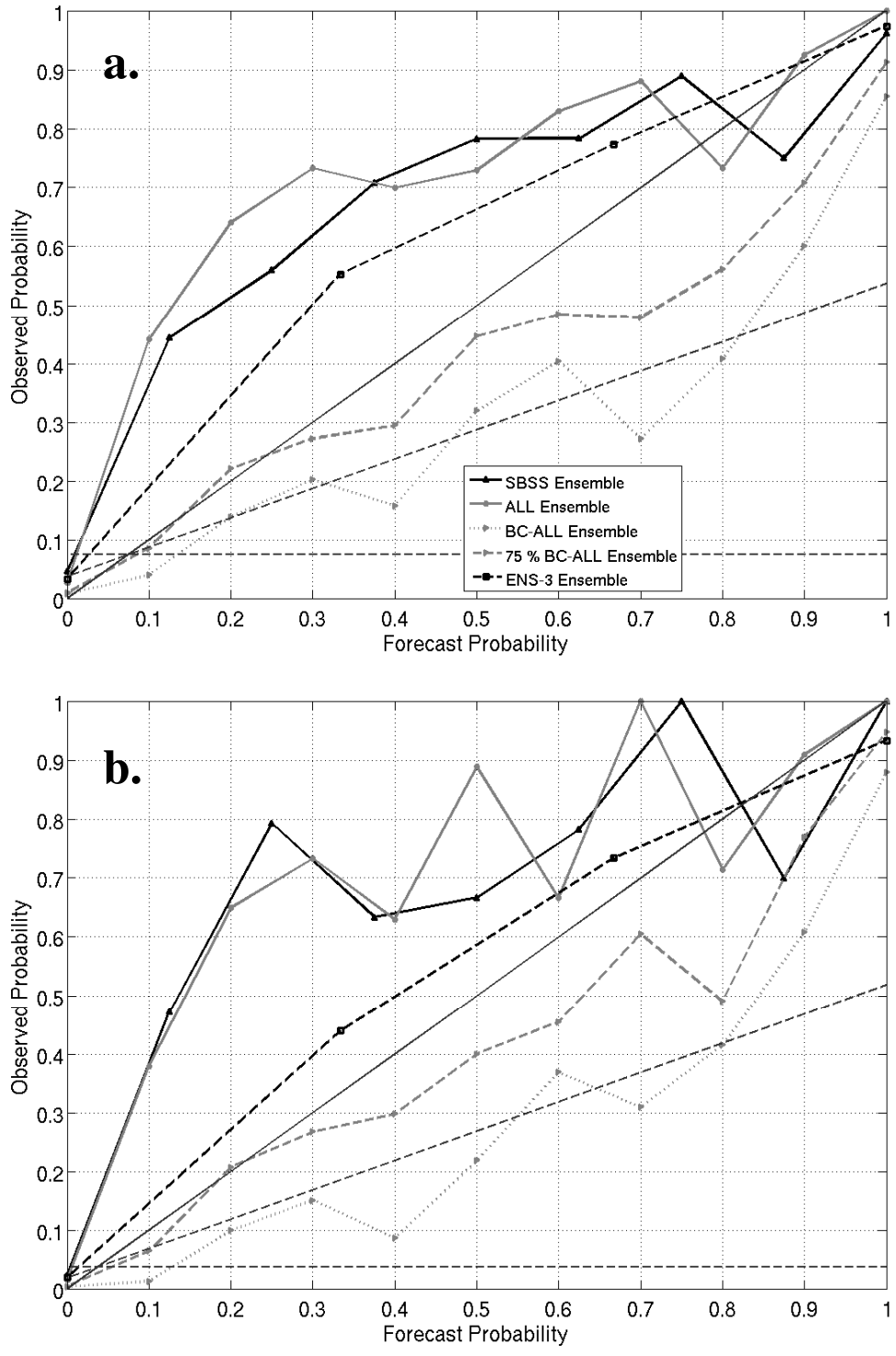


Figure 4.13. (a) Same as 4.11 (a) except the inclusion of a 75% BC-ALL member (gray dashed) which replaces the BC-ALL (gray dotted) in 4.11 (a). (b) Same as (a) except for surge observations greater than 0.4 m.

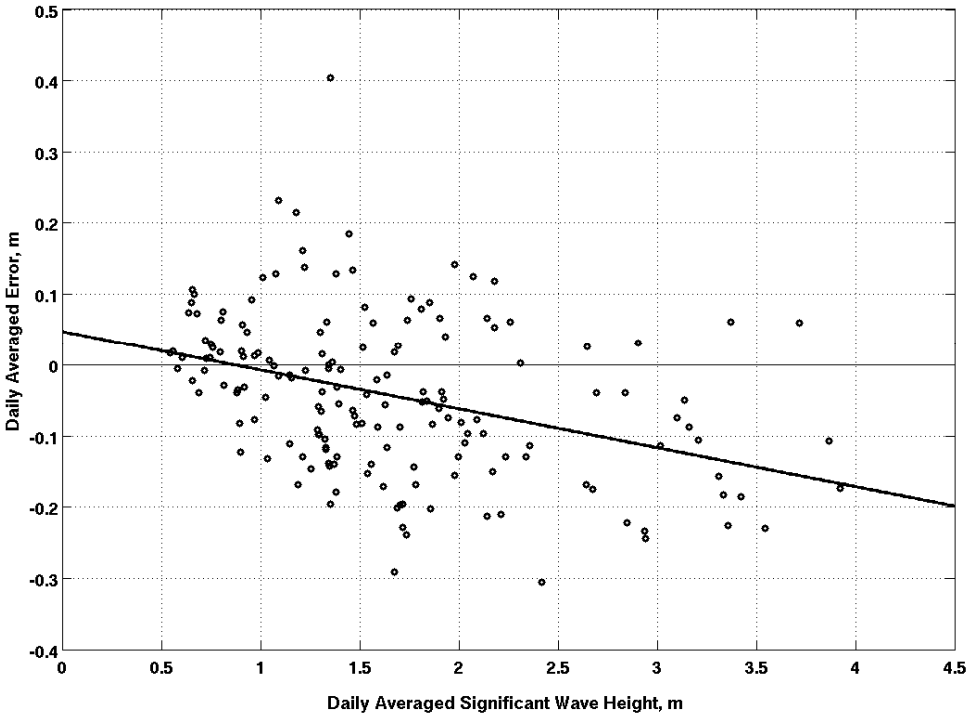


Figure 4.14. Storm error (in m) averaged for all stations using the SBSS ensemble mean between forecast hours 24-48for all 75 days. Seasonal mean error (gray) and ± 1 standard deviation lines (gray) are also included.

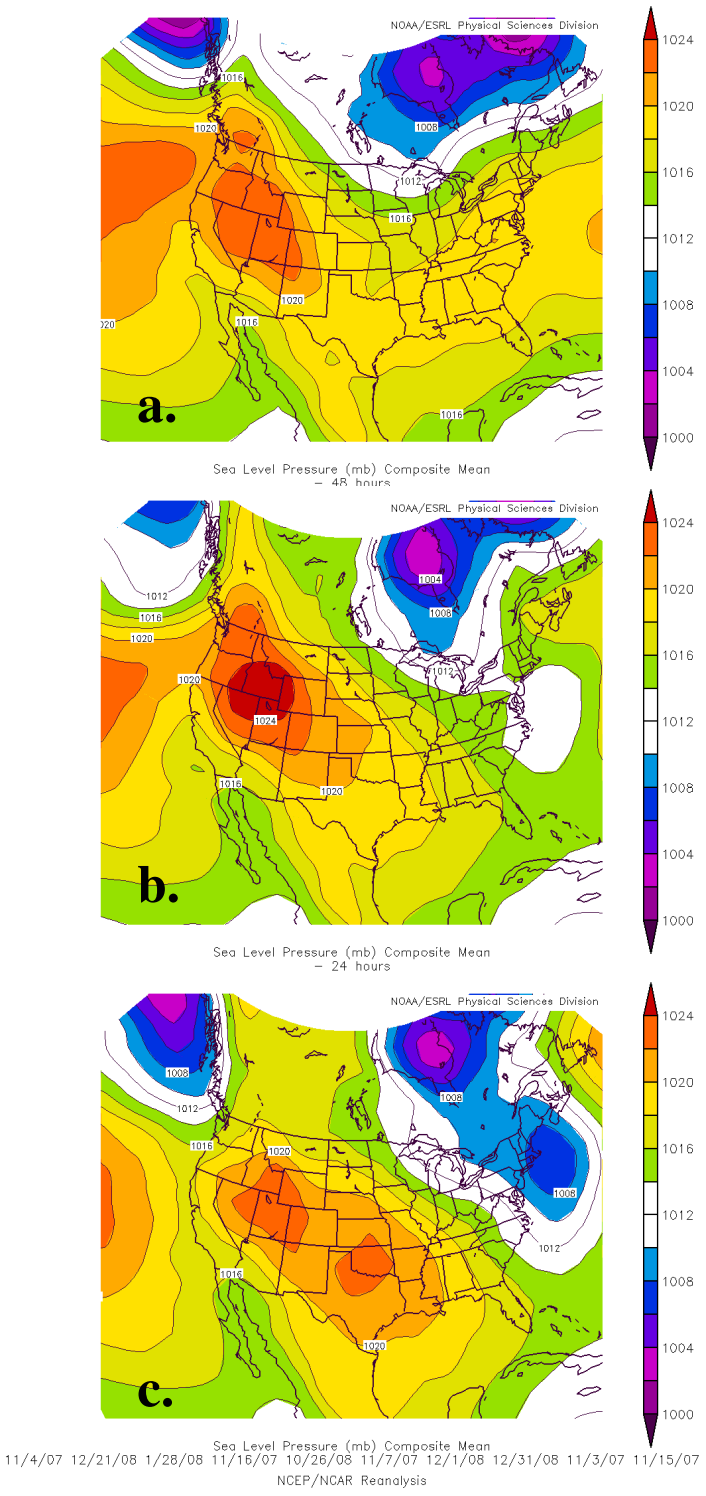


Figure 4.15. (a) NCAR-NCEP reanalysis sea-level pressure (SLP; shaded) composite for the 10 largest negative storm surge error day two (24-48 h) forecasts for SBSS control member 9a at the Battery. (b) Same as (a) except composite for 24 hours prior to largest negative error day. (c) Same as (a) except for 48 hours prior.

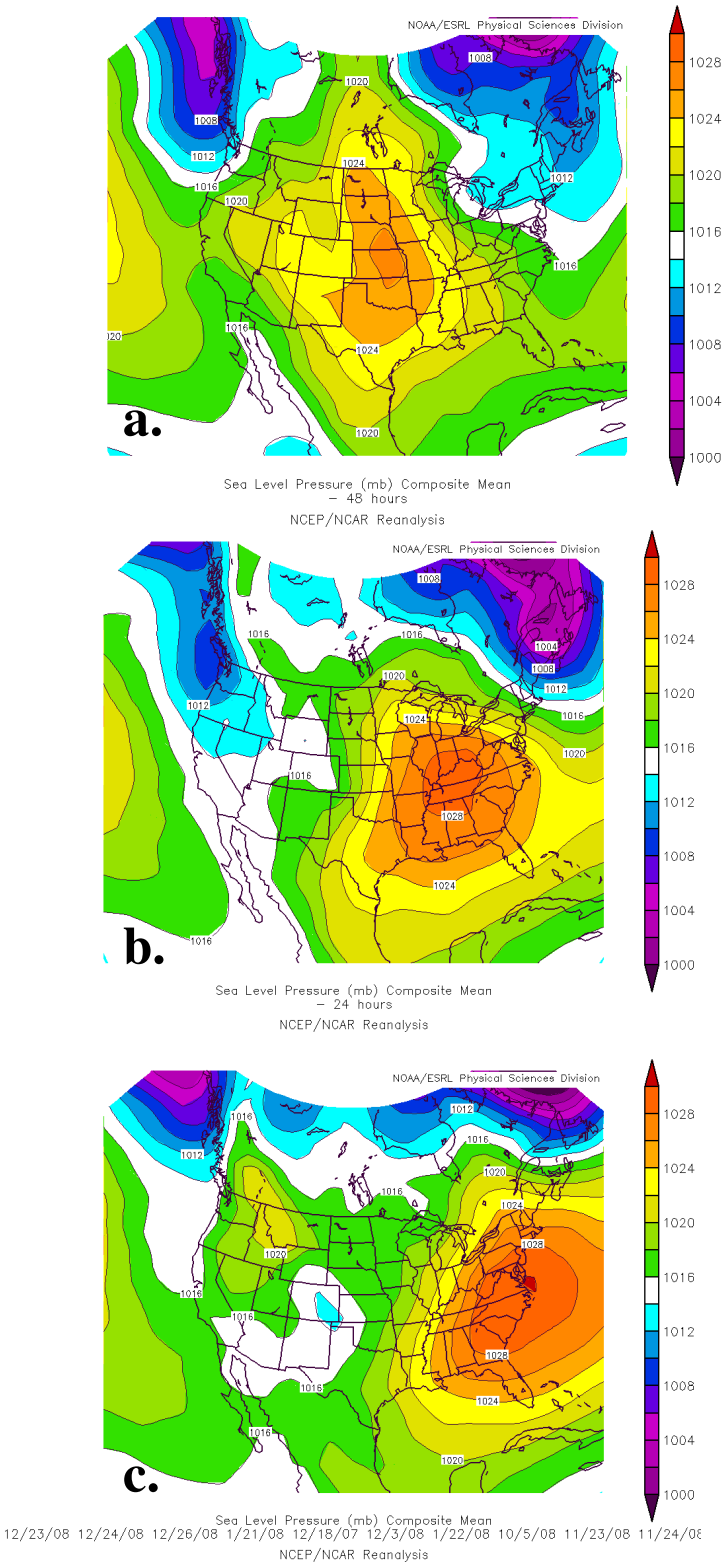


Figure 4.16. Same as figure 4.13 except for the 10 largest positive storm surge error day 2 (24-48 h) forecasts for SBSS control member 9a at the Battery.

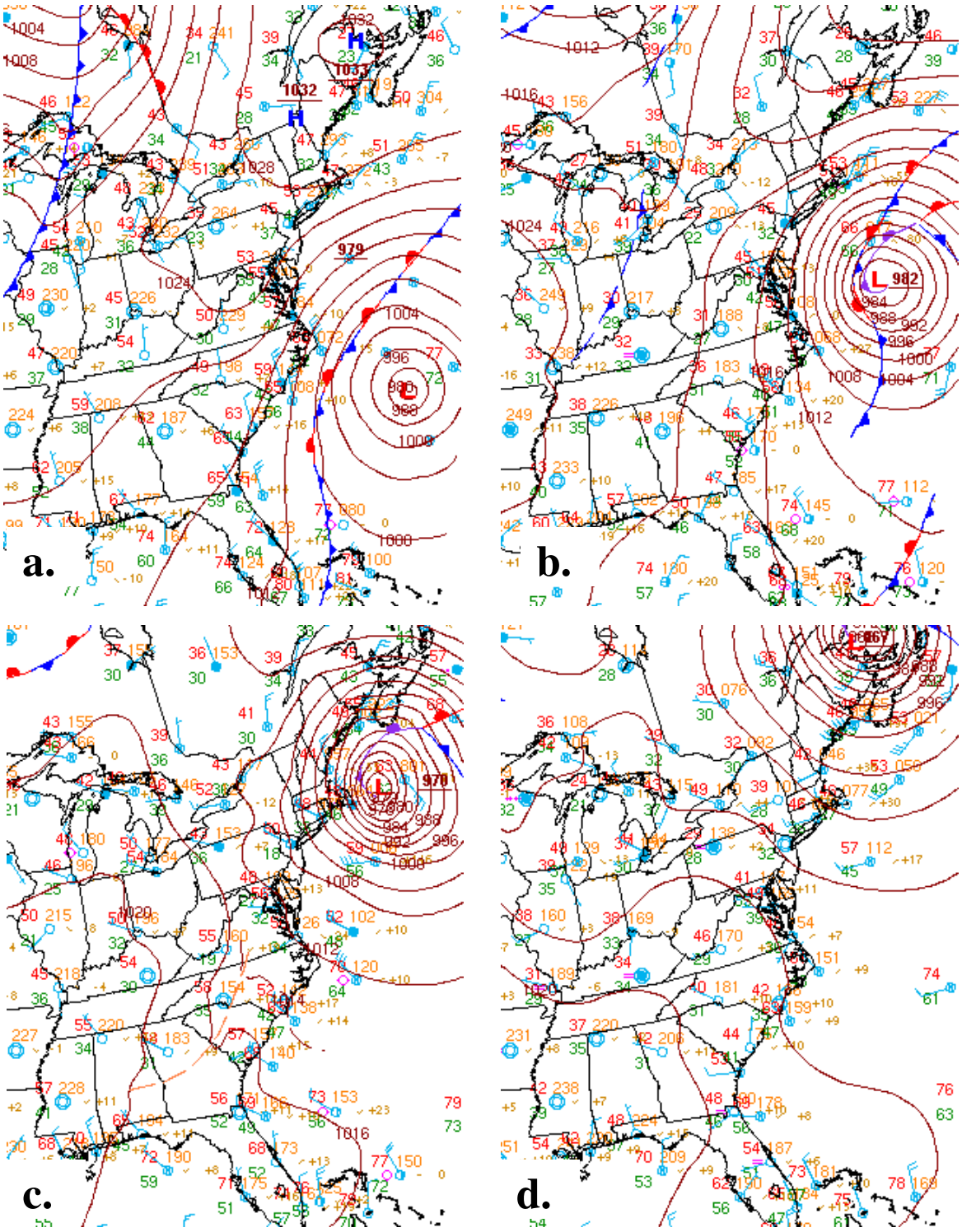


Figure 4.17. (a) NOAA-National weather service surface analysis at 0000 UTC 3 November with SLP (contoured), station plots, winds (full barb = 10 kts), and frontal analysis. (b) Same as (a) except for 1200 UTC 3 November 2007. (c) Same as (a) except 0000 UTC 4 November 2007. (d) Same as (a) except 1200 UTC 4 November 2007.

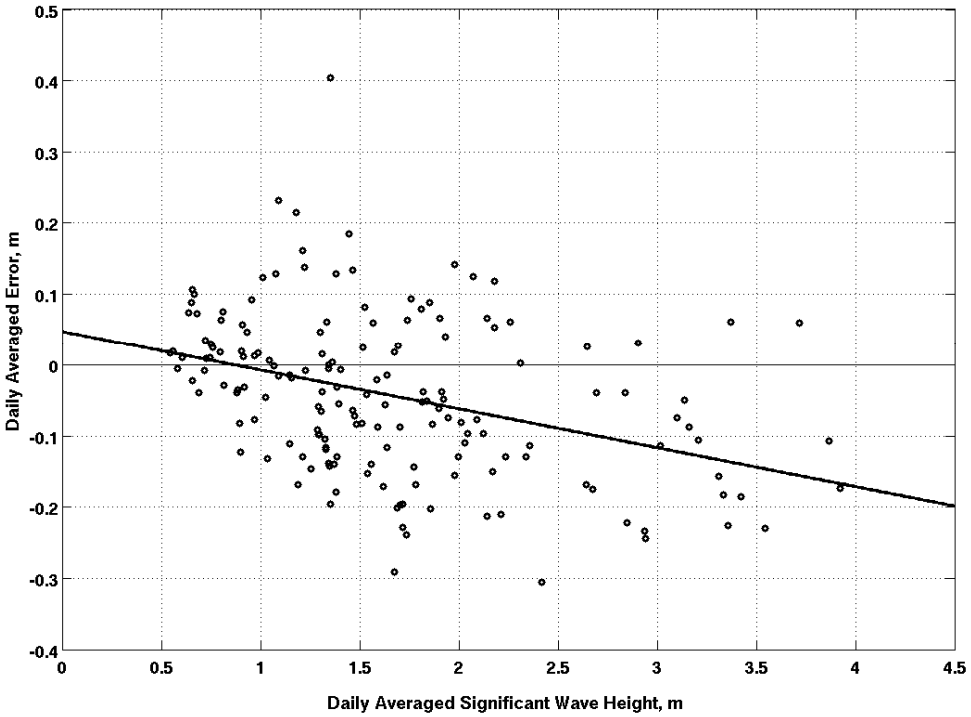


Figure 4.18. Averaged daily significant wave height at buoy 44017 compared with daily averaged mean error at Montauk Point, NY for the same corresponding days. See figure 4.1 for buoy 44017 location.