

1 Simulating storm surge and compound flooding events with 2 a creek-to-ocean model: importance of baroclinic effects 3

4
5 Fei Ye^{a1}, Yinglong J. Zhang^a, Haocheng Yu^a, Weiling Sun^a, Saeed Moghimi^b, Edward Myers^b, Karinna
6 Nunez^a, Ruoyin Zhang^c, Harry V. Wang^a, Aron Roland^d, Kevin Martins^{e, f}, Xavier Bertin^f, Jiabi Du^g,
7 Zhuo Liu^a
8

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

10 b. Coast Survey Development Laboratory, National Oceanic and Atmospheric Administration, Silver Spring, MD
11 20910, USA

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

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

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

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

17 g. Department of Marine Sciences, Texas A&M University at Galveston, Galveston, TX 7754, USA
18
19

20 Abstract

21 We present a creek-to-ocean 3D baroclinic model based on unstructured grids that aims to unite
22 traditional hydrologic and ocean models in a single modeling platform, by taking full advantage of the
23 polymorphism (i.e. a single model grid can seamlessly morph between full 3D, 2DV, 2DH and quasi-
24 1D configurations). Using Hurricane Irene (2011)'s impact on the Delaware Bay as an example, a
25 seamless 2D-3D model grid is implemented to include the entire US East Coast and Gulf of Mexico
26 with a highly resolved Delaware Bay (down to 20-m resolution). The model is forced by flows from a
27 hydrological model (National Water Model) at the landward boundary. We demonstrate the model's
28 accuracy, stability and robustness with the simulation of the storm surge and subsequent river flooding
29 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.

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

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

36

37 **1 Introduction**

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

43 In general, three types of inundation processes are of concern to coastal communities: coastal storm
44 surge and inundation, pluvial inundation (precipitation driven flooding), and fluvial inundation (river
45 flooding). Traditionally, these processes have been studied separately using different types of models:
46 hydrodynamic models for storm surge and hydrological/hydraulic models for pluvial/fluvial flooding. A
47 hydrodynamic model generally excludes the watershed mostly because of the constraints induced by
48 numerical stability and/or computational cost, whereas a hydrologic model's capability roughly stops at
49 mean sea level (MSL) because it is not designed for estuarine/oceanic processes. A plethora of models
50 are available for these two types of simulations (e.g., Kerr et al. 2013; Chen et al. 2013; Sharma et al.
51 2019). These models, however, often neglect the important interaction between coastal and inland
52 processes such as the compound flooding and backflow effects. The compound effects from the
53 combination of all these processes, including the 3D baroclinic effects, have not been studied in detail
54 before and are the subject of this study.

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

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

63 To this end, National Oceanic and Atmospheric Administration (NOAA) has initiated the NOAA
64 Water Initiative² to understand, develop, demonstrate and implement an improved total water level
65 prediction that includes signals propagating up and downstream in coastal, estuarine and riverine
66 environments, particularly during storm events. The goal is to establish an integrated water forecasting
67 system that covers inland and coastal waters to provide vital information to end-users and stakeholders.
68 As our contributions to the Water Initiative, this study uses a seamless creek-to-ocean 3D baroclinic
69 unstructured-grid (UG) model (SCHISM, schism.wiki, last accessed in July 2019; Zhang et al. 2016),
70 driven by the predicted flows from a continental-scale hydrological model (National Water Model³, or
71 “NWM”), to holistically simulate the total water level and its individual components. The flexibility
72 and robustness of SCHISM are indispensable for traversing large contrasts of temporal and spatial
73 scales from oceanic processes (such as the Gulf Stream) to inland flooding in small creeks. In doing so,
74 we have built a very reconfigurable and flexible modeling platform that can be extended to cover a
75 larger domain with sufficient resolution in the areas of interest. Effort is on-going to extend the current
76 model to cover all major estuaries and bays in the US East Coast and Gulf of Mexico.

77 In this study, we focus on the importance of including the 3D effects in the storm surge simulations.
78 The model is first validated using observational data derived from NOAA, USGS and satellite products.
79 Sensitivity tests are then conducted to examine the importance of 3D effects, in particular, the
80 baroclinicity. Minato (1998) was among the first to elucidate the implication of incorporating 3D
81 effects in simulating storm surges, and he gave a simple explanation on why finer vertical resolution led
82 to higher surges in Tosa Bay (Japan). Zheng et al. (2013) compared results from 2D and 3D barotropic
83 models for the storm surges in the Gulf of Mexico and showed that both models can adequately
84 simulate the surge provided that the bottom frictions were properly calibrated. Li et al. (2006) and Cho
85 et al. (2012) studied the impact of Hurricane Floyd (1999) and Isabel (2003) in the Chesapeake Bay
86 using 3D baroclinic models but did not explicitly expound the influence of baroclinicity on surface
87 elevation, partly because of the small model domain used. Orton et al. (2012) used a 3D baroclinic
88 model on a small domain to simulate storm surges near New York City and estimated that neglecting
89 water density variations led to typical reductions of 1–13% in the peak surge. But as suggested by
90 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

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

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

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

114 **2 Study case**

115 **2.1 Hurricane Irene (2011)**

116 Hurricane Irene, the first major Atlantic hurricane in 2011, was selected here as a case study. The
117 hurricane made its landfall along the US East Coast at Outer Bank, NC on Aug 27, 2011 as a Category
118 1 hurricane. The hurricane re-entered Atlantic coastal waters in Virginia, Delaware and New Jersey,
119 and weakened to a tropical storm before making the second landfall in New Jersey and the third landfall
120 in New York City. As seen from the weather maps (Fig. 1), while there was only one major wind event
121 related to Irene, there were two large precipitation events associated with Irene and the subsequent

122 Tropical Storm Lee; the latter originated from Gulf of Mexico and swept over land over eastern states.
123 These events led to two streamflow peaks in Delaware Bay’s largest tributary, the Delaware River (Fig.
124 2b). The first peak flow of about $4000 \text{ m}^3 \text{ s}^{-1}$ occurred on Aug 28, 2011, around the landfall of Irene.
125 The second peak, about $1500 \text{ m}^3 \text{ s}^{-1}$ higher than the first, occurred 10 days later under Tropical Storm
126 Lee (2011), mainly resulting from the precipitation and the subsequent pluvial/fluviial processes rather
127 than a storm surge. The streamflow of the second largest tributary, the Schuylkill River, exhibits similar
128 patterns (Fig. 2b). The different flood origins (ocean and inland) made this event an ideal test for
129 compound-flood modeling.

130 **2.2 Observation**

131 The observational datasets used in this study included 9 tide gauges maintained by NOAA⁴, 2
132 streamflow gauges from USGS⁵, and 2 wave buoys from NOAA’s National Data Buoy Center
133 (NDBC)⁶. The locations of these stations are shown in Fig. 3. In addition, salinity survey data inside the
134 Delaware Bay collected during a 1984 intensive survey was used to assess the model’s performance in
135 simulating 3D baroclinic processes, because of the lack of salinity observation during Irene. Also,
136 NASA’s GHRSSST Level 4 G1SST Global Foundation Sea Surface Temperature Analysis⁷ was used to
137 assess the model skills for large-scale processes in the open ocean, including the Gulf Stream. The use
138 of all these observation datasets ensured a thorough assessment of the model skill.

139 **3 Numerical model**

140 **3.1 Model and domain**

141 SCHISM is a flexible, primitive equation, hydrostatic model grounded on hybrid Finite-
142 element/Finite-volume method and hybrid triangular-quadrangular UGs in the horizontal and hybrid
143 Localized Sigma Coordinates with Shaved Cells (LSC²) grid in the vertical (Zhang et al. 2015, 2016).
144 The model uses a semi-implicit time stepping scheme to enhance robustness and efficiency, and the
145 numerical dissipation is kept low with a judicious combination of higher-order, monotone schemes (Ye
146 et al. 2018; Ye et al. 2019) and the semi-implicit finite-element formulation.

147 A SCHISM-based regional model has been developed, which covers the Northwest Atlantic Ocean,
148 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.

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

152 The focus site in this pilot study is the Delaware Bay, which is a major estuary on the U.S. East
153 Coast. Tidal portions of the Delaware River start at Trenton, New Jersey, and the total length of the
154 estuary from Trenton to the mouth is ~210 km, draining a watershed area of $4.2 \times 10^5 \text{ km}^2$ (Sharp 1984).
155 A key characteristic of the estuary, as far as the tides are concerned, is that it constricts rapidly from the
156 widest point in the lower Bay (near the NOAA station Brandywine; Fig. 3b) to the mid-Bay. The
157 funneling effect from width variation leads to a ‘hypersynchronous’ system with tides generally
158 increasing landward (Friedrichs 2010), but the channel meandering in the mid-Bay (near the NOAA
159 station Reedy Point; Fig. 3b) considerably complicates the dynamics there (Section 4.1). The Bay is
160 mostly shallow with a mean depth of 7 m (Harleman 1966). The Delaware River provides 58% of the
161 freshwater inflow; the confluence of the Schuylkill River below Philadelphia adds another 14%; other
162 tributaries collectively account for the remaining 28% (Sharp 1983). The average freshwater inflow is
163 $570 \text{ m}^3 \text{ s}^{-1}$, with high flow conditions occurring during the spring freshets (Whitney and Garvine 2006).
164 Even under peak freshwater inflow, the estuary is vertically mixed by the tides most of time (Wong
165 1995). The freshwater outflow from the Bay is generally weak; the plume is typically in contact with
166 the frictional bottom boundary layer and confined within 20 kilometers offshore under normal
167 conditions (Münchow and Garvine 1993; Wong and Münchow 1995; Yankovsky and Chapman 1997).

168 The model domain includes a part of the Delaware Bay watershed, up to the 10-m isobath above
169 mean sea level (MSL). In the Delaware River (the largest tributary of the Delaware Bay), the domain
170 extends to the USGS gauge of Riegelsville (Fig. 3b) at 40 m above MSL. Such a domain choice
171 requires that the model directly simulates some pluvial and fluvial processes normally handled by
172 hydrological models. The bathymetry information (Fig. 3) is derived from two DEM (Digital Elevation
173 Model) sources: the global relief model ETOPO1 (Amante and Eakins, 2009) for the ocean, and the 1-
174 m USGS Coastal National Elevation Database (Danielson, 2018) for the Delaware Bay. The inclusion
175 of a large portion of Delaware Bay watershed complicates the use of a curved vertical datum such as
176 NGVD29 and therefore, the model is based on a flat datum of NAVD88. The latter is convenient
177 because (1) the USGS DEM is given in NAVD88; (2) most new instruments use this datum.
178 Accordingly, all model-data comparisons on elevation are also based on NAVD88. A datum conversion
179 from NAVD88 to NGVD29 (e.g. using the VDatum⁸ tool, which itself relies on model simulation)

⁸ url: <https://vdatum.noaa.gov/>, last accessed in June 2019.

180 would introduce uncertainties especially in the upper Bay near the fall line, so we choose to rely on
181 observation instead. Among the nine NOAA tide stations used (Fig. 3b), the differences between the
182 local MSLs and the NAVD88 datum are available at three stations: the lower Bay station “Lewes” (-
183 0.121 m; i.e., the local MSL is below the NAVD88 datum), the mid-Bay station Reedy Point (-0.015
184 m), and the upper Bay station Philadelphia (0.118 m). As shown later (Section 4.1), the model is able to
185 correctly set up the surface slope, with the local MSL asymptotically increasing toward upper Bay as
186 observed. The datum differences between MSL and NAVD88 at the three stations are linearly
187 interpolated onto other stations based on the along-thalweg distance (with linear extrapolation for
188 Berlington and Newbold, which are up-estuary from Philadelphia). This procedure is expected to
189 introduce uncertainties on the order of a few centimeters in the model-data comparison of elevation.

190 **3.2 Grid generation**

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

200 The grid generation software SMS (Surface-water Modeling System)⁹ was used to generate the
201 horizontal grid. The horizontal spatial domain was discretized by an UG with 667 K nodes and 1,273 K
202 elements, including 39 K quadrangular elements used to represent the shipping channel (Fig. 4). A
203 quasi-uniform triangular grid with a resolution of 6-7 km was applied in the open ocean, which was
204 smoothly transitioned to about 2-km resolution near the coastline. Locally high resolution was applied
205 in the Delaware Bay, with a typical resolution of 600 m in the lower Bay channel, 50 m in the upper
206 Bay channel, 150 m in the watershed areas above MSL, and down to 20 m in some small creeks.
207 During the grid generation process, “feature arcs” (Fig. 4a) were used mainly to (1) explicitly
208 incorporate ‘features’ such as the NWM segments into SCHISM’s horizontal grid (red arcs in Fig. 4b);
209 (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.

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

212 The vertical discretization used in the model took full advantage of the hybrid terrain-following-like
213 LSC² coordinate (Zhang et al., 2015), with variable number of layers at different horizontal locations.
214 The average number of layers was 18.3, with a maximum of 44 layers in the deepest ocean and only 1
215 layer in shallow areas with depths less than 0.5 m (Fig. 5). As a result, 2D representation was applied
216 for about 57% of the Delaware Bay watershed or 30% of the total grid elements. An element was
217 deemed wet when the local water depth, calculated from the implicit finite-element solver, exceeded 10⁻⁶
218 m; such a small threshold was needed to accurately capture the very thin layer of fluid initially formed
219 on dry land during precipitation events. A smaller value (10⁻⁸) does not significantly change the results.
220 The model was stable even with this choice of small threshold for wetting and drying, courtesy of the
221 implicit scheme. Because the vertical flow structure can be safely ignored for the pluvial and inundation
222 processes, the single-layer configuration greatly reduced the computational cost and meanwhile
223 enhanced the robustness of the model (Zhang et al. 2016).

224 **3.3 Baseline model setup**

225 **3.3.1 Forcing and parameters**

226 The “baseline” setup used a 3D baroclinic model. Wave effects were excluded in the baseline model
227 but examined in subsequent sensitivity analysis (Section 5). Atmospheric forcing applied at the air-sea
228 interface consisted of two sources. The first source was derived from ECWMF’s ERA5 reanalysis
229 dataset¹⁰, and the variables included air temperature, air pressure (reduced to MSL), humidity, wind
230 speed and direction at 10 m above MSL, downward short-wave and long-wave radiations, and
231 precipitation rate. This product has a spatial resolution of 30 km and temporal resolution of 1 hour. The
232 second source was a high-resolution product from ECMWF, with a spatial resolution of 5 km
233 (Magnusson et al. 2014). The comparison shown in Fig. 2 suggests that the atmospheric forcing used in
234 the model is sufficiently accurate for simulating the storm surge. Occasionally, there are noticeable
235 mismatches (about 4 m s⁻¹) in wind speed between the forcing and the observation, e.g., at the peak of
236 Irene and one day afterwards (Day 32-33 in Fig. 2a) and during the river flooding period (Day 42.5 in
237 Fig. 2a). These mismatches contribute to the uncertainties in the simulated water level. The surface
238 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.

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

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

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

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

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

278 **3.3.2 Coupling to NWM**

279 The freshwater delivery into the Delaware Bay was derived from the NWM, i.e. a reanalysis
280 product¹¹ from NWM v1.2, provided by the NOAA team. We first determined all NWM segments that
281 intersected the SCHISM land boundary, and then the streamflow at each segment was then imposed in
282 the adjacent SCHISM elements as a point source (for inflow segments) or sink (for outflow segments)
283 (Fig. 7). Implementation of volume and mass sources and sinks inside SCHISM was rather
284 straightforward via simple volume integrations in the finite-element equations. Although we have only
285 considered the one-way coupling between NWM and SCHISM so far, the simplicity of this coupling
286 strategy bodes well for the eventual two-way coupling between the two models. The precipitation
287 effects are not discussed in this paper and are left for a future study.

288 **3.4 Sensitivity runs**

289 Sensitivity runs were conducted to examine the importance of different processes, including
290 baroclinic response, 3D barotropic processes, wind wave effects, etc. Table 1 shows the setups of
291 important sensitivity runs used in this paper. In the “3D Barotropic” run, tracer transport was turned off
292 and the baroclinic force from temperature and salinity gradient was excluded. To assess the importance
293 of 3D barotropic processes, a “2D Barotropic” run was also included. As explained in Zheng et al.
294 (2013), both 2D and 3D models can simulate a storm surge well, if bottom friction coefficients are
295 properly adjusted in each model. The near-bottom velocities in 2D and 3D models differ by orders of
296 magnitude, so drag coefficients need to be adjusted accordingly to achieve comparable bottom stress,
297 which is one of the key controls for surges. In general, 3D baroclinic models should use a much smaller
298 coefficient than 2D models, because the former produce larger near-bottom velocity inside stratified
299 regions due to the two-layer exchange flow. This considerably complicates the inter-comparison of 2D
300 and 3D models (Zheng et al. 2013).

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

301 A key difference between the current and previous studies is that we calibrated the three
 302 configurations (Table 1) of the model separately to achieve best possible results (in terms of overall
 303 Mean Absolute Error, or ‘MAE’) before the results were inter-compared. Different choices of surface
 304 stress formulations, time steps, and other parameters were tested, but the dominant control was found to
 305 be the bottom friction. This approach largely removes the ambiguity of the bottom friction
 306 parameterizations used in each configuration and ensures that the findings in subsequent sections are
 307 not unduly influenced by parameter choices. For the 2D model, we used a uniform Manning’s n of
 308 $0.019 \text{ s m}^{-1/3}$; for the 3D barotropic model, the same roughness height as in the “baseline” was found to
 309 give best results.

310 Wave effects are also important in nearshore regions (Kennedy et al. 2012; Guerin et al. 2018). In
 311 particular, wave breaking induces a setup near the shoreline and embayment and alters the mean
 312 circulation nearshore. To examine this, we included a run called “base+wave”, where the base model
 313 was fully coupled with the Wind Wave Model (WWM; Roland et al., 2012) on the same horizontal
 314 grid. The wave effect was incorporated into SCHISM via the 3D vortex formalism of Bennis et al.
 315 (2011) as implemented and validated in Guerin et al. (2018). In addition, the formulation of wave-
 316 enhanced bottom boundary layer was from Soulsby (1997) and the formulation of wave breaking
 317 induced turbulence followed that of Craig and Banner (1994). The wave model was initialized using a
 318 global hindcast product based on WWIII¹² (Rascle and Ardhuin, 2013), and was also forced at the
 319 ocean boundary by the same product; five variables (significant wave height, peak frequency, mean
 320 wave period (TM02) and direction, and directional spread) were extracted to construct wave spectra at
 321 the ocean boundary using the JONSWAP formula (Hasselmann et al. 1973). WWM was run with the
 322 implicit mode, and the coupling time step was set to be 600 s. Thirty-six bins were used to resolve the
 323 directional and frequency domain.

324 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.

325 **4 Model validation**

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

330 **4.1 Surface elevation**

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

343 A tidal harmonic analysis is also conducted on major tidal constituents, including M_2 , K_1 , O_1 , and N_2
344 (Fig. 9), from Day 10 to Day 50 of the simulation period. Although this period includes flooding/surge,
345 the model data comparison is still valid because the same period is applied to both model and data. The
346 “baseline” again shows good agreement with observation. The average MAEs for the M_2 constituent
347 (which accounts for 93% of the total tidal energy) of all stations are: 4.0 cm in amplitude and 7.5 degree
348 (15.5 minutes) in phase. The observed M_2 amplitude increases from the mouth to the mid-Bay, and then
349 slightly decreases from the mid-Bay from Ship John Shoal to Reedy Point (see station locations in Fig.
350 3b), and increases again into the upper Bay. This longitudinal variability is attributed to the balance
351 between the funneling effects due to the trumpet shape of the channel and the bottom friction (Du et al.
352 2018), but the meandering near Reedy Point further complicates the dynamics and leads to the local
353 extrema. In general, the model captures this trend, but has slightly larger errors near the mid-Bay
354 extrema than elsewhere. The model also tends to lag the observation at upper Bay stations.

355 **4.2 Sea-surface temperature (SST)**

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

360 To ensure that the baroclinic processes in the open ocean (mainly controlled by temperature
361 gradient) are well captured by the baseline model, the simulated SST is compared with a satellite
362 derived product (NASA’s GHRSSST Level 4 G1SST). During the passage of Hurricane Irene, the Gulf
363 Stream is greatly disturbed, as evidenced by the fluctuations in its volume transport (cf. Fig. 17b). The
364 restoration of the coastal ocean takes several days (Ezer 2018). By Sept 7, 2011 the Gulf Stream is
365 largely restored to its pre-storm condition (Fig. 10). SCHISM qualitatively captures the restored SST
366 condition, with the free meandering north of Cape Hatteras slightly improved from that in HYCOM.
367 Also apparent in both observation and model results is a cold wake in the middle of Atlantic left by
368 another storm (Hurricane Katia; also see the bottom-left panel of Fig. 1), which did not land on the US
369 east coast. The complex eddies and meanders as shown in Fig. 10 cannot be reproduced by barotropic
370 models and are responsible for setting up large surface slopes along the path of the Gulf Stream (cf. Fig.
371 18 and Fig. 19), which has implications for the rebounding water level inside the Bay (Section 6). More
372 detailed assessment of the model skill including the vertical structure of the Gulf Stream has been
373 reported in Ye et al. (2019).

374 **4.3 Salinity**

375 Delaware Bay is a weakly stratified estuary with a nearly linear axial salinity distribution (Garvine et
376 al. 1992). Previously, a multi-model comparison for this system using the observational data collected
377 during the 1984 hydrographic surveys has been shown in Patchen (2007). During the surveys, several
378 stations reported salinity measurements, three of which had data at more than one depth. Note that the
379 measurements at different stations were collected at different time periods.

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

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

391 **5 Wind wave effects**

392 To assess the wave effects on the water surface elevation, we restarted the baseline simulation on
 393 August 21, 2017 and ran the “base+wave” (Table 1) model for 10 days. We first validate the
 394 “base+wave” model using the observation at two nearby NDBC buoys (see locations in Fig. 3). The
 395 modeled significant wave heights and peak periods matches the observation well (Fig. 13), with the
 396 average MAEs being 24 cm for the significant wave height and 2.1 seconds for the peak period
 397 respectively. During the storm, large waves of relatively longer periods (Fig. 13) entered the Bay from
 398 the south and large wave breaking occurred near the steep bathymetric slopes near the entrance (Fig.
 399 14).

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

407 **6 Discussion on 3D effects and baroclinic adjustment**

408 **6.1 Overall comparison for total water level**

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

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

	Total water level MAE (m) Entire period adjustment period	M₂ amplitude MAE (m)	M₂ 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
---------------	-------------	------	------

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

425 6.2 Role of large-scale processes

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

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

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

450 **6.3 Baroclinic vs barotropic pressure gradients**

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

$$455 \overline{F_{BC}} = \frac{1}{H} \int_{-h}^{\eta} F_{BC}(z) dz, \text{ with } F_{BC}(z) = -\frac{g}{\rho_0} \int_z^{\eta} \nabla \rho d\zeta,$$

456 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
 457 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
 458 bathymetry in $[m]$; $H = \eta + h$ is the total water depth in $[m]$. The barotropic PGF is denoted as $F_{BT} =$
 459 $-g\nabla\eta$.

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

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

474 **6.4 Computational performance**

475 We briefly remark on the relative efficiency of the three configurations. The baseline model
476 achieves a performance of ~80 times faster than real time, using 1440 cores of NASA's Pleiades. The
477 3D barotropic model runs 190 time faster than real time on 960 cores, and the 2D barotropic model runs
478 230 time faster than real time on 72 cores. Therefore, the 2D and 3D barotropic models are
479 approximately 57 and 3.5 times faster than the baseline respectively.

480 **6.5 Summary and future work**

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

486 **7 Conclusion**

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

503 Acknowledgement

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

513 Supplementary materials

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

516 List of Figures

- 517 Fig. 1: Weather along the US east coast on Aug 28, 2011 (around Hurricane Irene) and Sept 8, 2011 (around
518 Tropical Storm Lee), showing large precipitations on both dates. The contours in the surface weather maps show
519 air pressure in millibar. The 24-hour precipitation is a record of the past 24 hours until the shown time. Hurricane
520 Katia (2011; bottom-left panel) did not land on the US east coast. Credit: NOAA Central Library U.S. Daily
521 Weather Maps Project (<https://www.wpc.ncep.noaa.gov/dailywxmap/explanation.html>); partial views of the
522 original online maps.
- 523 Fig. 2: The influence of Hurricane Irene on the Delaware Bay: (a) wind patterns near the Bay mouth (see Fig. 3
524 for the location of the station NDBC 44009); (b) streamflows at two USGS gauges (locations shown in Fig. 3; the
525 Delaware River station is at Trenton NJ), with two peaks corresponding to the landfall of Irene (2011) and the
526 subsequent river flooding under Tropical Storm Lee (2011).
- 527 Fig. 3: Model domain, bathymetry and locations of observation stations. Additionally, sample points (“#1”, “#2”,
528 “#3” and “A”, “B”, “C”) used in subsequent analysis are marked.
- 529 Fig. 4: Illustration of horizontal grid generation: (a) “feature arcs” in SMS, used to align the grid elements with
530 channels and follow the NWM segments; (b) zoomed-in view on the arcs and the grid, with the arcs
531 corresponding to NWM segments highlighted in red.
- 532 Fig. 5: Illustration of the vertical grid along a creek-bay-ocean transect: (a) transect location; (b) along-transect
533 view of the vertical grid; (c) zoomed-in view on the shallow portion of (b).
- 534 Fig. 6: Shapiro filter strength in a region (box in the inset) with steep bathymetry. The maximum strength is set as
535 0.5 (Zhang et al. 2016).
- 536 Fig. 7: The coupling between the NWM and SCHISM: (a) NWM segments and SCHISM land boundaries for the
537 Delaware Bay; (b) zoomed-in view on the NWM segments that intersect the SCHISM land boundary in the upper
538 Delaware Bay.
- 539 Fig. 8: Model-data comparison of surface elevations.
- 540 Fig. 9: Tidal harmonics of 4 major constituents calculated for Days 10-50 of the simulation period: (a) amplitude;
541 (b) phase. See Fig. 3 for station locations. The results from two sensitivity runs (3D barotropic and 2D barotropic)
542 are also included here for future reference.
- 543 Fig. 10: Comparison of SST on 2011-09-07. The MAEs throughout the domain are 0.61°C for HYCOM and
544 0.70°C for SCHISM.
- 545 Fig. 11: Comparison of salinity for 1984 survey, at multiple stations and multiple vertical positions (‘B’ denotes
546 bottom, ‘M’ denotes mid-depth; otherwise surface). Note that the time axes are different for different stations. The
547 overall MAE is 0.86 PSU. See Fig. 3 for station locations.
- 548 Fig. 12: Depth averaged salinity for the last 70 days of the 1984 simulation.
- 549 Fig. 13: Comparison of (a,b) significant wave height, and (c,d) peak period at two NDBC buoys. See Fig. 3 for
550 station locations.
- 551 Fig. 14: The wave effect: (a) differences in maximum elevation between ‘base+wave’ and baseline, with the
552 largest differences found in two regions and pointed out by black arrows; (b) strong wave breaking near the steep
553 slopes (cf. the bathymetry in Fig. 3). The locations of stations A-C in Fig. 3 are repeated here to correlate them
554 with the wave influence.
- 555 Fig. 15: Comparison of total water elevation at five stations between the baseline (no wave) and the “base+wave”
556 (with wave) results. See Fig. 3b and Fig. 14 for station locations.

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.

574 **References**

- 575 Amante, C. and Eakins, B.W., 2009. ETOPO1 arc-minute global relief model: procedures, data sources
576 and analysis.
- 577 Ardhuin, F., Rogers, E., Babanin, A.V., Filipot, J.F., Magne, R., Roland, A., Van Der Westhuysen, A.,
578 Queffeuilou, P., Lefevre, J.M., Aouf, L. and Collard, F., 2010. Semiempirical dissipation source
579 functions for ocean waves. Part I: Definition, calibration, and validation. *Journal of Physical*
580 *Oceanography*, 40(9), pp.1917-1941.
- 581 Bennis, A.-C., Ardhuin, F., Dumas, F., 2011. On the coupling of wave and three-dimensional
582 circulation models: choice of theoretical framework, practical implementation and adiabatic tests.
583 *Ocean Model.* 40 (3), 260–272.
- 584 Blake, E.S., 2007. The deadliest, costliest and most intense United States tropical cyclones from 1851
585 to 2006 (and other frequently requested hurricane facts). NOAA Technical Memorandum NWS TPC
586 5, 43 pp.
- 587 Carrere, L., Lyard, F., Cancet, M., Guillot, A. and Picot, N., 2016, May. FES 2014, a new tidal model—
588 Validation results and perspectives for improvements. In *Proceedings of the ESA living planet*
589 *symposium* (pp. 9-13).
- 590 Chen, C., Beardsley, R.C., Luettich, R.A., Westerink, J.J., Wang, H., Perrie, W., Xu, Q., Donahue,
591 A.S., Qi, J., Lin, H. and Zhao, L., 2013. Extratropical storm inundation testbed: Intermodel
592 comparisons in Scituate, Massachusetts. *Journal of Geophysical Research: Oceans*, 118(10),
593 pp.5054-5073.
- 594 Cho, K.H., Wang, H.V., Shen, J., Valle-Levinson, A. and Teng, Y.C., 2012. A modeling study on the
595 response of Chesapeake Bay to hurricane events of Floyd and Isabel. *Ocean Modelling*, 49, pp.22-
596 46.
- 597 Craig, P. D., and M. L. Banner (1994), Modeling wave-enhanced turbulence in the ocean surface layer,
598 *J. Phys. Oceanogr.*, 24(12), 2546–2559.
- 599 Danielson, J.J., Poppenga, S.K., Tyler, D.J., Palaseanu-Lovejoy, M. and Gesch, D.B., 2018. Coastal
600 National Elevation Database (No. 2018-3037). US Geological Survey.
- 601 Danilov, S., 2012. Two finite-volume unstructured mesh models for large-scale ocean modeling. *Ocean*
602 *Model.* 47, 14–25.
- 603 Du, J., Shen, J., Zhang, Y.J., Ye, F., Liu, Z., Wang, Z., Wang, Y.P., Yu, X., Sisson, M. and Wang,
604 H.V., 2018. Tidal Response to Sea - Level Rise in Different Types of Estuaries: The Importance of
605 Length, Bathymetry, and Geometry. *Geophysical Research Letters*, 45(1), pp.227-235.
- 606 Du, J., Park, K., 2019. Estuarine salinity recovery from an extreme precipitation event: Hurricane
607 Harvey in Galveston Bay. *Sci. Total Environ.* 670, 1049–1059.
- 608 Ezer, T., 2013. Sea level rise, spatially uneven and temporally unsteady: Why the US East Coast, the
609 global tide gauge record, and the global altimeter data show different trends. *Geophysical Research*
610 *Letters*, 40(20), pp.5439-5444.
- 611 Ezer, T., 2018. On the interaction between a hurricane, the Gulf Stream and coastal sea level. *Ocean*
612 *Dynamics*, 68(10), pp.1259-1272.

- 613 Ezer, T., 2019. Numerical modeling of the impact of hurricanes on ocean dynamics: sensitivity of the
614 Gulf Stream response to storm's track, *Ocean Dynamics* (in press).
- 615 Friedrichs, C.T., 2010. Barotropic tides in channelized estuaries. *Contemporary issues in estuarine*
616 *physics*, pp.27-61.
- 617 Garvine, R.W., McCarthy, R.K. and Wong, K.C., 1992. The axial salinity distribution in the Delaware
618 estuary and its weak response to river discharge. *Estuarine, Coastal and Shelf Science*, 35(2),
619 pp.157-165.
- 620 Gebert, J.A. and Searfoss, R., 2012. Chapter 4 Sediment, Technical Report for the Delaware Estuary
621 and Basin, Partnership for the Delaware Estuary, Report No. 12-01, pp. 107-118.
- 622 Guérin, T., Bertin, X., Coulombier, T. and de Bakker, A., 2018. Impacts of wave-induced circulation in
623 the surf zone on wave setup. *Ocean Modelling*, 123, pp.86-97.
- 624 Harleman, D. R. F., 1966. Real estuaries. *Estuary and Coastline Hydrodynamics*, A. T. Ippen, Ed.,
625 McGraw-Hill, 522-545.
- 626 Hasselmann, K., Barnett, T.P., Bouws, E., Carlson, H., Cartwright, D.E., Enke, K., Ewing, J.A.,
627 Gienapp, H., Hasselmann, D.E., Kruseman, P. and Meerburg, A., 1973. Measurements of wind-
628 wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP).
629 *Ergänzungsheft* 8-12.
- 630 Hwang, P.A., 2018. High-wind drag coefficient and whitecap coverage derived from microwave
631 radiometer observations in tropical cyclones. *Journal of Physical Oceanography*, 48(10), pp.2221-
632 2232.
- 633 Kerr, P.C., Donahue, A.S., Westerink, J.J., Luettich Jr, R.A., Zheng, L.Y., Weisberg, R.H., Huang, Y.,
634 Wang, H.V., Teng, Y., Forrest, D.R. and Roland, A., 2013. US IOOS coastal and ocean modeling
635 testbed: Inter - model evaluation of tides, waves, and hurricane surge in the Gulf of Mexico. *Journal*
636 *of Geophysical Research: Oceans*, 118(10), pp.5129-5172.
- 637 Kennedy, A.B., Westerink, J.J., Smith, J.M., Hope, M.E., Hartman, M., Taflanidis, A.A., Tanaka, S.,
638 Westerink, H., Cheung, K.F., Smith, T. and Hamann, M., 2012. Tropical cyclone inundation
639 potential on the Hawaiian Islands of Oahu and Kauai. *Ocean Modelling*, 52, pp.54-68.
- 640 Knutson, T.R., McBride, J.L., Chan, J., Emanuel, K., Holland, G., Landsea, C., Held, I., Kossin, J.P.,
641 Srivastava, A.K. and Sugi, M., 2010. Tropical cyclones and climate change. *Nature geoscience*, 3(3),
642 p.157.
- 643 Le Roux, D.Y., Sene, A., Rostand, V. and Hanert, E., 2005. On some spurious mode issues in shallow-
644 water models using a linear algebra approach. *Ocean Modelling*, 10(1-2), pp.83-94.
- 645 Li, M., Zhong, L., Boicourt, W.C., Zhang, S. and Zhang, D.L., 2006. Hurricane-induced storm surges,
646 currents and destratification in a semi - enclosed bay. *Geophysical Research Letters*, 33(2).
- 647 Magnusson, L., Bidlot, J.R., Lang, S.T., Thorpe, A., Wedi, N. and Yamaguchi, M., 2014. Evaluation of
648 medium-range forecasts for Hurricane Sandy. *Monthly Weather Review*, 142(5), pp.1962-1981.
- 649 Minato, S., 1998. Storm Surge Simulation using POM and a Revisitation of Dynamics of Sea Surface
650 Elevation Short-Term Variation." *Papers in Meteorology and Geophysics* 48, no. 3 (1998): 79-88.
- 651 Münchow, A. and Garvine, R.W., 1993. Dynamical properties of a buoyancy - driven coastal current.
652 *Journal of Geophysical Research: Oceans*, 98(C11), pp.20063-20077.

- 653 Orton, P., Georgas, N., Blumberg, A. and Pullen, J., 2012. Detailed modeling of recent severe storm
654 tides in estuaries of the New York City region. *Journal of Geophysical Research: Oceans*, 117(C9).
- 655 Patchen, R., 2007, November. Establishment of a Delaware Bay Model Evaluation Environment.
656 Estuarine and Coastal Modeling. In *Proceedings of the Tenth International Conference*, by Malcolm
657 L. Spaulding, PE, Ph. D.,(editor).
- 658 Pond, S., Pickard, G.L., 1998. *Introductory Dynamical Oceanography*. Butterworth-Heinmann,
659 Stoneham, MA.
- 660 Rasche, N. and Ardhuin, F., 2013. A global wave parameter database for geophysical applications. Part
661 2: Model validation with improved source term parameterization. *Ocean Modelling*, 70, pp.174-188.
- 662 Roland, A., Zhang, Y.J., Wang, H.V., Meng, Y., Teng, Y.C., Maderich, V., Brovchenko, I., Dutour -
663 Sikiric, M. and Zanke, U., 2012. A fully coupled 3D wave - current interaction model on
664 unstructured grids. *Journal of Geophysical Research: Oceans*, 117(C11).
- 665 Shapiro, R., 1970. Smoothing, filtering, and boundary effects. *Reviews of geophysics*, 8(2), pp.359-
666 387.
- 667 Sharma, S., Siddique, R., Reed, S., Ahnert, P. and Mejia, A., 2019. Hydrological Model Diversity
668 Enhances Streamflow Forecast Skill at Short - to Medium - Range Timescales. *Water Resources*
669 *Research*, 55(2), pp.1510-1530.
- 670 Sharp, J.H., 1983. *The Delaware estuary: Research as background for estuarine management and*
671 *development*. Univ. Delaware and New Jersey Mar. Sci. Consort.
- 672 Soulsby, R. 1997. *Dynamics of marine sands*, Thomas Telford, London.
- 673 Teng, J., Jakeman, A.J., Vaze, J., Croke, B.F., Dutta, D. and Kim, S., 2017. Flood inundation
674 modelling: A review of methods, recent advances and uncertainty analysis. *Environmental*
675 *modelling & software*, 90, pp.201-216.
- 676 Umlauf, L. and Burchard, H., 2003. A generic length-scale equation for geophysical turbulence models.
677 *Journal of Marine Research*, 61(2), pp.235-265.
- 678 Wahl, T., Jain, S., Bender, J., Meyers, S.D. and Luther, M.E., 2015. Increasing risk of compound
679 flooding from storm surge and rainfall for major US cities. *Nature Climate Change*, 5(12), p.1093.
- 680 Whitney, M.M. and Garvine, R.W., 2006. Simulating the Delaware Bay buoyant outflow: Comparison
681 with observations. *Journal of Physical Oceanography*, 36(1), pp.3-21.
- 682 Wolf, J., 2009. Coastal flooding: impacts of coupled wave–surge–tide models. *Natural Hazards*, 49(2),
683 pp.241-260.
- 684 Wong, K.C., 1995. The hydrography at the mouth of Delaware Bay: Tidally averaged distribution and
685 intratidal variability. *Estuarine, Coastal and Shelf Science*, 41(6), pp.719-736.
- 686 Wong, K.C. and Münchow, A., 1995. Buoyancy forced interaction between estuary and inner shelf:
687 observation. *Continental Shelf Research*, 15(1), pp.59-88.
- 688 Yankovsky, A.E. and Chapman, D.C., 1997. A simple theory for the fate of buoyant coastal discharges.
689 *Journal of Physical oceanography*, 27(7), pp.1386-1401.
- 690 Ye, F., Zhang, Y.J., Wang, H.V., Friedrichs, M.A., Irby, I.D., Alteljevich, E., Valle-Levinson, A.,
691 Wang, Z., Huang, H., Shen, J. and Du, J., 2018. A 3D unstructured-grid model for Chesapeake Bay:
692 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.

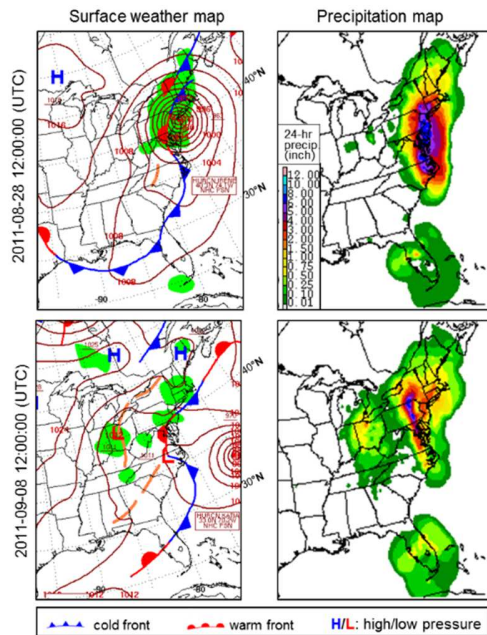


Fig. 1: Weather along the US east coast on Aug 28, 2011 (around Hurricane Irene) and Sept 8, 2011 (around Tropical Storm Lee), showing large precipitations on both dates. The contours in the surface weather maps show air pressure in millibar. The 24-hour precipitation is a record of the past 24 hours 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/explanation.html>); partial views of the original online maps.

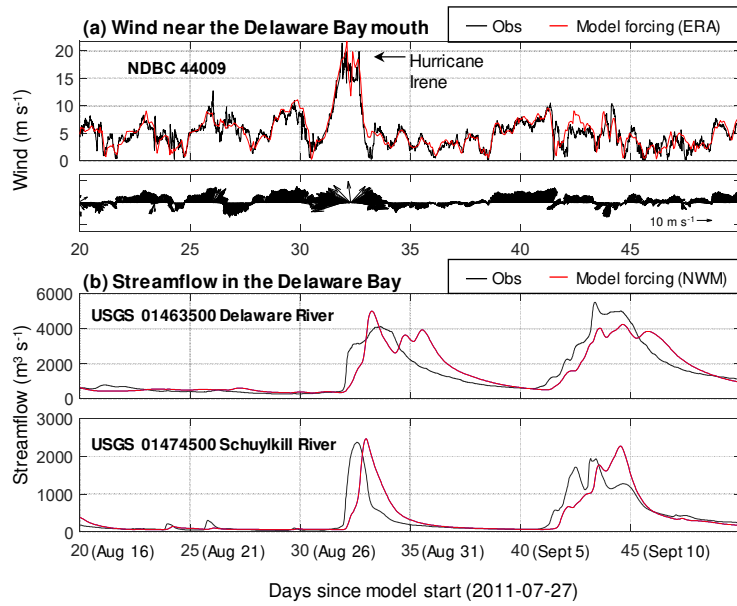


Fig. 2: The influence of Hurricane Irene on the Delaware Bay: (a) wind patterns near the Bay mouth (see Fig. 3 for the location of the station NDBC 44009); (b) streamflows at two USGS gauges (locations shown in Fig. 3; the Delaware River station is at Trenton NJ), with two peaks corresponding to the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

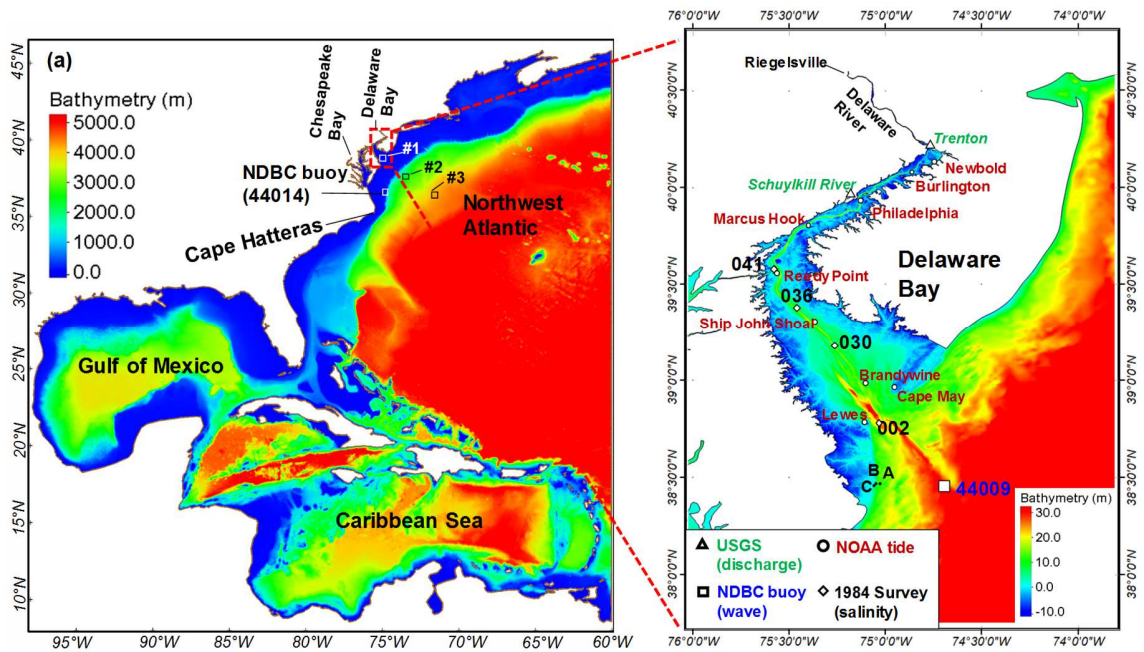


Fig. 3: Model domain, bathymetry and locations of observation stations. Additionally, sample points (“#1”, “#2”, “#3” and “A”, “B”, “C”) used in subsequent analysis are marked.

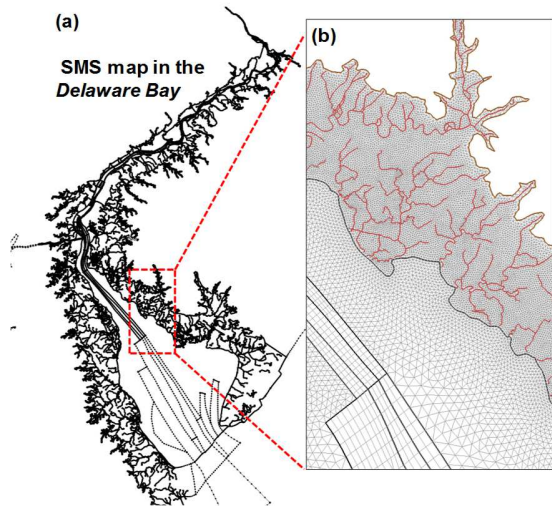


Fig. 4: Illustration of horizontal grid generation: (a) “feature arcs” in SMS, used to align the grid elements with channels and follow the NWM segments; (b) zoomed-in view on the arcs and the grid, with the arcs corresponding to NWM segments highlighted in red.

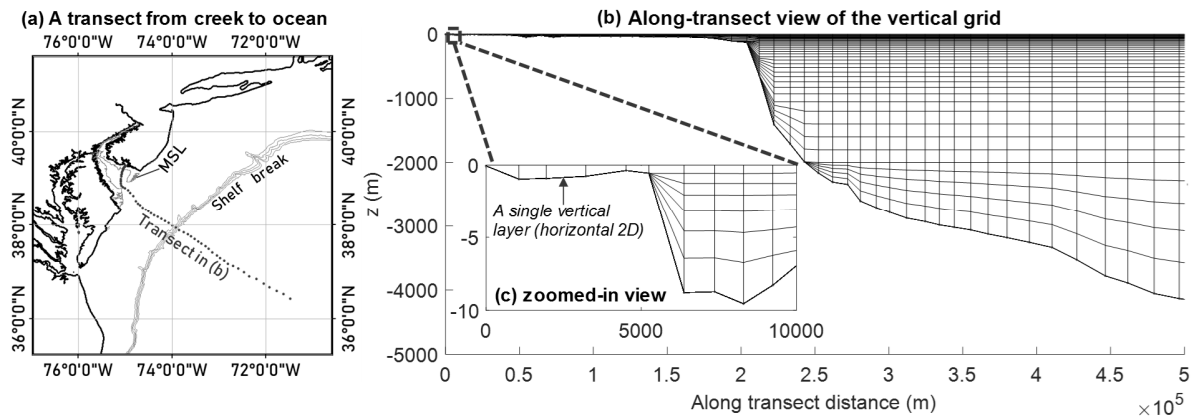


Fig. 5: Illustration of the vertical grid along a creek-bay-ocean transect: (a) transect location; (b) along-transect view of the vertical grid; (c) zoomed-in view on the shallow portion of (b).

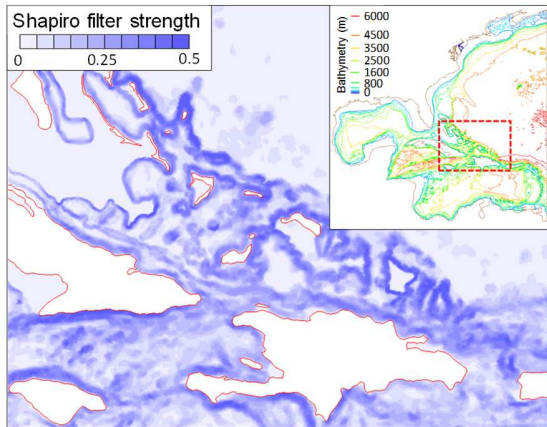


Fig. 6: Shapiro filter strength in a region (box in the inset) with steep bathymetry. The maximum strength is set as 0.5 (Zhang et al. 2016).

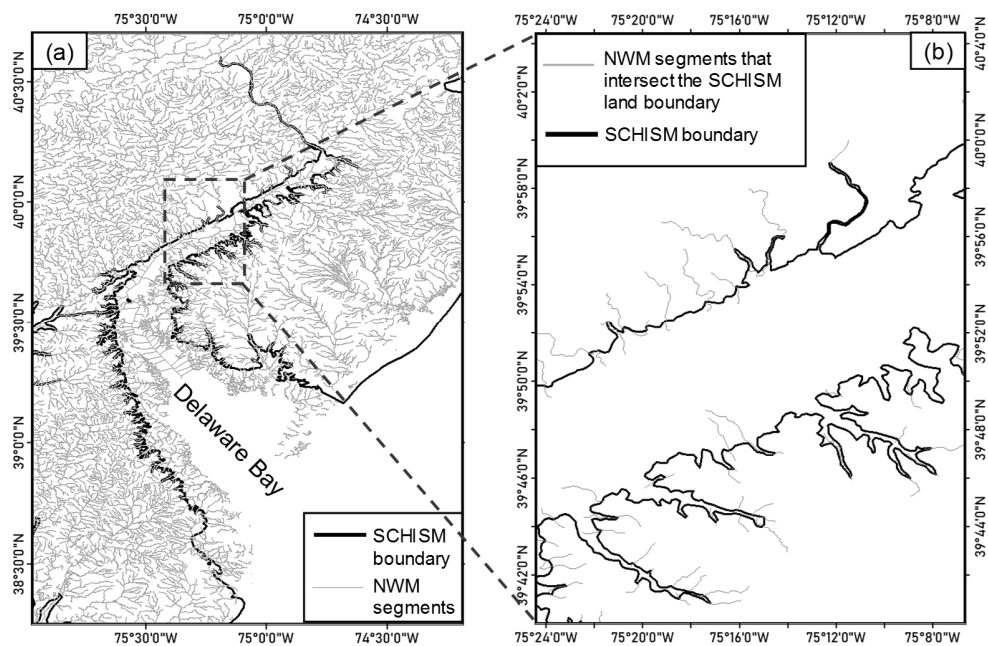


Fig. 7: The coupling between the NWM and SCHISM: (a) NWM segments and SCHISM land boundaries for the Delaware Bay; (b) zoomed-in view on the NWM segments that intersect the SCHISM land boundary in the upper Delaware Bay.

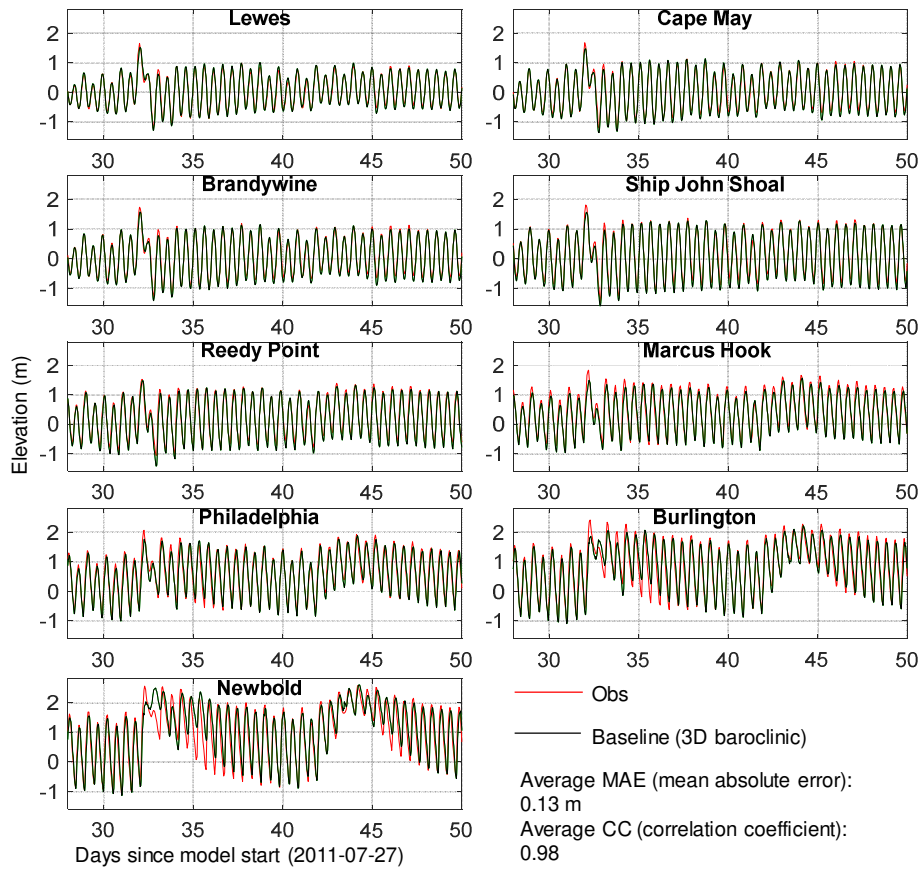


Fig. 8: Model-data comparison of surface elevations.

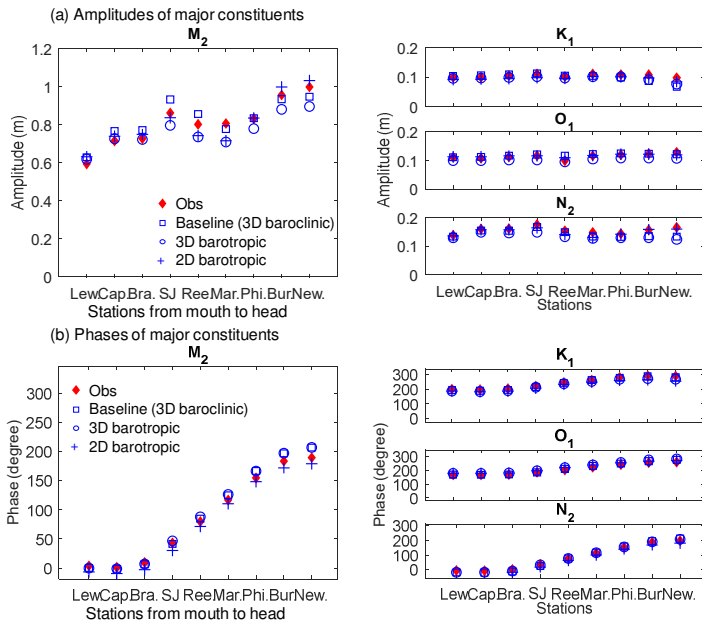


Fig. 9: Tidal harmonics of 4 major constituents calculated for Days 10-50 of the simulation period: (a) amplitude; (b) phase. See Fig. 3 for station locations. The results from two sensitivity runs (3D barotropic and 2D barotropic) are also included here for future reference.

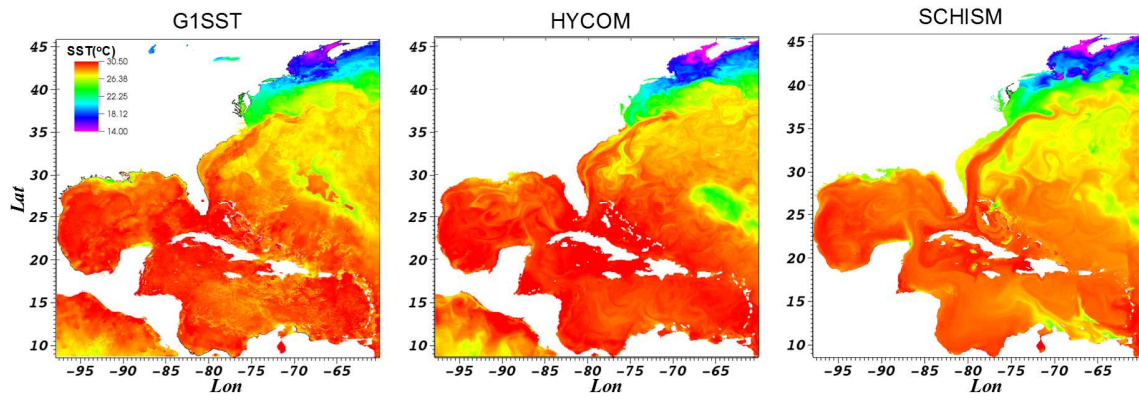


Fig. 10: Comparison of SST on 2011-09-07. The MAEs throughout the domain are 0.61°C for HYCOM and 0.70°C for SCHISM.

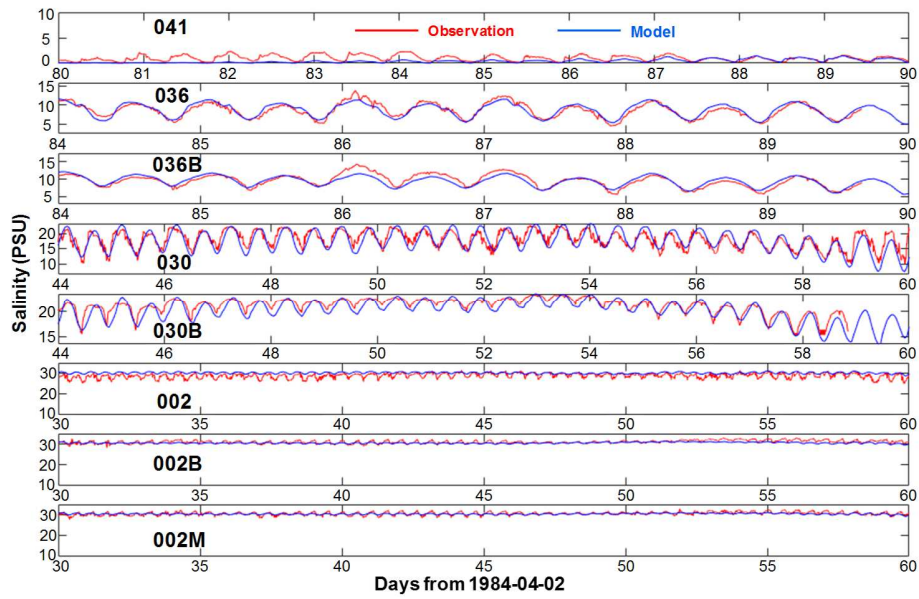


Fig. 11: Comparison of salinity for 1984 survey, at multiple stations and multiple vertical positions ('B' denotes bottom, 'M' denotes mid-depth; otherwise surface). Note that the time axes are different for different stations. The overall MAE is 0.86 PSU. See Fig. 3 for station locations.

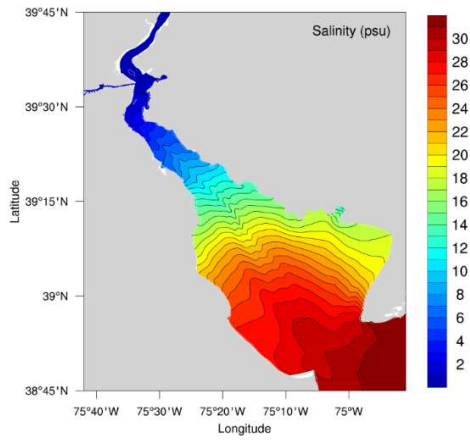


Fig. 12: Depth averaged salinity for the last 70 days of the 1984 simulation.

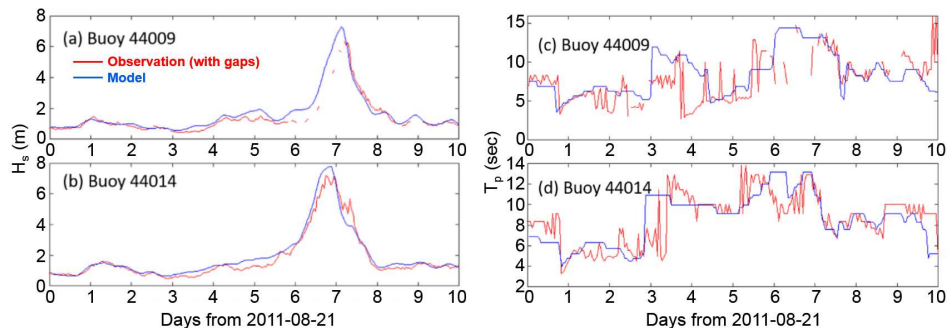


Fig. 13: Comparison of (a,b) significant wave height, and (c,d) peak period at two NDBC buoys. See Fig. 3 for station locations.

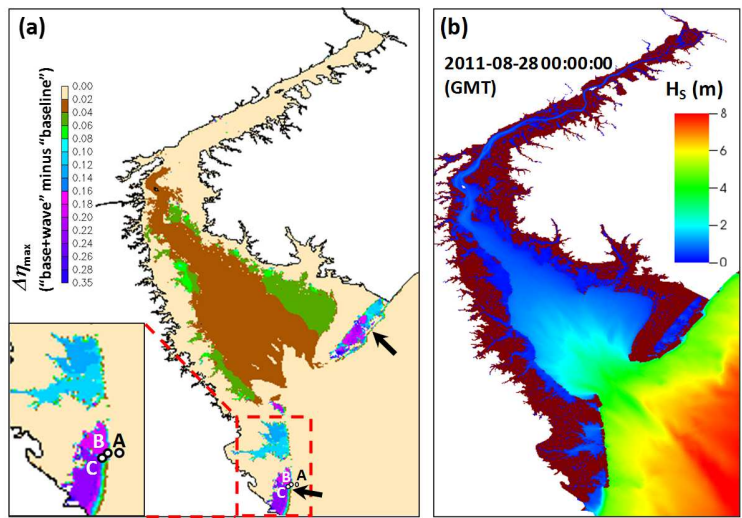


Fig. 14: The wave effect: (a) differences in maximum elevation between 'base+wave' and baseline, with the largest differences found in two regions and pointed out by black arrows; (b) strong wave breaking near the steep slopes (cf. the bathymetry in Fig. 3). The locations of stations A-C in Fig. 3 are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

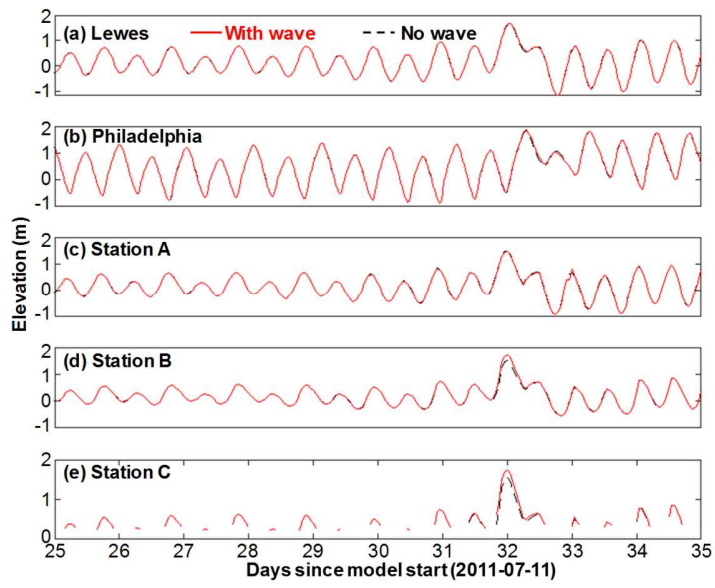


Fig. 15: Comparison of total water elevation at five stations between the “baseline” (no wave) and the “base+wave” (with wave) results. See Fig. 3b and Fig. 14 for station locations.

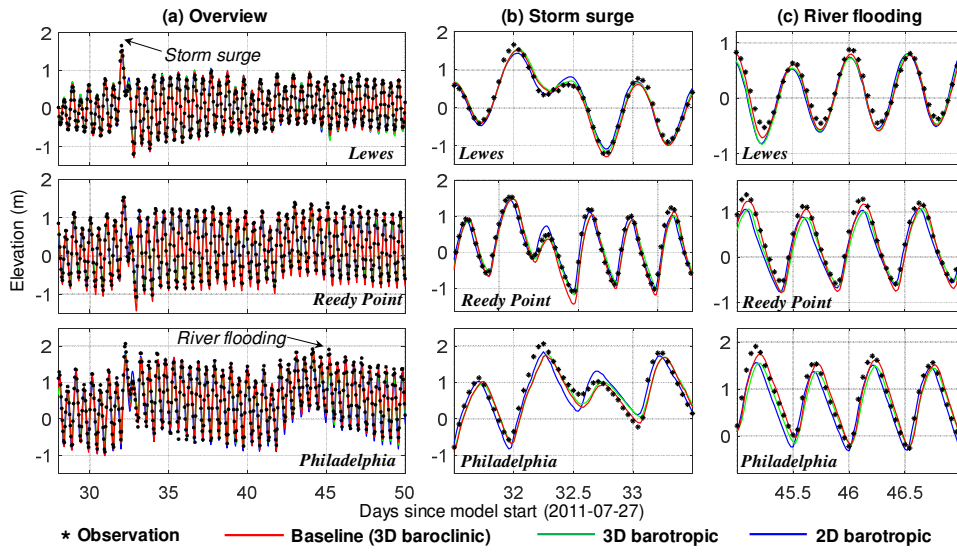


Fig. 16: Comparison of the simulated elevations at 3 stations in the lower, mid- and upper Bay between “baseline” and two sensitivity tests. (a) Overview; (b) zoomed-in view during the storm surge of Hurricane Irene; (c) zoomed-in view during the river flooding period. See Fig. 3b for station locations.

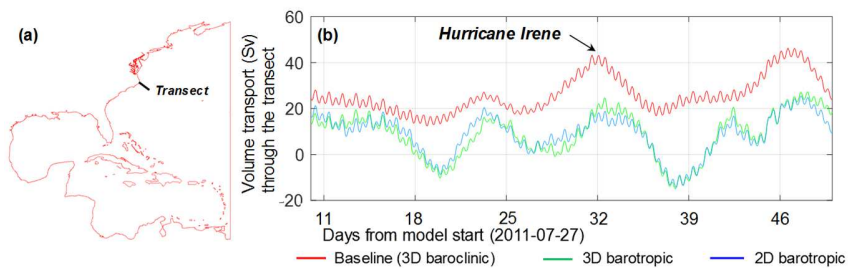


Fig. 17: Gulf Stream volume transport: (a) locations of the transect near Cape Hatteras; (b) time-series of the volume transport for the “baseline” and two sensitivity tests.

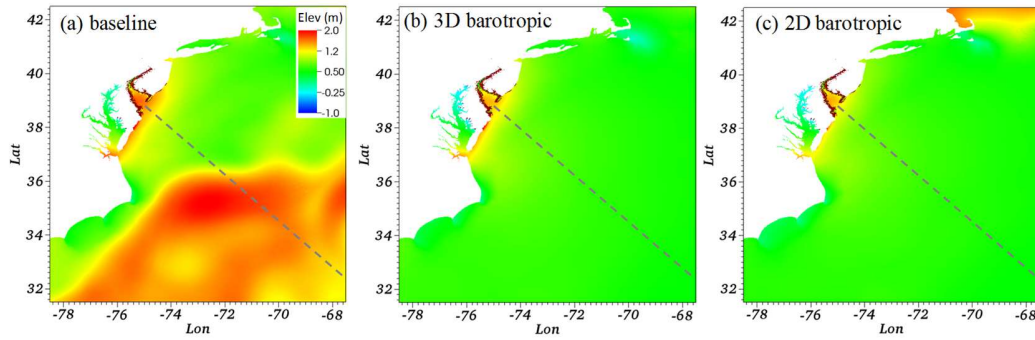


Fig. 18: Comparison of SSH at 2011-08-28 00:00:00 (UTC), calculated from: (a) baseline; (b) 3D barotropic; and (c) 2D barotropic models. The dash line in each sub-plot marks a transect used in the subsequent analysis.

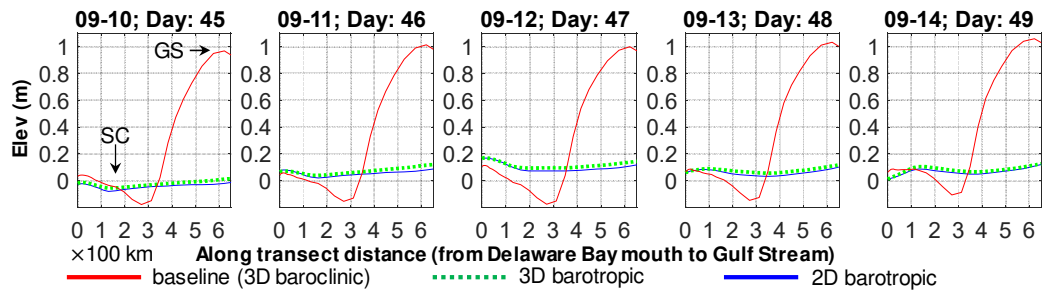


Fig. 19: Snapshots of sea surface slope along a cross-shore transect (location marked in Fig. 18) during a post-storm period. The MAB slope current (SC) and Gulf Stream (GS) are marked on the surface slope in the baseline. The two barotropic runs show larger temporal swings near the Delaware Bay mouth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

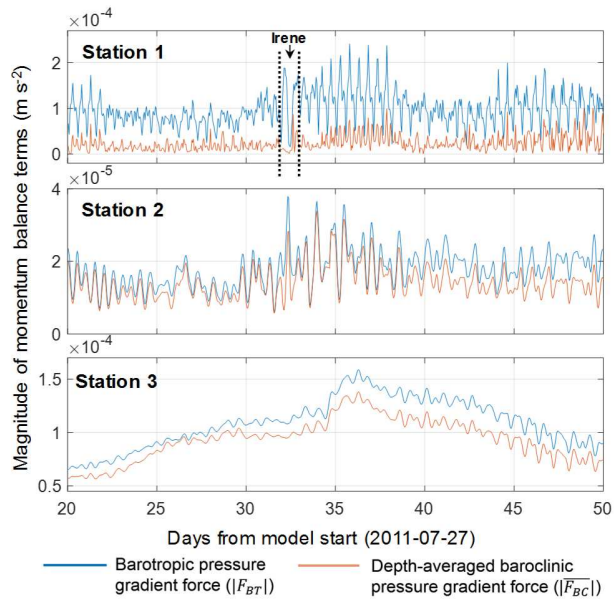


Fig. 20: Relative importance of the barotropic pressure gradient force and the depth-averaged baroclinic pressure gradient force ($\overline{F_{BC}}$): time-series at three representative stations, showing the *magnitudes* of the two terms. The station locations are shown in Fig. 3.

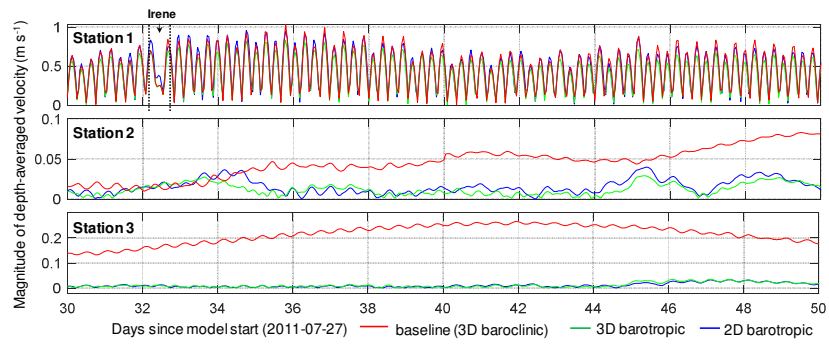


Fig. 21: Comparison of the magnitude of the depth averaged velocity between the baseline and two sensitivity runs at 3 stations. The station locations are shown in Fig. 3a.