Simulating storm surge and compound flooding events with a creek-to-ocean model: Importance of baroclinic effects

Fei Ye, Yinglong J. Zhang, Haocheng Yu, Weiling Sun, Saeed Moghimi, Edward Myers, Karinna Nunez, Ruoyin Zhang, Harry V. Wang, Aron Roland, et al.

Simulating storm surge and compound flooding events with a creek-to-ocean model: importance of baroclinic effects

Fei Ye[a1], Yinglong J. Zhang[a], Haocheng Yu[a], Weiling Sun[a], Saeed Moghimi[b], Edward Myers[b], Karinna Nunez[a], Ruoyin Zhang[c], Harry V. Wang[a], Aron Roland[d], Kevin Martins[e, f], Xavier Bertin[f], Jiabi Du[g], Zhuo Liu[a]

a. Virginia Institute of Marine Science, College of William & Mary, Gloucester Point, VA 23062, USA

b. Coast Survey Development Laboratory, National Oceanic and Atmospheric Administration, Silver Spring, MD 20910, USA

c. State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing 100084, China

d. Institute for Hydraulic and Water Resources Engineering, Technische Universität Darmstadt, Darmstadt, Germany.

e. EPOC (CNRS-Université de Bordeaux) - UMR5805, Allée Geoffroy Saint-Hilaire, 33615 Pessac, France

f. LIENSs (CNRS - La Rochelle Université) - UMR7266, 2 rue Olympe de Gouges, 17000 La Rochelle, France

g. Department of Marine Sciences, Texas A&M University at Galveston, Galveston, TX 7754, USA

Abstract

We present a creek-to-ocean 3D baroclinic model based on unstructured grids that aims to unite traditional hydrologic and ocean models in a single modeling platform, by taking full advantage of the polymorphism (i.e. a single model grid can seamlessly morph between full 3D, 2DV, 2DH and quasi-1D configurations). Using Hurricane Irene (2011)'s impact on the Delaware Bay as an example, a seamless 2D-3D model grid is implemented to include the entire US East Coast and Gulf of Mexico with a highly resolved Delaware Bay (down to 20-m resolution). The model is forced by flows from a hydrological model (National Water Model) at the landward boundary. We demonstrate the model's accuracy, stability and robustness with the simulation of the storm surge and subsequent river flooding events and compound surges. Through a series of sensitivity tests, we illustrate the importance of

[1] Corresponding author; e-mail: feiye@vims.edu; phone: (804) 684-7723; fax: (804) 684-7179.

including in the simulation the baroclinic effects, as provided by the large-scale Gulf Stream, in order to correctly capture the adjustment process following the main surge and the subsequent compound flooding events. The baroclinicity can explain up to 14% of the elevation error during the adjustment phase after the storm.

Keywords: Storm surge; 3D model; baroclinicity; SCHISM; National Water Model; Delaware Bay, USA

1 Introduction

The catastrophic loss from floods world-wide costs hundreds of billions of dollars each year, including property damages and loss of lives (Blake 2007). The flood and inundation hazard due to either coastal storm surges or river flooding has been well studied using physical and statistical models (Wolf 2009; Teng et al. 2017). However, a critical knowledge gap exists in the attempt to address the effects of compound flooding due to the combined effects of different flood sources.

In general, three types of inundation processes are of concern to coastal communities: coastal storm surge and inundation, pluvial inundation (precipitation driven flooding), and fluvial inundation (river flooding). Traditionally, these processes have been studied separately using different types of models: hydrodynamic models for storm surge and hydrological/hydraulic models for pluvial/fluvial flooding. A hydrodynamic model generally excludes the watershed mostly because of the constraints induced by numerical stability and/or computational cost, whereas a hydrologic model's capability roughly stops at mean sea level (MSL) because it is not designed for estuarine/oceanic processes. A plethora of models are available for these two types of simulations (e.g., Kerr et al. 2013; Chen et al. 2013; Sharma et al. 2019). These models, however, often neglect the important interaction between coastal and inland processes such as the compound flooding and backflow effects. The compound effects from the combination of all these processes, including the 3D baroclinic effects, have not been studied in detail before and are the subject of this study.

The issue of compound flooding has recently garnered a great deal of attention because of the increased concurrences of storms and heavy precipitation in coastal areas (Wahl et al. 2015). For example, during Hurricane Harvey (2017), Galveston Bay received freshwater inputs of approximately three times the bay's volume (Du et al., 2019), causing catastrophic flooding along the Texas coast. As the climate warms, many of the climate models predict increasing occurrences of such 'wet' storms in the coming decades (Knutson et al 2010), which sets the perfect conditions for compound flooding.

This trend highlights the urgency of understanding the detailed mechanism of compound flooding to accurately forecast its impact, for effective mitigation and planning.

To this end, National Oceanic and Atmospheric Administration (NOAA) has initiated the NOAA Water Initiative[2] to understand, develop, demonstrate and implement an improved total water level prediction that includes signals propagating up and downstream in coastal, estuarine and riverine environments, particularly during storm events. The goal is to establish an integrated water forecasting system that covers inland and coastal waters to provide vital information to end-users and stakeholders. As our contributions to the Water Initiative, this study uses a seamless creek-to-ocean 3D baroclinic unstructured-grid (UG) model (SCHISM, schism.wiki, last accessed in July 2019; Zhang et al. 2016), driven by the predicted flows from a continental-scale hydrological model (National Water Model[3], or "NWM"), to holistically simulate the total water level and its individual components. The flexibility and robustness of SCHISM are indispensable for traversing large contrasts of temporal and spatial scales from oceanic processes (such as the Gulf Stream) to inland flooding in small creeks. In doing so, we have built a very reconfigurable and flexible modeling platform that can be extended to cover a larger domain with sufficient resolution in the areas of interest. Effort is on-going to extend the current model to cover all major estuaries and bays in the US East Coast and Gulf of Mexico.

In this study, we focus on the importance of including the 3D effects in the storm surge simulations. The model is first validated using observational data derived from NOAA, USGS and satellite products. Sensitivity tests are then conducted to examine the importance of 3D effects, in particular, the baroclinicity. Minato (1998) was among the first to elucidate the implication of incorporating 3D effects in simulating storm surges, and he gave a simple explanation on why finer vertical resolution led to higher surges in Tosa Bay (Japan). Zheng et al. (2013) compared results from 2D and 3D barotropic models for the storm surges in the Gulf of Mexico and showed that both models can adequately simulate the surge provided that the bottom frictions were properly calibrated. Li et al. (2006) and Cho et al. (2012) studied the impact of Hurricane Floyd (1999) and Isabel (2003) in the Chesapeake Bay using 3D baroclinic models but did not explicitly expound the influence of baroclinicity on surface elevation, partly because of the small model domain used. Orton et al. (2012) used a 3D baroclinic model on a small domain to simulate storm surges near New York City and estimated that neglecting water density variations led to typical reductions of 1–13% in the peak surge. But as suggested by Zheng et al. (2013), these numbers are highly sensitive to the choices of bottom friction; in addition, the

[2] url: https://www.noaa.gov/water/explainers/noaa-water-initiative-vision-and-five-year-plan, last accessed in September 2019.
[3] url: https://water.noaa.gov/about/nwm, last accessed in July 2019

exclusion of large-scale baroclinic processes such as the Gulf Stream led to additional uncertainties in their model. Therefore, the impact of 3D baroclinic effects on storm surge needs to be further assessed.

One of the key differences between the current study and previous studies is that we compare the 'best calibrated' results from different configurations (2D, 3D barotropic and 3D baroclinic) by adjusting model parameters independently to achieve best possible results under each configuration. This ascertains that the differences are not due to calibration issues or of numerical origin. Our results from numerous sensitivity tests indicate that the 3D baroclinic effects do not significantly alter the main surge (as the latter is mostly governed by large-scale barotropic processes) but play an important role in the restoration process afterward, mainly through the large-scale oceanic response as found in the Gulf Stream. Results from sensitivity tests clearly indicate that the restoration process in the water surface elevation ('rebounding waves') cannot be properly captured by barotropic models. The importance of the Gulf Stream on coastal inundation has been reported recently by Ezer (2013, 2018) but the focus there was on the remote connection between storms and coastal flooding far away from the storm path. Our results provide direct evidence on the importance of including the baroclinic oceanic response in storm surge simulations for bays and estuaries.

In what follows, Section 2 briefly describes the hurricane event (Irene 2011) used in this study as well as available observational data collected by multiple agencies. Section 3 presents the details of our model setups including the baseline simulation using a 3D baroclinic configuration, as well as other sensitivity simulations. Section 4 presents the validation of the baseline setup for surface elevation, salinity and temperature. Section 5 discusses the effects from wind waves. Then in Section 6, we use results from sensitivity experiments to elucidate the effects from 3D processes including baroclinicity, highlighting the stabilization effect of the Gulf Stream in the restoration process after the storm. A summary is given in Section 7.

2 Study case

2.1 Hurricane Irene (2011)

Hurricane Irene, the first major Atlantic hurricane in 2011, was selected here as a case study. The hurricane made its landfall along the US East Coast at Outer Bank, NC on Aug 27, 2011 as a Category 1 hurricane. The hurricane re-entered Atlantic coastal waters in Virginia, Delaware and New Jersey, and weakened to a tropical storm before making the second landfall in New Jersey and the third landfall in New York City. As seen from the weather maps (Fig. 1), while there was only one major wind event related to Irene, there were two large precipitation events associated with Irene and the subsequent

Tropical Storm Lee; the latter originated from Gulf of Mexico and swept over land over eastern states. These events led to two streamflow peaks in Delaware Bay's largest tributary, the Delaware River (Fig. 2b). The first peak flow of about 4000 $m^3 s^{-1}$ occurred on Aug 28, 2011, around the landfall of Irene. The second peak, about 1500 $m^3 s^{-1}$ higher than the first, occurred 10 days later under Tropical Storm Lee (2011), mainly resulting from the precipitation and the subsequent pluvial/fluvial processes rather than a storm surge. The streamflow of the second largest tributary, the Schuylkill River, exhibits similar patterns (Fig. 2b). The different flood origins (ocean and inland) made this event an ideal test for compound-flood modeling.

2.2 Observation

The observational datasets used in this study included 9 tide gauges maintained by NOAA[4], 2 streamflow gauges from USGS[5], and 2 wave buoys from NOAA's National Data Buoy Center (NDBC)[6]. The locations of these stations are shown in Fig. 3. In addition, salinity survey data inside the Delaware Bay collected during a 1984 intensive survey was used to assess the model's performance in simulating 3D baroclinic processes, because of the lack of salinity observation during Irene. Also, NASA's GHRSST Level 4 G1SST Global Foundation Sea Surface Temperature Analysis[7] was used to assess the model skills for large-scale processes in the open ocean, including the Gulf Stream. The use of all these observation datasets ensured a thorough assessment of the model skill.

3 Numerical model

3.1 Model and domain

SCHISM is a flexible, primitive equation, hydrostatic model grounded on hybrid Finite-element/Finite-volume method and hybrid triangular-quadrangular UGs in the horizontal and hybrid Localized Sigma Coordinates with Shaved Cells (LSC^2) grid in the vertical (Zhang et al. 2015, 2016). The model uses a semi-implicit time stepping scheme to enhance robustness and efficiency, and the numerical dissipation is kept low with a judicious combination of higher-order, monotone schemes (Ye et al. 2018; Ye et al. 2019) and the semi-implicit finite-element formulation.

A SCHISM-based regional model has been developed, which covers the Northwest Atlantic Ocean, the Gulf of Mexico, and the Caribbean Sea (Fig. 3a). The large spatial domain used here accommodates

[4] url: https://tidesandcurrents.noaa.gov/tide_predictions.html/, last accessed in June 2019.
[5] url: https://waterdata.usgs.gov/nwis/uv/, last accessed in June 2019.
[6] url: https://www.ndbc.noaa.gov/, last accessed in June 2019.
[7] url: https://podaac.jpl.nasa.gov/dataset/JPL_OUROCEAN-L4UHfnd-GLOB-G1SST, last accessed in June 2019.

different storm paths that make landfalls in this region and includes the path of the most important western boundary current in this region, the Gulf Stream, in order to study its baroclinic responses during storms (Ezer 2018, 2019).

The focus site in this pilot study is the Delaware Bay, which is a major estuary on the U.S. East Coast. Tidal portions of the Delaware River start at Trenton, New Jersey, and the total length of the estuary from Trenton to the mouth is ~210 km, draining a watershed area of 4.2×10^5 km^2 (Sharp 1984). A key characteristic of the estuary, as far as the tides are concerned, is that it constricts rapidly from the widest point in the lower Bay (near the NOAA station Brandywine; Fig. 3b) to the mid-Bay. The funneling effect from width variation leads to a 'hypersynchronous' system with tides generally increasing landward (Friedrichs 2010), but the channel meandering in the mid-Bay (near the NOAA station Reedy Point; Fig. 3b) considerably complicates the dynamics there (Section 4.1). The Bay is mostly shallow with a mean depth of 7 m (Harleman 1966). The Delaware River provides 58% of the freshwater inflow; the confluence of the Schuylkill River below Philadelphia adds another 14%; other tributaries collectively account for the remaining 28% (Sharp 1983). The average freshwater inflow is 570 m^3 s^{-1}, with high flow conditions occurring during the spring freshets (Whitney and Garvine 2006). Even under peak freshwater inflow, the estuary is vertically mixed by the tides most of time (Wong 1995). The freshwater outflow from the Bay is generally weak; the plume is typically in contact with the frictional bottom boundary layer and confined within 20 kilometers offshore under normal conditions (Münchow and Garvine 1993; Wong and Münchow 1995; Yankovsky and Chapman 1997).

The model domain includes a part of the Delaware Bay watershed, up to the 10-m isobath above mean sea level (MSL). In the Delaware River (the largest tributary of the Delaware Bay), the domain extends to the USGS gauge of Riegelsville (Fig. 3b) at 40 m above MSL. Such a domain choice requires that the model directly simulates some pluvial and fluvial processes normally handled by hydrological models. The bathymetry information (Fig. 3) is derived from two DEM (Digital Elevation Model) sources: the global relief model ETOPO1 (Amante and Eakins, 2009) for the ocean, and the 1-m USGS Coastal National Elevation Database (Danielson, 2018) for the Delaware Bay. The inclusion of a large portion of Delaware Bay watershed complicates the use of a curved vertical datum such as NGVD29 and therefore, the model is based on a flat datum of NAVD88. The latter is convenient because (1) the USGS DEM is given in NAVD88; (2) most new instruments use this datum. Accordingly, all model-data comparsions on elevation are also based on NAVD88. A datum conversion from NAVD88 to NGVD29 (e.g. using the VDatum[8] tool, which itself relies on model simulation)

[8] url: https://vdatum.noaa.gov/, last accessed in June 2019.

would introduce uncertainties especially in the upper Bay near the fall line, so we choose to rely on observation instead. Among the nine NOAA tide stations used (Fig. 3b), the differences between the local MSLs and the NAVD88 datum are available at three stations: the lower Bay station "Lewes" (-0.121 m; i.e., the local MSL is below the NAVD88 datum), the mid-Bay station Reedy Point (-0.015 m), and the upper Bay station Philadelphia (0.118 m). As shown later (Section 4.1), the model is able to correctly set up the surface slope, with the local MSL asymptotically increasing toward upper Bay as observed. The datum differences between MSL and NAVD88 at the three stations are linearly interpolated onto other stations based on the along-thalweg distance (with linear extrapolation for Berlington and Newbold, which are up-estuary from Philadelphia). This procedure is expected to introduce uncertainties on the order of a few centimeters in the model-data comparison of elevation.

3.2 Grid generation

One of the key steps and challenges in UG modeling is the grid generation. The stability and robustness of SCHISM greatly simplifies the grid generation process: effort is mostly focused on resolving key bathymetric and geometric features without worrying about computational cost or numerical instability, courtesy of the implicit scheme used. Furthermore, polymorphism allows a very faithful representation of the underlying bathymetry and topography without the need for any smoothing as required by many other terrain-following coordinate models (Zhang et al. 2016). As explained in Ye et al. (2018), bathymetry smoothing in an estuarine regime should be avoided, because it alters fundamental aspects of estuarine circulation such as salt intrusion, channel-shoal contrast, and the related lateral circulation.

The grid generation software SMS (Surface-water Modeling System)[9] was used to generate the horizontal grid. The horizontal spatial domain was discretized by an UG with 667 K nodes and 1,273 K elements, including 39 K quadrangular elements used to represent the shipping channel (Fig. 4). A quasi-uniform triangular grid with a resolution of 6-7 km was applied in the open ocean, which was smoothly transitioned to about 2-km resolution near the coastline. Locally high resolution was applied in the Delaware Bay, with a typical resolution of 600 m in the lower Bay channel, 50 m in the upper Bay channel, 150 m in the watershed areas above MSL, and down to 20 m in some small creeks. During the grid generation process, "feature arcs" (Fig. 4a) were used mainly to (1) explicitly incorporate 'features' such as the NWM segments into SCHISM's horizontal grid (red arcs in Fig. 4b); (2) align the quadrilateral elements with main channels thus resembling a structured grid locally (black

[9] url: https://aquaveo.com/, last accessed in June 2019.

arcs in Fig. 4b). After the grid was generated, the DEMs were linearly interpolated onto the computational grid *without* any bathymetry smoothing.

The vertical discretization used in the model took full advantage of the hybrid terrain-following-like LSC^2 coordinate (Zhang et al., 2015), with variable number of layers at different horizontal locations. The average number of layers was 18.3, with a maximum of 44 layers in the deepest ocean and only 1 layer in shallow areas with depths less than 0.5 m (Fig. 5). As a result, 2D representation was applied for about 57% of the Delaware Bay watershed or 30% of the total grid elements. An element was deemed wet when the local water depth, calculated from the implicit finite-element solver, exceeded 10^{-6} m; such a small threshold was needed to accurately capture the very thin layer of fluid initially formed on dry land during precipitation events. A smaller value (10^{-8}) does not significantly change the results. The model was stable even with this choice of small threshold for wetting and drying, courtesy of the implicit scheme. Because the vertical flow structure can be safely ignored for the pluvial and inundation processes, the single-layer configuration greatly reduced the computational cost and meanwhile enhanced the robustness of the model (Zhang et al. 2016).

3.3 Baseline model setup

3.3.1 Forcing and parameters

The "baseline" setup used a 3D baroclinic model. Wave effects were excluded in the baseline model but examined in subsequent sensitivity analysis (Section 5). Atmospheric forcing applied at the air-sea interface consisted of two sources. The first source was derived from ECWMF's ERA5 reanalysis dataset[10], and the variables included air temperature, air pressure (reduced to MSL), humidity, wind speed and direction at 10 m above MSL, downward short-wave and long-wave radiations, and precipitation rate. This product has a spatial resolution of 30 km and temporal resolution of 1 hour. The second source was a high-resolution product from ECMWF, with a spatial resolution of 5 km (Magnusson et al. 2014). The comparison shown in Fig. 2 suggests that the atmospheric forcing used in the model is sufficiently accurate for simulating the storm surge. Occasionally, there are noticeable mismatches (about 4 m s^{-1}) in wind speed between the forcing and the observation, e.g., at the peak of Irene and one day afterwards (Day 32-33 in Fig. 2a) and during the river flooding period (Day 42.5 in Fig. 2a). These mismatches contribute to the uncertainties in the simulated water level. The surface wind stress and heat exchange were calculated from the bulk aerodynamic model of Zeng et al. (1998).

[10] url: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5, last accessed in June 2019.

Other parameterizations for surface stress, e.g. from the bulk formulas of Pond & Pichard (1998), Hwang (2018), or from the wave model (Ardhuin et al. 2010), yielded similar results.

A major calibration parameter for surface elevation was the bottom friction. Although some information on bottom sediment composition was available and suggested different bottom characteristics in the lower and upper Bay (Gebert & Searfoss 2012), the "total" bed roughness that accounts for macro roughness such as bed form was not available and it is also expected to be temporally varying. Therefore, the roughness used in the model was selected through calibration. We used a bottom roughness of 0.5 mm in the ocean and the lower Delaware Bay, and then transitioned it to 0.05 mm in the mid- and upper Bay along the main channel; in upland areas (3 m above MSL), we used a uniform 1-mm roughness.

A bi-harmonic viscosity was added to the horizontal momentum equation (Zhang et al., 2016) to control the spurious inertial modes that often arise in large-scale UG models (Le Roux et al., 2005; Danilov, 2012). In addition, a Laplacian viscosity in the form of Shapiro filter (Shapiro 1970; Zhang et al. 2016) was locally added for the steep bathymetry (Fig. 6), where spurious modes would otherwise be exacerbated by the pressure gradient errors. The specified Shapiro filter strength (γ; non-dimensional; Zhang et al. 2016) was a function of the local bathymetric slopes (α; non-dimensional), expressed as $\gamma = 0.5 \tanh(\alpha/\alpha_0)$, where α_0 was a reference slope chosen as 0.5 in the current setup. This led to a maximum filter strength of 0.5 for the steepest bathymetric slope. Horizontal diffusivity was not explicitly added because the 3rd-order WENO transport scheme used is essentially monotone (Ye et al. 2019). In shallow waters with depths less than 5 m, the 3rd-order WENO scheme was replaced by a 1st-order but more efficient upwind scheme. The vertical viscosity and diffusivity were calculated by the generic length-scale model (*k-kl*; Umlauf and Burchard, 2003).

The model was initialized from the HYCOM reanalysis product on July 27, 2011 and run with a non-split time step of 150 seconds in a fully implicit mode (i.e. with implicitness factor of 1). An implicitness factor of 0.6 gives essentially the same results. The boundary conditions (B.C.) for temperature, salinity, and the sub-tidal components of the sea-surface height (SSH) and horizontal velocity along the ocean boundary was also derived from HYCOM. Note that HYCOM used an unknown vertical datum, and therefore the SSH was adjusted by +0.5 m based on the calibration results at a coastal gauge (Lewes, DE). The tidal components of the B.C. for the elevation and barotropic velocity were then added using the FES2014 product (Carrere et al. 2016). To prevent long-term drift, the tracer field (salinity and temperature) was relaxed to HYCOM results in a region within ~1° from the ocean boundary, with a maximum relaxation constant of 1 day.

The simulation period starts on July 27, 2011, i.e. one month before Irene's landfall on the US east coast. One month was deemed sufficient for spin-up, because the initial conditions were from the fully dynamic conditions provided by the data-assimilated HYCOM product (cf. Zeng and He 2016); separate runs with an additional 30-day spin-up led to essentially same results (not shown). The simulation covered 50 days that included the main surge and the subsequent river flooding events. For the purpose of salinity validation, we used another period in 1984 (cf. Section 4.3) because salinity observation is unavailable during the Irene period.

3.3.2 Coupling to NWM

The freshwater delivery into the Delaware Bay was derived from the NWM, i.e. a reanalysis product[11] from NWM v1.2, provided by the NOAA team. We first determined all NWM segments that intersected the SCHISM land boundary, and then the streamflow at each segment was then imposed in the adjacent SCHISM elements as a point source (for inflow segments) or sink (for outflow segments) (Fig. 7). Implementation of volume and mass sources and sinks inside SCHISM was rather straightforward via simple volume integrations in the finite-element equations. Although we have only considered the one-way coupling between NWM and SCHISM so far, the simplicity of this coupling strategy bodes well for the eventual two-way coupling between the two models. The precipitation effects are not discussed in this paper and are left for a future study.

3.4 Sensitivity runs

Sensitivity runs were conducted to examine the importance of different processes, including baroclinic response, 3D barotropic processes, wind wave effects, etc. Table 1 shows the setups of important sensitivity runs used in this paper. In the "3D Barotropic" run, tracer transport was turned off and the baroclinic force from temperature and salinity gradient was excluded. To assess the importance of 3D barotropic processes, a "2D Barotropic" run was also included. As explained in Zheng et al. (2013), both 2D and 3D models can simulate a storm surge well, if bottom friction coefficients are properly adjusted in each model. The near-bottom velocities in 2D and 3D models differ by orders of magnitude, so drag coefficients need to be adjusted accordingly to achieve comparable bottom stress, which is one of the key controls for surges. In general, 3D baroclinic models should use a much smaller coefficient than 2D models, because the former produce larger near-bottom velocity inside stratified regions due to the two-layer exchange flow. This considerably complicates the inter-comparison of 2D and 3D models (Zheng et al. 2013).

[11] url: https://registry.opendata.aws/nwm-archive/, last accessed in June 2019.

A key difference between the current and previous studies is that we calibrated the three configurations (Table 1) of the model separately to achieve best possible results (in terms of overall Mean Absolute Error, or 'MAE') before the results were inter-compared. Different choices of surface stress formulations, time steps, and other parameters were tested, but the dominant control was found to be the bottom friction. This approach largely removes the ambiguity of the bottom friction parameterizations used in each configuration and ensures that the findings in subsequent sections are not unduly influenced by parameter choices. For the 2D model, we used a uniform Manning's n of 0.019 s $m^{-1/3}$; for the 3D barotropic model, the same roughness height as in the "baseline" was found to give best results.

Wave effects are also important in nearshore regions (Kennedy et al. 2012; Guerin et al. 2018). In particular, wave breaking induces a setup near the shoreline and embayment and alters the mean circulation nearshore. To examine this, we included a run called "base+wave", where the base model was fully coupled with the Wind Wave Model (WWM; Roland et al., 2012) on the same horizontal grid. The wave effect was incorporated into SCHISM via the 3D vortex formalism of Bennis et al. (2011) as implemented and validated in Guerin et al. (2018). In addition, the formulation of wave-enhanced bottom boundary layer was from Soulsby (1997) and the formulation of wave breaking induced turbulence followed that of Craig and Banner (1994). The wave model was initialized using a global hindcast product based on WWIII[12] (Rascle and Ardhuin, 2013), and was also forced at the ocean boundary by the same product; five variables (significant wave height, peak frequency, mean wave period (TM02) and direction, and directional spread) were extracted to construct wave spectra at the ocean boundary using the JONSWAP formula (Hasselmann et al. 1973). WWM was run with the implicit mode, and the coupling time step was set to be 600 s. Thirty-six bins were used to resolve the directional and frequency domain.

Table 1: Sensitivity runs

Run name	**Set-up**
baseline	3D baroclinic (cf. Section 3.3)
3D barotropic	Barotropic, otherwise same as "baseline"
2D barotropic	Based on "3D barotropic", but using a single vertical layer and a different bottom friction formulation (Manning's n)
base+wave	Two-way coupling with the wind wave model (WWM), otherwise same as "baseline"

[12] url: ftp://ftp.ifremer.fr/ifremer/ww3/HINDCAST, last accessed in June 2019.

4 Model validation

In this section we first validate the "baseline" model using the observational datasets shown in Section 2.2. The elevation skills at all NOAA tide stations in the Delaware Bay are examined first, followed by salinity inside the Bay and sea surface temperature (SST) in the ocean; the last two are important for the baroclinic processes.

4.1 Surface elevation

The simulated total water levels generally agree well with the observations (Fig. 8). The mean absolute error (MAE) and correlation coefficient (CC), averaged over all stations, are 13 cm and 0.98 respectively. Larger errors are found at upstream stations, most likely caused by a combination of model errors, uncertainties/errors in the DEM and the NWM-predicted flows (cf. Fig. 2). These uncertainties generally have more impact on the narrower and shallower channels in the upper Bay than on the wider and deeper lower Bay. The average MAE in the lower and mid-Bay (first 5 stations) is only 9 cm. Most interestingly, even though the model has errors in predicting both the primary first surge (around Day 32.5) and the second surge (around Day 44.5) at the most upstream station (Newbold), it correctly predicts that the second surge is higher than the first surge there (Fig. 8). The observation suggests that the second surge is 5 cm (versus 11 cm as suggested by the model) higher than the first, indicating very strong river influence in the upper Delaware Bay. On the other hand, this strong river influence also implies that the errors in NWM (Fig. 2) explain part of the model errors.

A tidal harmonic analysis is also conducted on major tidal constituents, including M_2, K_1, O_1, and N_2 (Fig. 9), from Day 10 to Day 50 of the simulation period. Although this period includes flooding/surge, the model data comparison is still valid because the same period is applied to both model and data. The "baseline" again shows good agreement with observation. The average MAEs for the M_2 constituent (which accounts for 93% of the total tidal energy) of all stations are: 4.0 cm in amplitude and 7.5 degree (15.5 minutes) in phase. The observed M_2 amplitude increases from the mouth to the mid-Bay, and then slightly decreases from the mid-Bay from Ship John Shoal to Reedy Point (see station locations in Fig. 3b), and increases again into the upper Bay. This longitudinal variability is attributed to the balance between the funneling effects due to the trumpet shape of the channel and the bottom friction (Du et al. 2018), but the meandering near Reedy Point further complicates the dynamics and leads to the local extrema. In general, the model captures this trend, but has slightly larger errors near the mid-Bay extrema than elsewhere. The model also tends to lag the observation at upper Bay stations.

4.2 Sea-surface temperature (SST)

As explained by Ezer (2013, 2018, 2019), Gulf Stream, as a major western boundary current, plays an important role in the coastal response to global sea-level rise and tropical cyclones. In particular, the weakening of the Gulf Stream transport is often responsible for 'fair weather' flooding events along Mid-Atlantic Bight (MAB) coast.

To ensure that the baroclinic processes in the open ocean (mainly controlled by temperature gradient) are well captured by the baseline model, the simulated SST is compared with a satellite derived product (NASA's GHRSST Level 4 G1SST). During the passage of Hurricane Irene, the Gulf Stream is greatly disturbed, as evidenced by the fluctuations in its volume transport (cf. Fig. 17b). The restoration of the coastal ocean takes several days (Ezer 2018). By Sept 7, 2011 the Gulf Stream is largely restored to its pre-storm condition (Fig. 10). SCHISM qualitatively captures the restored SST condition, with the free meandering north of Cape Hatteras slightly improved from that in HYCOM. Also apparent in both observation and model results is a cold wake in the middle of Atlantic left by another storm (Hurricane Katia; also see the bottom-left panel of Fig. 1), which did not land on the US east coast. The complex eddies and meanders as shown in Fig. 10 cannot be reproduced by barotropic models and are responsible for setting up large surface slopes along the path of the Gulf Stream (cf. Fig. 18 and Fig. 19), which has implications for the rebounding water level inside the Bay (Section 6). More detailed assessment of the model skill including the vertical structure of the Gulf Stream has been reported in Ye et al. (2019).

4.3 Salinity

Delaware Bay is a weakly stratified estuary with a nearly linear axial salinity distribution (Garvine et al. 1992). Previously, a multi-model comparison for this system using the observational data collected during the1984 hydrographic surveys has been shown in Patchen (2007). During the surveys, several stations reported salinity measurements, three of which had data at more than one depth. Note that the measurements at different stations were collected at different time periods.

Because of the lack of salinity measurements during Irene, we re-ran the base model for the period in 1984 when intensive hydrographic surveys were conducted by NOAA. The model was initialized on April 2, 1984 and run for 100 days. The comparison shown in Fig. 11 indicates that the model is able to capture the salinity variation (with an averaged MAE of 0.86 PSU overall) as well as the larger stratification found in the mid-Bay. Therefore, the model can capture the 3D density structure inside the Bay, which is important for simulating the baroclinic response. The depth averaged salinity in the Bay from the last 70 days of the model results is shown in Fig. 12. The pattern is qualitatively similar to that

in Whitney and Garvine (2006) and shows clear lateral gradients between channel and shoal. As discussed in Garvine et al. (1992), the lateral variation of salinity generally exceeds its weak vertical stratification, which is one of the main reasons that the salt intrusion in the Bay is relatively insensitive to the variations in the river inflow.

5 Wind wave effects

To assess the wave effects on the water surface elevation, we restarted the baseline simulation on August 21, 2017 and ran the "base+wave" (Table 1) model for 10 days. We first validate the "base+wave" model using the observation at two nearby NDBC buoys (see locations in Fig. 3). The modeled significant wave heights and peak periods matches the observation well (Fig. 13), with the average MAEs being 24 cm for the significant wave height and 2.1 seconds for the peak period respectively. During the storm, large waves of relatively longer periods (Fig. 13) entered the Bay from the south and large wave breaking occurred near the steep bathymetric slopes near the entrance (Fig. 14).

The influence of the waves on the elevation inside the Bay is shown in Fig. 14 and Fig. 15. The comparison of the model results with and without waves at the NOAA stations shows mostly anemic differences that occur during the main surge (Fig. 14ab); the increase in the main surge due to the wave effects is less than 5 cm. On the other hand, large wave breaking occurs at some steep slopes near the Bay entrance, which in turn results in a larger impact on the surface elevation there up to 30 cm (Fig. 14 and Fig. 15c-e). As shown in Fig. 14a, the wave effects are mostly negligible (<2 cm) in the upper Bay and in the watershed during Hurricane Irene.

6 Discussion on 3D effects and baroclinic adjustment

6.1 Overall comparison for total water level

The effects of including/excluding 3D barotropic processes and 3D baroclinic effects are elucidated through a comparison among "baseline", "3D barotropic" and "2D barotropic" (Table 1). The error statistics from the three configurations are listed in Table 2. The best overall skill is achieved by "baseline"; larger differences are found during the adjustment period from Day 45 to 50.

Table 2: Overall model errors on the simulated water level.

	Total water level MAE (m) **Entire period | adjustment period**	**M_2 amplitude MAE (m)**	**M_2 phase MAE (degree)**
baseline	0.13 | 0.12	0.04	7.5
3D barotropic	0.14 | 0.14	0.06	8.7

2D barotropic	0.15 | 0.15	0.04	10.0

A closer look at the time-series of total elevation also indicates that the largest differences among the three runs occur during the post-surge adjustment period (Fig. 16bc). Besides the main surge that occurred around Aug 28, a second surge mostly attributed to river flooding induced by Tropical Storm Lee occurred around Sept 8. During Irene, while all three configurations have produced the maximum surge well, the baseline reproduces the observed water level fluctuations best on average, especially during the rebounding phase after the peak surge (Fig. 16b), whereas the 2D barotropic run has the largest error for the rebounding phase. Similarly, the baseline is best at sustaining the high water-level during the river flooding phase, especially at the upper Bay stations (e.g., Philadelphia in Fig. 16c), whereas the flood recedes too quickly in the other two configurations. As shown in Table 2, the difference between the MAEs of the two 3D configurations suggests that neglecting baroclinic effects increases the error in elevation by 14% during the adjustment period.

6.2 Role of large-scale processes

To examine the effects of large-scale processes on the water level inside the Bay, we first look at the impact of the hurricane on these processes themselves. The volume transports calculated from the three model runs at a coastal transect along the path of the hurricane are compared in Fig. 17. Overall, the volume transports calculated from the two barotropic runs are close to each other and are significantly lower but with larger fluctuations than that from the baseline baroclinic run. These findings suggest that the baroclinic adjustment is significant in the coastal ocean, which is consistent with the findings of Ezer (2018). Moreover, the 3D barotropic effect is not negligible either, which produces a mean absolute difference of 2.1 Sv between the 3D and 2D barotropic runs, or 19% of the latter's mean. The differences are attributed to 3D barotropic processes (e.g. Ekman transport), which have apparently led to the small differences in the elevations inside the Bay between the two barotropic models as seen in Fig. 16.

Compared to the barotropic runs, the smaller water level fluctuations found in the baseline run are attributed to the stabilizing effects of the large-scale baroclinic processes. The prevailing northward current in the Gulf Stream core has partially buffered the disruption caused by the passage of the hurricane. The northward current is accompanied by a significant surface slope in the shelf seas due to geostrophic balance, which is absent in the barotropic runs (Fig. 18). The existence of the surface slope aids in accelerating the restoration after the passage of the storm, as it works against the prevailing surface slope induced by the storm (with higher elevation nearshore; Fig. 18). As soon as the Gulf Stream restores to its pre-storm condition (Day 37, Fig. 17), the accompanying surface slope works

actively to sustain the elevation in the Delaware Bay. Without the contribution from the Gulf Stream, the elevations from the two barotropic runs show larger swings especially in the river flooding period (Fig. 19), as compared to the baroclinic run in which the Gulf Stream helps sustain the high water-level up-estuary (e.g., Philadelphia in Fig. 16c). As a result, the largest discrepancies in the simulated Bay elevations between baroclinic and barotropic runs are found in the adjustment period.

6.3 Baroclinic vs barotropic pressure gradients

To quantify the importance of baroclinicity during and after the storm, we compare the baroclinic pressure gradient force (PGF) with the barotropic PGF at three representative stations from the Bay mouth to the ocean (#1-#3 in Fig. 3a). The results from the baseline 3D baroclinic run are used. To make these two forces comparable to each other, the baroclinic PGF is depth-averaged as:

$$\overline{F_{BC}} = \frac{1}{H}\int_{-h}^{\eta} F_{BC}(z)\,\mathrm{d}z, \text{ with } F_{BC}(z) = -\frac{g}{\rho_0}\int_{z}^{\eta} \nabla\rho\,\mathrm{d}\zeta,$$

where ∇ is the horizontal gradient operator $(\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$; g is gravity acceleration in [m^2 s^{-1}]; ρ is water density in [kg m^{-3}]; ρ_0 is a reference water density in [kg m^{-3}]; η is surface elevation in [m]; h is the bathymetry in [m]; $H = \eta + h$ is the total water depth in [m]. The barotropic PGF is denoted as $F_{BT} = -g\nabla\eta$.

As shown in Fig. 20, the barotropic PGF is dominant near the Delaware Bay mouth. At the height of Hurricane Irene (Day 32, Station 1 in Fig. 20), barotropic PGF peaks while baroclinic PGF drops to a minimum, confirming the dominance of the barotropic PGF in generating the main surge. But after the passage of the storm (Day 33, Station 1 in Fig. 20), the importance of baroclinicity increases considerably as it works actively to restore (increase) the water level. The magnitude of the baroclinic PGF reaches up to 67% of the barotropic PGF during the subsequent restoration phase. Further offshore, the baroclinic PGF becomes increasingly important as expected (Station 2 and 3 in Fig. 20).

Without the baroclinic gradient in the momentum equation, the depth-averaged velocities calculated from the two barotropic models are different from those from the baseline (Fig. 21). At the two offshore stations, the discrepancies among the three setups are obvious, because the signals from the MAB slope current and the Gulf Stream are absent in the two barotropic runs. But even at the nearshore station, the difference is up to 17% between '3D barotropic' and 'baseline', and 54% between '2D barotropic' and 'baseline'. Therefore, the results here confirm the importance of baroclinicity during the restoration phase.

6.4 Computational performance

We briefly remark on the relative efficiency of the three configurations. The baseline model achieves a performance of ~80 times faster than real time, using 1440 cores of NASA's Pleiades. The 3D barotropic model runs 190 time faster than real time on 960 cores, and the 2D barotropic model runs 230 time faster than real time on 72 cores. Therefore, the 2D and 3D barotropic models are approximately 57 and 3.5 times faster than the baseline respectively.

6.5 Summary and future work

The focus of the discussion is on the interaction between oceanic and estuarine processes. The 3D baroclinic effects are shown to play an important role in the restoration phase. On the other hand, results for smaller-scale hydrological and hydraulic processes in the watershed (including backwater effect, precipitation induced flash flooding, etc.) are not discussed in this paper but will be the focus of a follow-up paper.

7 Conclusion

We have successfully applied a creek-to-ocean 3D baroclinic model to study the response of the Delaware Bay during and after Hurricane Irene (2011). A single unstructured grid was constructed to cover a large domain and provide high resolution in the Bay to accurately simulate the riverine response. The model was forced by the National Water Model at the landward boundary, located at 10 m above sea level. The model was shown to exhibit good skill in predicting the total water levels as well as the 3D density structure. Through comparison among the baseline 3D baroclinic model and sensitivity tests with a 3D barotropic and 2D barotropic model, we examined the importance of baroclinicity during and after the storm including a subsequent river flooding period. The largest differences in elevation were found during the post-surge adjustment period that lasted more than 2 weeks. The baroclinic model better captured the rebounding water level and the sustained high water-level during the ensuing river flooding. The difference was attributed to the stabilizing force provided by the large-scale Gulf Stream. Therefore, our results confirmed that the baroclinicity is a major driving force behind 'fair weather' flooding events as suggested by Ezer (Ezer, T., 2018. On the interaction between a hurricane, the Gulf Stream and coastal sea level. Ocean Dynamics, 68(10), 1259-1272). The coupled modeling system bridges a critical knowledge gap between the hydrological and hydrodynamic regimes, and greatly simplifies the eventual two-way coupling between the two types of models.

Acknowledgement

This study was funded by NOAA under the Water Initiative. The authors thank Dr. L. Magnusson of ECMWF for providing the high-resolution atmospheric model results. Simulations presented in this paper were conducted using the following computational facilities: (1) Sciclone at the College of William and Mary which were provided with assistance from the National Science Foundation, the Virginia Port Authority, Virginia's Commonwealth Technology Research Fund, and the Office of Naval Research; (2) the Extreme Science and Engineering Discovery Environment (XSEDE; Grant TG-OCE130032), which is supported by National Science Foundation grant number OCI-1053575; (3) the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center; and (4) US Department of Energy's Scientific Computing Center.

Supplementary materials

To ensure reproducibility, all input files of the baseline model and more details on input file generation are provided in the supplemental materials (https://doi.org/10.25773/d6x5-wb43).

List of Figures

557

558 Fig. 16: Comparison of the simulated elevations at 3 stations in the lower, mid- and upper Bay between "baseline"
559 and two sensitivity tests. (a) Overview; (b) zoomed-in view during the storm surge of Hurricane Irene; (c)
560 zoomed-in view during the river flooding period. See Fig. 3b for station locations.

561 Fig. 17: Gulf Stream volume transport: (a) locations of the transect near Cape Hatteras; (b) time-series of the
562 volume transport for the "baseline" and two sensitivity tests.

563 Fig. 18: Comparison of SSH at 2011-08-28 00:00:00 (UTC), calculated from: (a) baseline; (b) 3D barotropic; and
564 (c) 2D barotropic models. The dash line in each sub-plot marks a transect used in the subsequent analysis.

565 Fig. 19: Snapshots of sea surface slope along a cross-shore transect (location marked in Fig. 18) during a post-
566 storm period. The MAB slope current (SC) and Gulf Stream (GS) are marked on the surface slope in the baseline.
567 The two barotropic runs show larger temporal swings near the Delaware Bay mouth (the origin of the horizontal
568 axis) than the "baseline".

569 Fig. 20: Relative importance of the barotropic pressure gradient force and the depth-averaged baroclinic pressure
570 gradient force ($\overline{F_{BC}}$): (a) time-series at three representative stations, showing the *magnitudes* of the two terms; (b)
571 station locations. The station locations are shown in Fig. 3a.

572 Fig. 21: Comparison of the magnitude of the depth averaged velocity between the baseline and two sensitivity
573 runs at 3 stations. The station locations are shown in Fig. 3a.

References

Amante, C. and Eakins, B.W., 2009. ETOPO1 arc-minute global relief model: procedures, data sources and analysis.

Ardhuin, F., Rogers, E., Babanin, A.V., Filipot, J.F., Magne, R., Roland, A., Van Der Westhuysen, A., Queffeulou, P., Lefevre, J.M., Aouf, L. and Collard, F., 2010. Semiempirical dissipation source functions for ocean waves. Part I: Definition, calibration, and validation. Journal of Physical Oceanography, 40(9), pp.1917-1941.

Bennis, A.-C., Ardhuin, F., Dumas, F., 2011. On the coupling of wave and three-dimensional circulation models: choice of theoretical framework, practical implementation and adiabatic tests. Ocean Model. 40 (3), 260–272.

Blake, E.S., 2007. The deadliest, costliest and most intense United States tropical cyclones from 1851 to 2006 (and other frequently requested hurricane facts). NOAA Technical Memorandum NWS TPC 5, 43 pp.

Carrere, L., Lyard, F., Cancet, M., Guillot, A. and Picot, N., 2016, May. FES 2014, a new tidal model—Validation results and perspectives for improvements. In Proceedings of the ESA living planet symposium (pp. 9-13).

Chen, C., Beardsley, R.C., Luettich, R.A., Westerink, J.J., Wang, H., Perrie, W., Xu, Q., Donahue, A.S., Qi, J., Lin, H. and Zhao, L., 2013. Extratropical storm inundation testbed: Intermodel comparisons in Scituate, Massachusetts. Journal of Geophysical Research: Oceans, 118(10), pp.5054-5073.

Cho, K.H., Wang, H.V., Shen, J., Valle-Levinson, A. and Teng, Y.C., 2012. A modeling study on the response of Chesapeake Bay to hurricane events of Floyd and Isabel. Ocean Modelling, 49, pp.22-46.

Craig, P. D., and M. L. Banner (1994), Modeling wave-enhanced turbulence in the ocean surface layer, J. Phys. Oceanogr., 24(12), 2546–2559.

Danielson, J.J., Poppenga, S.K., Tyler, D.J., Palaseanu-Lovejoy, M. and Gesch, D.B., 2018. Coastal National Elevation Database (No. 2018-3037). US Geological Survey.

Danilov, S., 2012. Two finite-volume unstructured mesh models for large-scale ocean modeling. Ocean Model. 47, 14–25.

Du, J., Shen, J., Zhang, Y.J., Ye, F., Liu, Z., Wang, Z., Wang, Y.P., Yu, X., Sisson, M. and Wang, H.V., 2018. Tidal Response to Sea - Level Rise in Different Types of Estuaries: The Importance of Length, Bathymetry, and Geometry. Geophysical Research Letters, 45(1), pp.227-235.

Du, J., Park, K, 2019. Estuarine salinity recovery from an extreme precipitation event: Hurricane Harvey in Galveston Bay. Sci. Total Environ. 670, 1049–1059.

Ezer, T., 2013. Sea level rise, spatially uneven and temporally unsteady: Why the US East Coast, the global tide gauge record, and the global altimeter data show different trends. Geophysical Research Letters, 40(20), pp.5439-5444.

Ezer, T., 2018. On the interaction between a hurricane, the Gulf Stream and coastal sea level. Ocean Dynamics, 68(10), pp.1259-1272.

Ezer, T.,2019. Numerical modeling of the impact of hurricanes on ocean dynamics: sensitivity of the Gulf Stream response to storm's track, Ocean Dynamics (in press).

Friedrichs, C.T., 2010. Barotropic tides in channelized estuaries. Contemporary issues in estuarine physics, pp.27-61.

Garvine, R.W., McCarthy, R.K. and Wong, K.C., 1992. The axial salinity distribution in the Delaware estuary and its weak response to river discharge. Estuarine, Coastal and Shelf Science, 35(2), pp.157-165.

Gebert, J.A. and Searfoss, R., 2012. Chapter 4 Sediment, Technical Report for the Delaware Estuary and Basin, Partnership for the Delaware Estuary, Report No. 12-01, pp. 107-118.

Guérin, T., Bertin, X., Coulombier, T. and de Bakker, A., 2018. Impacts of wave-induced circulation in the surf zone on wave setup. Ocean Modelling, 123, pp.86-97.

Harleman, D. R. F., 1966. Real estuaries. Estuary and Coastline Hydrodynamics, A. T. Ippen, Ed., McGraw-Hill, 522–545.

Hasselmann, K., Barnett, T.P., Bouws, E., Carlson, H., Cartwright, D.E., Enke, K., Ewing, J.A., Gienapp, H., Hasselmann, D.E., Kruseman, P. and Meerburg, A., 1973. Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). Ergänzungsheft 8-12.

Hwang, P.A., 2018. High-wind drag coefficient and whitecap coverage derived from microwave radiometer observations in tropical cyclones. Journal of Physical Oceanography, 48(10), pp.2221-2232.

Kerr, P.C., Donahue, A.S., Westerink, J.J., Luettich Jr, R.A., Zheng, L.Y., Weisberg, R.H., Huang, Y., Wang, H.V., Teng, Y., Forrest, D.R. and Roland, A., 2013. US IOOS coastal and ocean modeling testbed: Inter - model evaluation of tides, waves, and hurricane surge in the Gulf of Mexico. Journal of Geophysical Research: Oceans, 118(10), pp.5129-5172.

Kennedy, A.B., Westerink, J.J., Smith, J.M., Hope, M.E., Hartman, M., Taflanidis, A.A., Tanaka, S., Westerink, H., Cheung, K.F., Smith, T. and Hamann, M., 2012. Tropical cyclone inundation potential on the Hawaiian Islands of Oahu and Kauai. Ocean Modelling, 52, pp.54-68.

Knutson, T.R., McBride, J.L., Chan, J., Emanuel, K., Holland, G., Landsea, C., Held, I., Kossin, J.P., Srivastava, A.K. and Sugi, M., 2010. Tropical cyclones and climate change. Nature geoscience, 3(3), p.157.

Le Roux, D.Y., Sene, A., Rostand, V. and Hanert, E., 2005. On some spurious mode issues in shallow-water models using a linear algebra approach. Ocean Modelling, 10(1-2), pp.83-94.

Li, M., Zhong, L., Boicourt, W.C., Zhang, S. and Zhang, D.L., 2006. Hurricane-induced storm surges, currents and destratification in a semi - enclosed bay. Geophysical Research Letters, 33(2).

Magnusson, L., Bidlot, J.R., Lang, S.T., Thorpe, A., Wedi, N. and Yamaguchi, M., 2014. Evaluation of medium-range forecasts for Hurricane Sandy. Monthly Weather Review, 142(5), pp.1962-1981.

Minato, S., 1998. Storm Surge Simulation using POM and a Revisitation of Dynamics of Sea Surface Elevation Short-Term Variation." Papers in Meteorology and Geophysics 48, no. 3 (1998): 79-88.

Münchow, A. and Garvine, R.W., 1993. Dynamical properties of a buoyancy - driven coastal current. Journal of Geophysical Research: Oceans, 98(C11), pp.20063-20077.

Orton, P., Georgas, N., Blumberg, A. and Pullen, J., 2012. Detailed modeling of recent severe storm tides in estuaries of the New York City region. Journal of Geophysical Research: Oceans, 117(C9).

Patchen, R., 2007, November. Establishment of a Delaware Bay Model Evaluation Environment. Estuarine and Coastal Modeling. In Proceedings of the Tenth International Conference, by Malcolm L. Spaulding, PE, Ph. D.,(editor).

Pond, S., Pickard, G.L., 1998. Introductory Dynamical Oceanography. Butterworth-Heinmann, Stoneham, MA.

Rascle, N. and Ardhuin, F., 2013. A global wave parameter database for geophysical applications. Part 2: Model validation with improved source term parameterization. Ocean Modelling, 70, pp.174-188.

Roland, A., Zhang, Y.J., Wang, H.V., Meng, Y., Teng, Y.C., Maderich, V., Brovchenko, I., Dutour - Sikiric, M. and Zanke, U., 2012. A fully coupled 3D wave - current interaction model on unstructured grids. Journal of Geophysical Research: Oceans, 117(C11).

Shapiro, R., 1970. Smoothing, filtering, and boundary effects. Reviews of geophysics, 8(2), pp.359-387.

Sharma, S., Siddique, R., Reed, S., Ahnert, P. and Mejia, A., 2019. Hydrological Model Diversity Enhances Streamflow Forecast Skill at Short - to Medium - Range Timescales. Water Resources Research, 55(2), pp.1510-1530.

Sharp, J.H., 1983. The Delaware estuary: Research as background for estuarine management and development. Univ. Delaware and New Jersey Mar. Sci. Consort.

Soulsby, R. 1997. Dynamics of marine sands, Thomas Telford, London.

Teng, J., Jakeman, A.J., Vaze, J., Croke, B.F., Dutta, D. and Kim, S., 2017. Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. Environmental modelling & software, 90, pp.201-216.

Umlauf, L. and Burchard, H., 2003. A generic length-scale equation for geophysical turbulence models. Journal of Marine Research, 61(2), pp.235-265.

Wahl, T., Jain, S., Bender, J., Meyers, S.D. and Luther, M.E., 2015. Increasing risk of compound flooding from storm surge and rainfall for major US cities. Nature Climate Change, 5(12), p.1093.

Whitney, M.M. and Garvine, R.W., 2006. Simulating the Delaware Bay buoyant outflow: Comparison with observations. Journal of Physical Oceanography, 36(1), pp.3-21.

Wolf, J., 2009. Coastal flooding: impacts of coupled wave–surge–tide models. Natural Hazards, 49(2), pp.241-260.

Wong, K.C., 1995. The hydrography at the mouth of Delaware Bay: Tidally averaged distribution and intratidal variability. Estuarine, Coastal and Shelf Science, 41(6), pp.719-736.

Wong, K.C. and Münchow, A., 1995. Buoyancy forced interaction between estuary and inner shelf: observation. Continental Shelf Research, 15(1), pp.59-88.

Yankovsky, A.E. and Chapman, D.C., 1997. A simple theory for the fate of buoyant coastal discharges. Journal of Physical oceanography, 27(7), pp.1386-1401.

Ye, F., Zhang, Y.J., Wang, H.V., Friedrichs, M.A., Irby, I.D., Alteljevich, E., Valle-Levinson, A., Wang, Z., Huang, H., Shen, J. and Du, J., 2018. A 3D unstructured-grid model for Chesapeake Bay: Importance of bathymetry. Ocean Modelling, 127, pp.16-39.

693 Ye, F., Zhang, Y., He, R., Wang, Z.G., Wang, H.V., and Du, J., 2019. Third-order WENO transport
694 scheme for simulating the baroclinic eddying ocean on an unstructured grid, Ocean Modelling
695 (2019)101466, https://doi.org/10.1016/j.ocemod.2019.101466.

696 Zeng, X., Zhao, M. and Dickinson, R.E., 1998. Intercomparison of bulk aerodynamic algorithms for the
697 computation of sea surface fluxes using TOGA COARE and TAO data. Journal of Climate, 11(10),
698 pp.2628-2644.

699 Zeng, X. and He, R., 2016. Gulf Stream variability and a triggering mechanism of its large meander in
700 the South Atlantic Bight. Journal of Geophysical Research: Oceans, 121(11), pp.8021-8038.

701 Zhang, Y.J., Ateljevich, E., Yu, H.C., Wu, C.H. and Jason, C.S., 2015. A new vertical coordinate
702 system for a 3D unstructured-grid model. Ocean Modelling, 85, pp.16-31.

703 Zhang, Y.J., Ye, F., Stanev, E.V. and Grashorn, S., 2016. Seamless cross-scale modeling with
704 SCHISM. Ocean Modelling, 102, pp.64-81.

705 Zheng, L., Weisberg, R.H., Huang, Y., Luettich, R.A., Westerink, J.J., Kerr, P.C., Donahue, A.S.,
706 Crane, G. and Akli, L., 2013. Implications from the comparisons between two - and three -
707 dimensional model simulations of the Hurricane Ike storm surge. Journal of Geophysical Research:
708 Oceans, 118(7), pp.3350-3369.

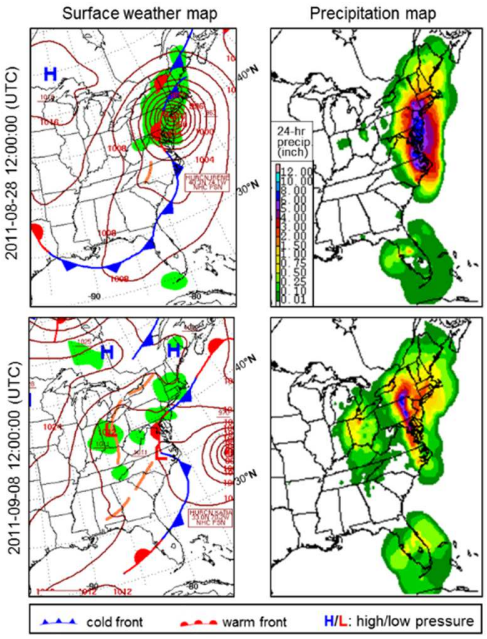

Fig. 1: Weather along the US east coast on Aug 28, 2011 (around Hurricane Irene) and Sept 8, 2011 (around Tropical Storm Lee), showing large precipitations on both dates. The contours in the surface weather maps show air pressure in millibar. The 24-hour precipitation is a record of the past 24 hours until the shown time. Hurricane Katia (2011; bottom-left panel) did not land on the US east coast. Credit: NOAA Central Library U.S. Daily Weather Maps Project (https://www.wpc.ncep.noaa.gov/dailywxmap/explaination.html); partial views of the original online maps.

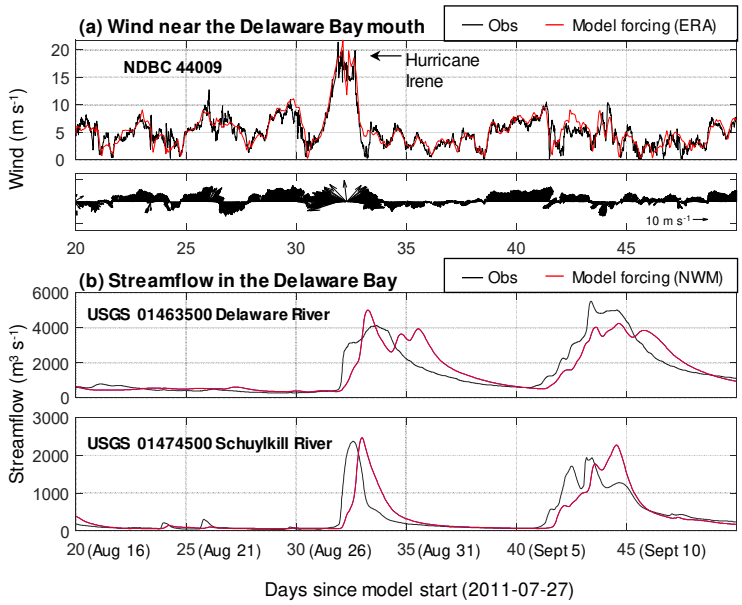

Fig. 2: The influence of Hurricane Irene on the Delaware Bay: (a) wind patterns near the Bay mouth (see Fig. 3 for the location of the station NDBC 44009); (b) streamflows at two USGS gauges (locations shown in Fig. 3; the Delaware River station is at Trenton NJ), with two peaks corresponding to the landfall of Irene (2011) and the subsequent river flooding under Tropical Storm Lee (2011).

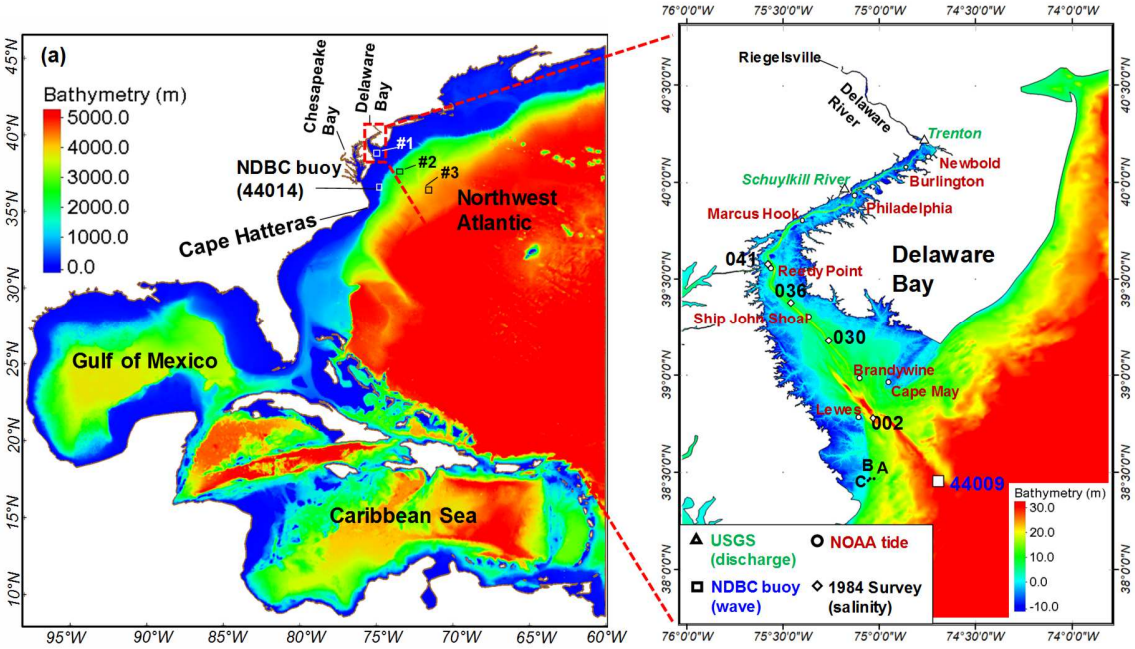

Fig. 3: Model domain, bathymetry and locations of observation stations. Additionally, sample points ("#1", "#2", "#3" and "A", "B", "C") used in subsequent analysis are marked.

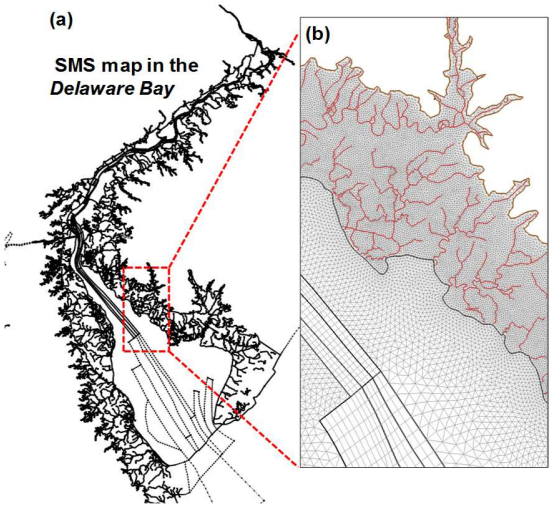

Fig. 4: Illustration of horizontal grid generation: (a) “feature arcs” in SMS, used to align the grid elements with channels and follow the NWM segments; (b) zoomed-in view on the arcs and the grid, with the arcs corresponding to NWM segments highlighted in red.

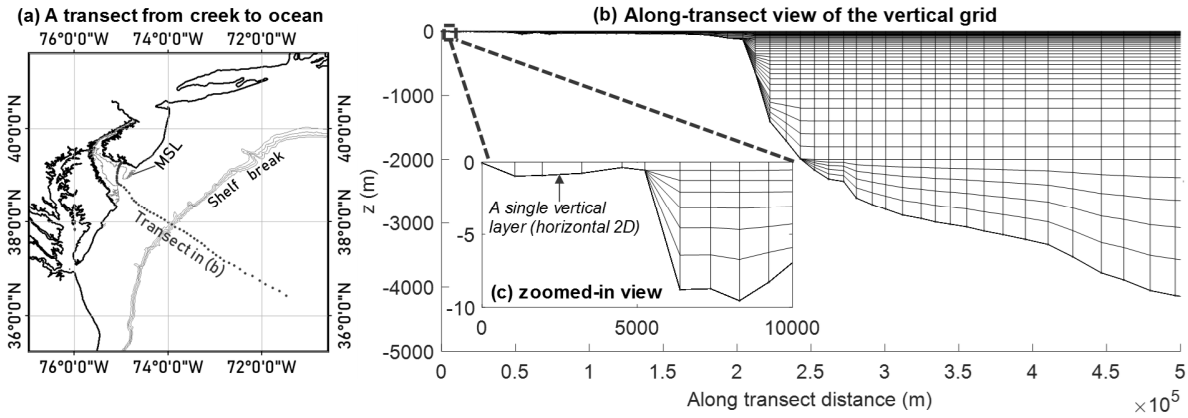

Fig. 5: Illustration of the vertical grid along a creek-bay-ocean transect: (a) transect location; (b) along-transect view of the vertical grid; (c) zoomed-in view on the shallow portion of (b).

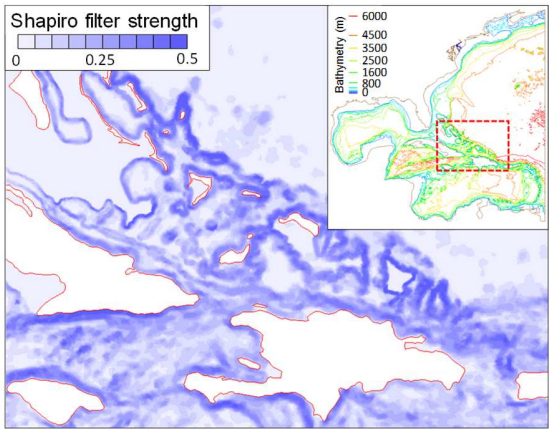

Fig. 6: Shapiro filter strength in a region (box in the inset) with steep bathymetry. The maximum strength is set as 0.5 (Zhang et al. 2016).

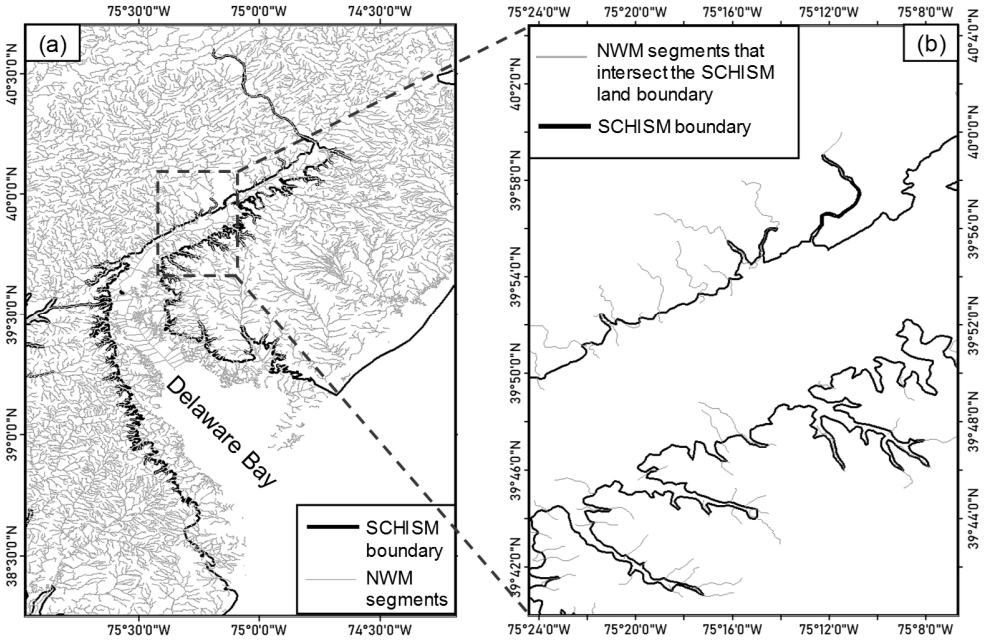

Fig. 7: The coupling between the NWM and SCHISM: (a) NWM segments and SCHISM land boundaries for the Delaware Bay; (b) zoomed-in view on the NWM segments that intersect the SCHISM land boundary in the upper Delaware Bay.

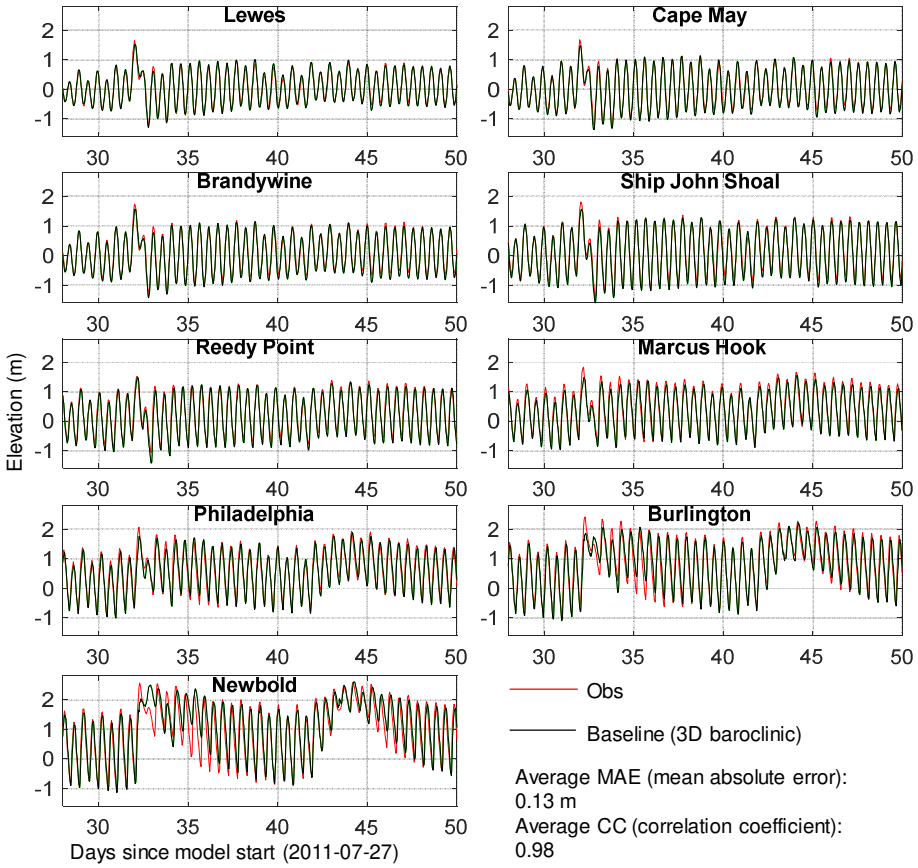

Fig. 8: Model-data comparison of surface elevations.

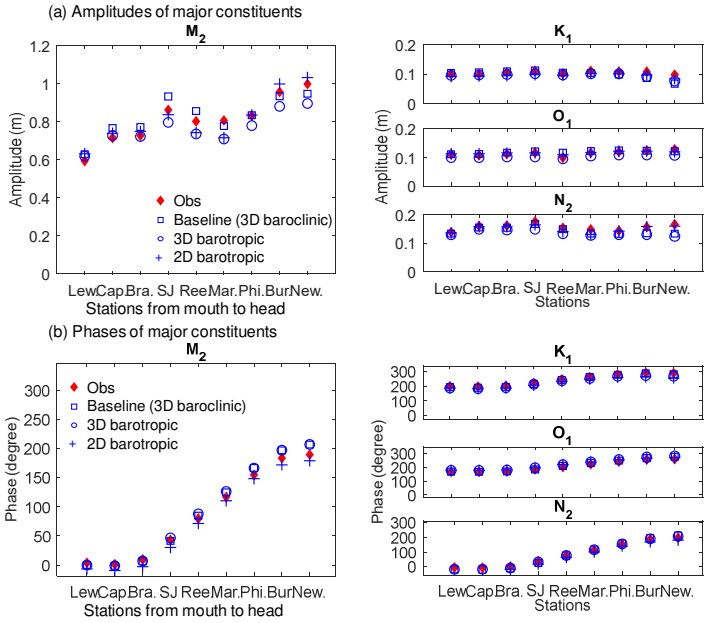

Fig. 9: Tidal harmonics of 4 major constituents calculated for Days 10-50 of the simulation period: (a) amplitude; (b) phase. See Fig. 3 for station locations. The results from two sensitivity runs (3D barotropic and 2D barotropic) are also included here for future reference.

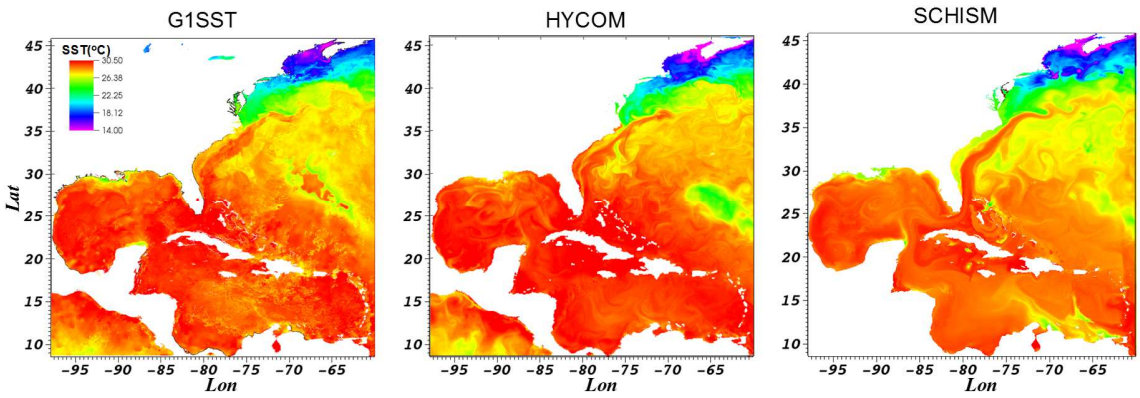

Fig. 10: Comparison of SST on 2011-09-07. The MAEs throughout the domain are 0.61°C for HYCOM and 0.70°C for SCHISM.

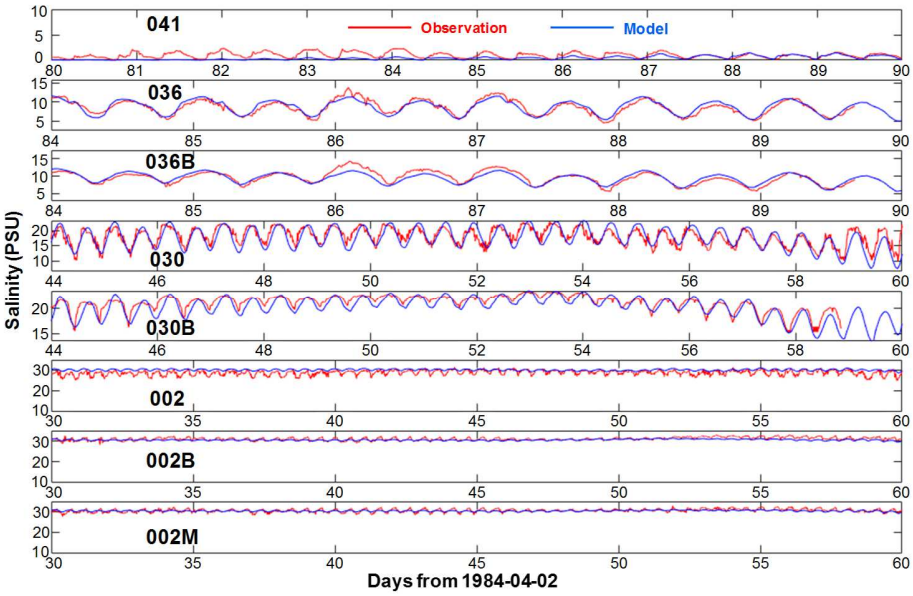

Fig. 11: Comparison of salinity for 1984 survey, at multiple stations and multiple vertical positions ('B' denotes bottom, 'M' denotes mid-depth; otherwise surface). Note that the time axes are different for different stations. The overall MAE is 0.86 PSU. See Fig. 3 for station locations.

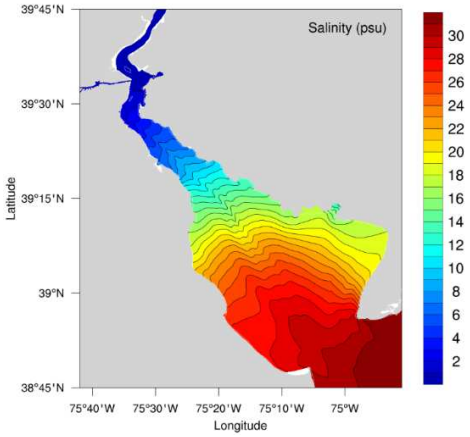

Fig. 12: Depth averaged salinity for the last 70 days of the 1984 simulation.

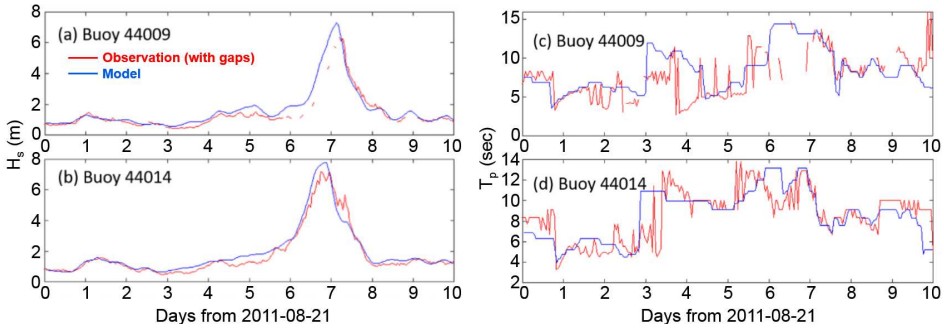

Fig. 13: Comparison of (a,b) significant wave height, and (c,d) peak period at two NDBC buoys. See Fig. 3 for station locations.

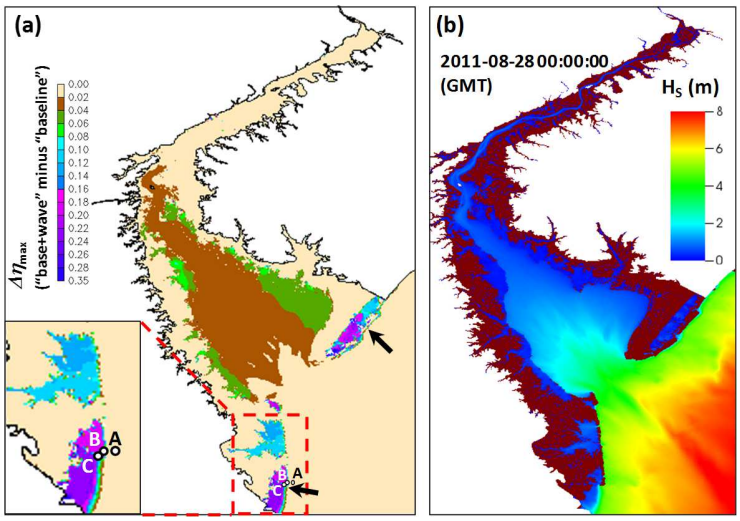

Fig. 14: The wave effect: (a) differences in maximum elevation between 'base+wave' and baseline, with the largest differences found in two regions and pointed out by black arrows; (b) strong wave breaking near the steep slopes (cf. the bathymetry in Fig. 3). The locations of stations A-C in Fig. 3 are repeated here to correlate them with the wave influence.

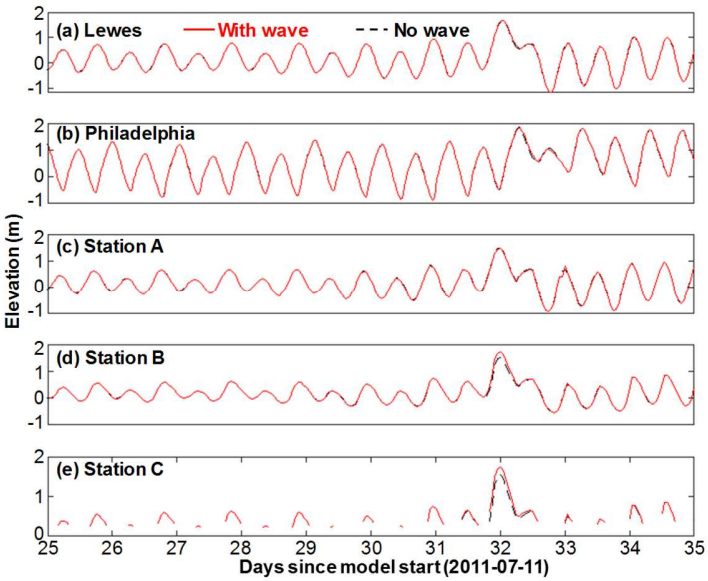

Fig. 15: Comparison of total water elevation at five stations between the "baseline" (no wave) and the "base+wave" (with wave) results. See Fig. 3b and Fig. 14 for station locations.

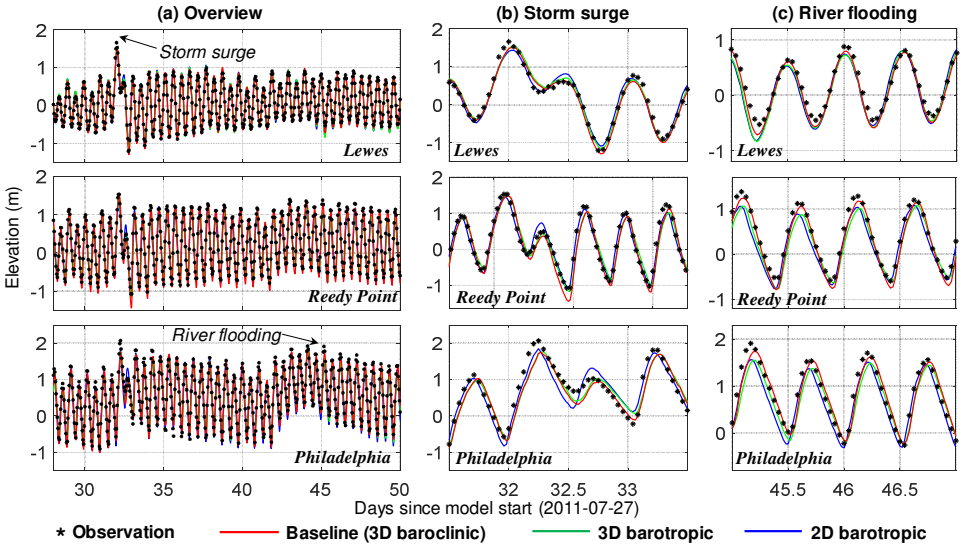

Fig. 16: Comparison of the simulated elevations at 3 stations in the lower, mid- and upper Bay between "baseline" and two sensitivity tests. (a) Overview; (b) zoomed-in view during the storm surge of Hurricane Irene; (c) zoomed-in view during the river flooding period. See Fig. 3b for station locations.

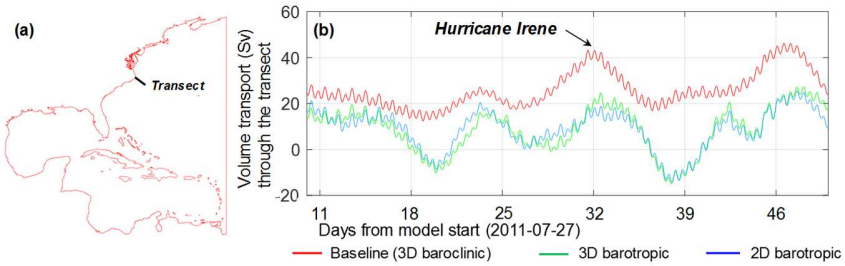

Fig. 17: Gulf Stream volume transport: (a) locations of the transect near Cape Hatteras; (b) time-series of the volume transport for the "baseline" and two sensitivity tests.

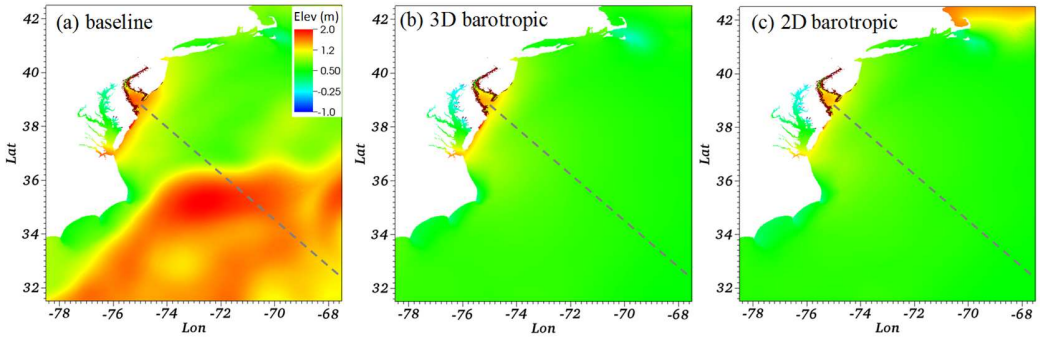

Fig. 18: Comparison of SSH at 2011-08-28 00:00:00 (UTC), calculated from: (a) baseline; (b) 3D barotropic; and (c) 2D barotropic models. The dash line in each sub-plot marks a transect used in the subsequent analysis.

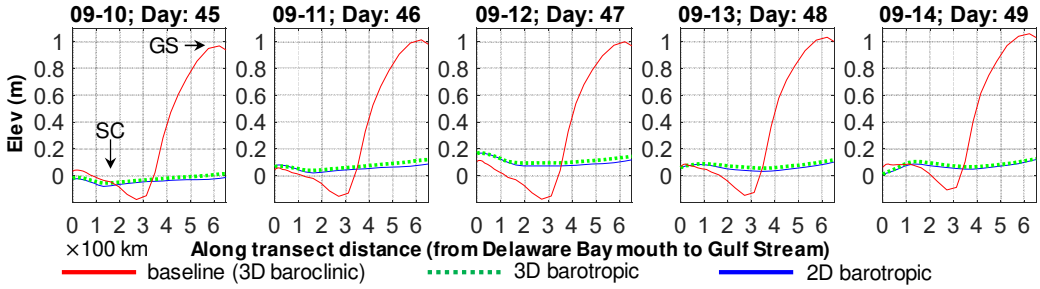

Fig. 19: Snapshots of sea surface slope along a cross-shore transect (location marked in Fig. 18) during a post-storm period. The MAB slope current (SC) and Gulf Stream (GS) are marked on the surface slope in the baseline. The two barotropic runs show larger temporal swings near the Delaware Bay mouth (the origin of the horizontal axis) than the "baseline".

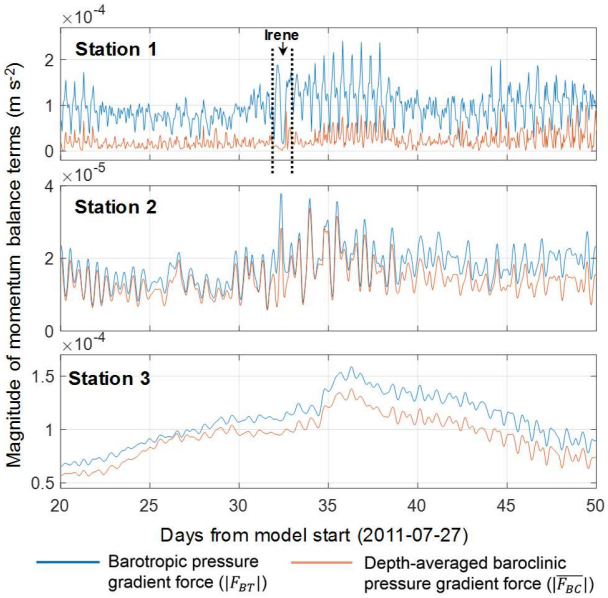

Fig. 20: Relative importance of the barotropic pressure gradient force and the depth-averaged baroclinic pressure gradient force ($\overline{F_{BC}}$): time-series at three representative stations, showing the *magnitudes* of the two terms. The station locations are shown in Fig. 3.

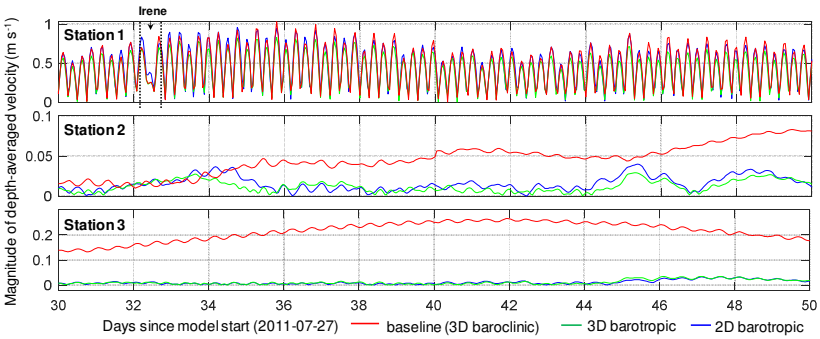

Fig. 21: Comparison of the magnitude of the depth averaged velocity between the baseline and two sensitivity runs at 3 stations. The station locations are shown in Fig. 3a.