Weather and Climate Extremes 38 (2022) 100504

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Ensemble forecast for storm tide and resurgence from Tropical Cyclone Isaias

Mahmoud Ayyad *, Philip M. Orton, Hoda El Safty, Ziyu Chen, Muhammad R. Hajj

Davidson Laboratory, Department of Civil, Environmental and Ocean Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA

ARTICLE INFO

Keywords:
Ensemble forecasting
Tropical cyclone
Hurricane Isaias
Resurgence
Storm tide

ABSTRACT

Ensemble-based probabilistic forecasting of storm surge is increasingly being used to provide metrics for emergency management decisions such as the near-worst case scenario. The Stevens Flood Advisory System is an ensemble prediction system used to forecast total water levels over a broad coastal region and street-scale flood levels for several New York Harbor (NYH) critical infrastructure sites. As a part of our continuous assessment of this system's performance, we evaluate its prediction of storm tide and resurgence during Tropical Cyclone Isaias (2020), which tracked northward along the Pennsylvania/New Jersey border and caused the largest storm surge in NYH since Hurricane Sandy. Isaias specific track and speed generated an unusual flood event consisting of a storm surge, a blowout, then a significant resurgence that caused minor flooding. The analysis shows that the super-ensemble spread provided an equal or better estimate of uncertainties than sub-ensembles based only on any single meteorological forcing system. Because of ensemble averaging, the central forecast under-predicted peak water levels and the resurgence peak though these were predicted by some of the ensemble members. The impacts of errors in forecast storm arrival time and resolution-related biases in coarse global atmospheric models on the predictions are noted. A limited comparison for this single storm with the National Hurricane Center's forecast show SFAS providing better accuracy and spread. Advantages and challenges of SFAS and other similar mid-latitude flood forecast systems are identified along with recommendations for analysis and improvement.

1. Introduction

Ensemble-based probabilistic forecasting of storm surge based on a cluster of "member" forecasts generated from perturbed meteorological forcing conditions and/or different forecasting models are, in general, expected to cover uncertainties associated with deterministic forecasts that are based on a specific forecast model and a single set of forcing conditions. Probabilistic forecasts are particularly useful for events with higher uncertainties, such as Tropical Cyclones (TCs) where a small change in the storm speed or direction can lead to large variations in storm surge over a broad coastal region (Ayyad et al., 2021). This is even more true for areas like the US Northeast where moderate-to-large tide ranges significantly impact total water levels (Georgas et al., 2014; Colle et al., 2015). Uncertainties generated from ensemble forecasts may be used to address risk thresholds based on specific criteria. For instance, evacuation decisions, management of port operations, and/or deployment of flood gates or barriers can be set based on the 90th percentile event (10% chance exceedance) rather than on a single prediction from a central forecast.

Ensemble-based prediction has substantially advanced coastal flood forecasting and early warning systems in mid-latitude and TC flood forecast systems. Mid-latitude forecast systems are developed for Extra-Tropical Cyclones (ETC) using regional or global meteorological forcing. TC coastal flood forecast systems operate only during TC conditions and use storm-following atmospheric model-based forcing that has sufficient resolution to capture small-scale variations such as hurricane eye walls. Fitting the latter category is the Probabilistic Hurricane Storm Surge (P-Surge) (Taylor and Glahn, 2008) developed by the National Oceanic and Atmospheric Administration (NOAA). P-Surge is based on the SLOSH (Sea, Lake, and Overland Surges from Hurricanes) model (Jelesnianski, 1992) and uses data from the National Hurricane Center (NHC)'s official advisory to create a set of synthetic storms by perturbing the storm's position, size, and intensity based on past errors of the advisories (Taylor and Glahn, 2008; Gonzalez and Taylor, 2021). P-Surge computes the probable storm surge 102 h in advance. It provides the hourly maximum water level at 10 percentile, and 50 to 90 percentiles with 10% interval. P-Surge generated a good forecast spread and accuracy of storm surge for Hurricane Sandy (Forbes et al., 2014).

In the mid-latitude category, the UK Meteorological Office has been running an ensemble-based forecasting system for coastal water

* Corresponding author.
E-mail address: mayyad@stevens.edu (M. Ayyad).

https://doi.org/10.1016/j.wace.2022.100504
Received 30 May 2022; Received in revised form 5 September 2022; Accepted 6 September 2022
Available online 10 September 2022

levels since 2006 that was made operational in 2009 (Flowerdew et al., 2010). Its lead time was initially 54 h and later extended to 124 h (Flowerdew et al., 2013). The European Center for Medium-Range Weather Forecasts (ECMWF-ENS) (Molteni et al., 1996) is used for meteorological forcing. The Dutch Meteorological Institute developed a 48-hour ensemble-based forecasting system for water levels along the Dutch coast (de Vries, 2009) also based on ECMWF-ENS. The Meteorological Development Laboratory (MDL) developed the Extra-Tropical Storm Surge (ETSS) forecast system based on SLOSH with some modifications using the Global Forecast System (GFS) wind and pressure input (Kim et al., 1996). In 2017, MDL implemented the Probabilistic Extra-Tropical Storm Surge (P-ETSS) forecast system (Liu et al., 2018) using ensemble forcing from the Global Ensemble Forecast System (GEFS). In 2019, P-ETSS started to utilize ensemble members from both GEFS and Canadian Meteorological Center (CMC), calling this combination of ensembles the North American Ensemble Forecast System (NAEFS) (Liu et al., 2019).

This paper focuses on the Stevens Flood Advisory System (SFAS, http://stevens.edu/SFAS), an ensemble-based forecasting system developed at the Stevens Institute of Technology. SFAS is the evolution of the publicly-available coastal ocean forecasting system initiated in 2007 with the New York Harbor Observing and Prediction System (Georgas and Blumberg, 2010; Orton et al., 2012; Georgas et al., 2014). The system models rainfall-driven hydrology, tide, and storm surge to predict total water levels (Georgas et al., 2016; Jordi et al., 2019). The "super-ensemble" forecasts have been running since late 2015 and are presently forced by a combination of ensemble and deterministic weather forecast products; namely GFS, GEFS, ECMWF-ENS, ECMWF-HRES (high-resolution), North American Mesoscale forecast system (NAM), and CMC. The CMC product is called the Canadian Global Ensemble Prediction System (GEPS). SFAS provides the 90% confidence intervals and central estimate time series of water level for the upcoming 108 h. The National Weather Service local weather forecast offices in Upton (New York) and Mt. Holly (NJ/Eastern Pennsylvania), and two other offices have been using SFAS' 5th–50th–95th percentile time series data alongside NOAA's forecast products year-round to develop their forecast guidance. The system's output is also used to set the boundary conditions for Stevens Institute's local two-dimensional street-scale flood modeling and mapping for localized domains around transportation and port facilities of the New York City metro area (Jordi et al., 2019).

Building on prior assessments of the deterministic forecast system (Georgas and Blumberg, 2010), Georgas et al. (2016) assessed SFAS ensemble-based central forecast performance for winter-season forecasts. Saleh et al. (2016) and Jordi et al. (2019) demonstrated through retrospective ensemble forecasts of TCs Irene and Sandy, respectively, that SFAS is capable to predict impact of extreme hydrologic events and flood forecasting at street scale. Saleh et al. (2016) showed that SFAS improved predictions of river discharge forecasts due to extreme hydrologic events. Jordi et al. (2019) determined that the 95th percentile flood forecast at street scale for Sandy was in good agreement with observations three days before the peak in flood level, while the 50th percentile was negatively biased because of the low resolution of the meteorological forcing.

As a part of the continuous assessment of SFAS' capabilities, there is a need to evaluate forecasts of challenging unusual coastal flood events. An annual ensemble water level forecast assessment is available for the years 2020 and 2021 (Orton et al., 2021) available at http://stevens.edu/SFAS, and tropical storm Isaias (2020) stood out to be the most challenging event in those two years. TC Isaias affected the region on August 4th, 2020. This TC brought the largest storm surge of 1.37 m at the Battery tide gauge since Hurricane Sandy in 2012, and as large as 1.83 m in the Hackensack River. Furthermore, Isaias caused a rapidly undulating positive surge, a blowout with negative surge, then a second positive surge or resurgence. This resurgence caused peak water levels and flooding in certain areas of New York Harbor (NYH). We propose the term "resurgence" to refer more broadly to any secondary surge occurring after a storm and its peak surge pass and winds cease to support local wind setup. These secondary surges have been observed in NYH region after hurricane landfalls on Long Island. One example is the 1938 Long Island Express, which resulted in a peak surge of 0.94 m, followed by a resurgence of 0.56 m six hours later at the Battery.

Here, we evaluate the ability of SFAS to predict storm tide and resurgence, and we assess the benefits of the SFAS super-ensemble versus individual sub-ensembles in case of Isaias. Given that Isaias was a TC, we also compare SFAS and P-Surge forecasts. The assessment focuses on four locations that represent different sub-basins of NYH, namely Battery, Jamaica Bay at Inwood, Kings Point in NY and Hackensack River in NJ. The paper proceeds with a description of SFAS, ensemble processing, and performance evaluation metrics in Section 2. A summary of Isaias track and its effect is presented in Section 3. Evaluation of predicted surge levels by the super- and sub-ensembles against observation data is performed in Section 4. The causes of observed phenomena and associated modeling challenges are discussed in Section 5. Conclusions are drawn in Section 6.

2. Methods

2.1. Stevens flood advisory system (SFAS)

Coupled coastal-hydrologic total water level modeling

SFAS is a fully automated operational Hydrologic-Coastal Ensemble Prediction System that forecasts water levels across the US Mid-Atlantic and Northeast coastline. The hydrodynamic simulations are conducted using the Stevens Institute Estuarine and Coastal Ocean Model (sECOM) (Georgas and Blumberg, 2010; Orton et al., 2012; Georgas et al., 2016), which is a free-surface hydrostatic model with terrain-following (sigma) vertical coordinates and an orthogonal curvilinear Arakawa C-grid. For the central forecast region of New York Bight (NYB), the simulations are performed using a nested application. The New York Harbor Observing and Prediction System (NYHOPS) applies sECOM in a three-dimensional configuration on a domain that spans the continental shelf and several estuaries. Boundary conditions are applied to the NYHOPS domain at its offshore boundary (OBCs) and its interface with hydrologic models. The OBCs include the superposition of time series of tides from the ADCIRC East Coast tide constituent database (Szpilka et al., 2016) and storm surge modeled on the Stevens Northwest Atlantic Prediction (SNAP) domain (Georgas and Blumberg, 2010).

Two US Army Corps of Engineers (USACE) models, the Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS) and River Analysis System (HEC-RAS) are used for hydrologic modeling of precipitation-runoff processes for NY and northern NJ major and moderate-sized rivers and for dendritic watershed from area-scaled neighboring rivers, respectively (Georgas et al., 2016). NYHOPS also uses the forecast stream flows in other parts of the domain from the United Stated Geological Survey (USGS), National Weather Service Advanced Hydrologic Prediction Service, and Stevens-operated coastal sensors (Georgas et al., 2016).

SFAS uses an ensemble forecasting approach by running sECOM simulations for 96 different atmospheric forcing datasets. The ensemble meteorological forcing used in this study includes 21 members of GEFS with three hours temporal resolution and 0.5° spatial resolution, one member of GFS with three hours temporal resolution and 0.25° spatial resolution, one member of NAM with three hours temporal resolution and 0.1° spatial resolution, 21 members of CMC with six hours temporal resolution and 0.5° spatial resolution, 51 members of ECMWF-ENS with three hours temporal resolution and 0.25° spatial resolution, and one ECMWF-HRES with three hours temporal resolution and 0.125° spatial resolution. The meteorological data are spatially and temporally interpolated using bicubic and cubic splines, respectively, to create hourly forcing fields on each domain's grid.

SFAS provides public and private forecasts four times per day. At every cycle, SFAS updates water levels for five and a half days. This period is split into a hindcast day used for training purposes and a forecast period over four and a half days determined from ensemble processing methods as described below. The hydrologic and hydrodynamic simulations, ensemble analyses, and website graphics require approximately two hours to be processed. The time series of the 00:00, 06:00, 12:00 and 18:00 (UTC) forecasts for water level are posted on the SFAS website at 02:00, 08:00, 14:00 and 20:00 (UTC). Additional street-scale flood simulations and mapping are performed for the 5th, 50th, and 95th percentiles of temporal maxima.

Tide bias correction and ensemble processing methods

SFAS utilizes an innovative "tide bias correction" approach to improve the forecast results. Using a multi-year period of prior data, harmonic analysis is performed on every station's raw simulation output and observations. The difference in resulting tide time series is the tide bias correction dataset, and may be applied to all future forecasts (Georgas and Blumberg, 2010).

Different ensemble forecasting methods have been used in literature (Florescu and Tudor, 2013; Florescu, 2014; Raftery et al., 2005; Gneiting et al., 2005; Georgas et al., 2016; Whan et al., 2021; Cho et al., 2022). The adopted method in the SFAS is the model based on root mean square error. The super-ensemble (NYHOPS-E/central) forecast, η_w, is the weighted mean of the 96 members given by

$$\eta_w = \sum_{j=1}^{96} w^{(j)} \eta^{(j)} \tag{1}$$

where $w^{(j)}$ is the normalized weight factor of member j. The weights are calculated using the hindcast day. The normalized weight factor is defined as

$$w^{(j)} = \frac{f^{(j)}}{\sum_{j=1}^{96} f^{(j)}} \tag{2}$$

Here, $f^{(j)}$ is the weight factor of every member. It is calculated using the root mean square error ($RMSE$) given by (Georgas et al., 2016)

$$RMSE^{(j)} = \sqrt{\frac{1}{N^{(j)}} \sum_{i=1}^{N^{(j)}} \left(\eta_{m}^{(j)}{}_{i} - \eta_{oi}\right)^2} \tag{3}$$

where $N^{(j)}$ is the number of time series data points per member j in the one day hindcast, and η_m and η_o are the modeled and observed water levels, respectively. The weight factor is then calculated as

$$f^{(j)} = \frac{1}{\left(|\epsilon^{(j)}| + 0.05\right)\left(RMSE^{(j)} + 0.05\right)} \tag{4}$$

where $\epsilon^{(j)}$ is the mean bias between the observation and hindcast levels for the same period of time for every ensemble member. The small value 0.05 is added in the denominator to avoid singularities.

The 90th percentile confidence interval is calculated by finding the 5th and 95th percentiles empirically formed from all available and equally weighted ensemble members. The range of the ensemble accounts for atmospheric uncertainties across multiple ensembles, but does not cover ocean modeling uncertainties. As a rough approximation of the effect of additional ocean, wave and air–sea interaction uncertainties, the weighted RMSE from the hindcast period is added to the 95th percentile and subtracted from the 5th percentile to broaden the spread.

2.2. Performance metrics

Multiple metrics have been used in literature to assess the performance of numerical models with respect to available observations. Here, we adopt the Peak Relative Error (PRE) defined as

$$PRE = \frac{max(\eta_m) - max(\eta_o)}{max(\eta_o)} * 100\% \tag{5}$$

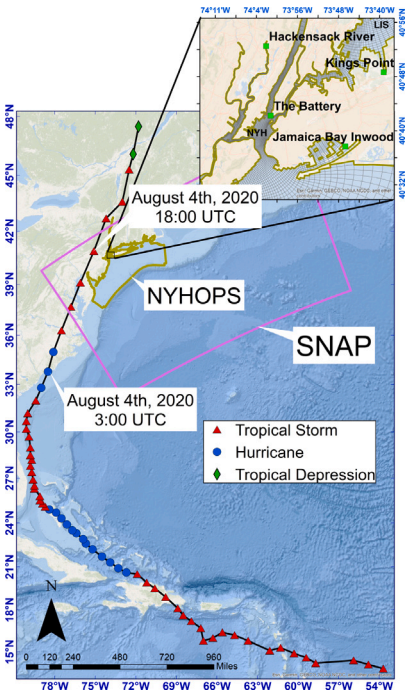


Fig. 1. Isaias track along the East Coast using three hours time period between consecutive readings. The locations of the four stations considered in this study are marked with green dots. All acronyms are defined in Section 2, except NYB is New York Bight, and LIS is Long Island Sound. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

A negative (positive) PRE means that the model is under- (over-) estimating the water level while zero PRE shows a perfect agreement between simulation results and observations. We also use the Coverage of Observation Uncertainties (COU) defined as

$$COU = \frac{n}{N} * 100\% \tag{6}$$

where n is the number of times the observed water levels fall within the specified confidence interval, and N is the total number of forecast time periods. Because we consider a confidence interval of 90%, between the 5th and 95th percentiles, the COU should ideally be equal to 90%.

3. Isaias track and observations

Isaias made two landfalls along its path with the first in the Bahamas on August 1st, 2020. As shown in Fig. 1, it then intensified to reach hurricane strength while moving parallel to Florida's coastline before making its second landfall in North Carolina at 3:00 (UTC) on August 4th. Afterwards, it weakened to a tropical storm as it tracked northward west of NYH along the Pennsylvania/New Jersey border on the afternoon of August 4th. As for its forward speed, it moved slowly before making landfall on the North Carolina coast, then it accelerated to a speed of 20 m s^{-1} as it passed by NYH. The rainfall gradually tapered off as it moved across New York and New England (Latto et al., 2021). The total rainfall in NYC was relatively low during the storm. For example, the total rain at Laguardia airport was about 1.35 cm. The performance assessment of SFAS, discussed in this paper, is based on comparing the ensemble-based predictions to observations at four coastal water level stations that are marked with green dots in the inset map of Fig. 1.

Fig. 2 shows time series of the wind vectors in panel *a* and of the pressure in panel *b* from NOAA's buoy 44 065 in New York Bight (NYB). Panel *a* also shows the wind vectors from the University of Connecticut's buoy at Execution Rocks (NOAA buoy code 44 022) in western Long Island Sound (LIS). The recorded maximum wind speeds

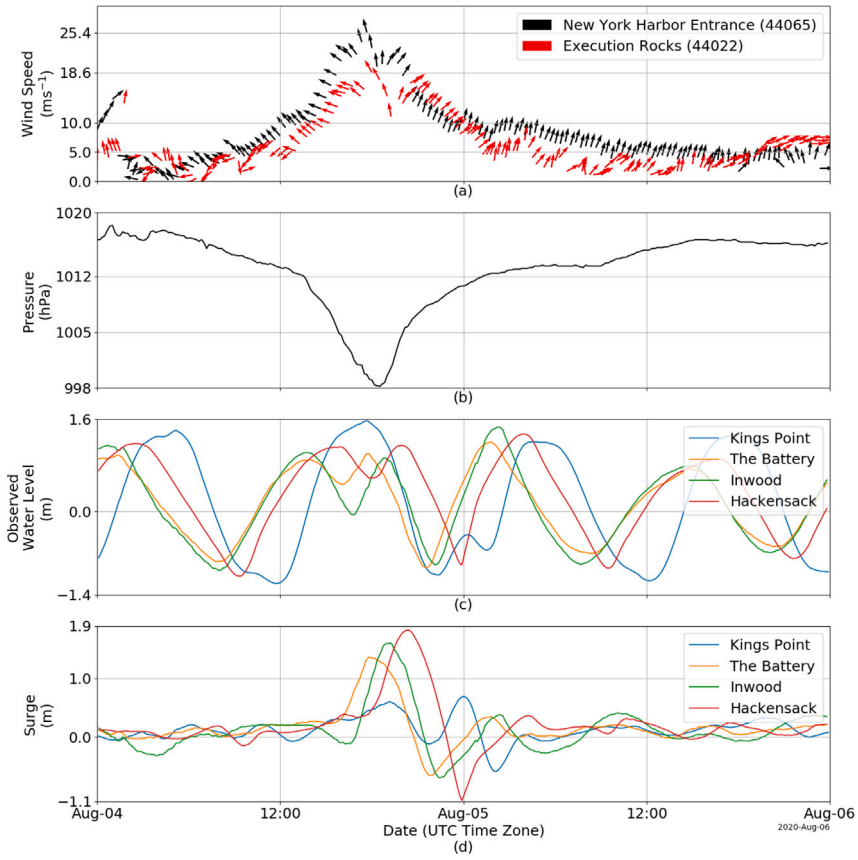


Fig. 2. Time series between August 4th and August 5th, 2020 of recorded (a) wind vectors showing the wind speed (m s^{-1}) and meteorological direction (clockwise from north) by NOAA's buoys offshore in NYB (44 065) and in LIS (44 022), (b) pressure (hPa) by NOAA's buoy in NYB, (c) water levels (m), and (d) surge (m) at the four stations.

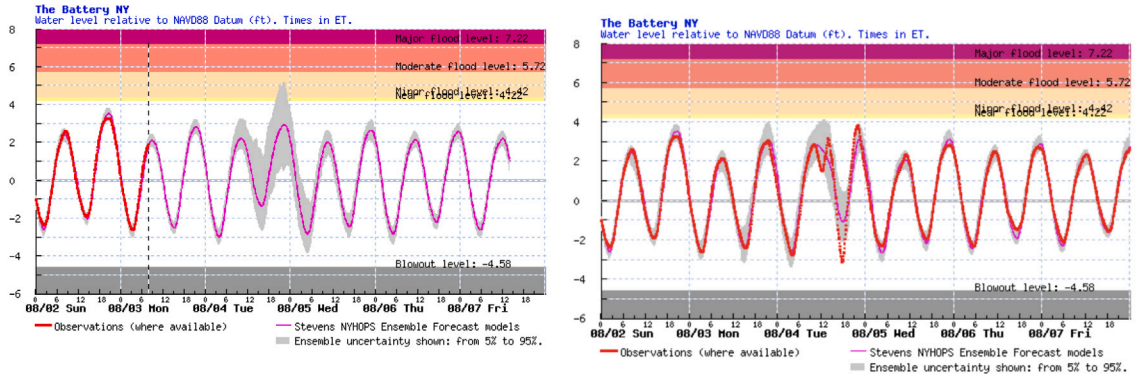
were respectively 25.4 and 18.6 m s^{-1} in NYB and LIS. Wind speeds above 10 m s^{-1} were maintained for about eight hours, which is a relatively short-duration event when compared with the most common surge-producing storms in this region (Orton et al., 2016). The wind speed, direction and pressure follow the movement of the hurricane. At 16:00 (UTC) on August 4th, the hurricane was located at lower latitudes with respect to NYB. The wind blew from the southeast (116 degrees) toward NYH and away from Long Island. The recorded pressure at NOAA's buoy offshore in NYB (44 065) was 1008 hPa and falling. Two hours later, at 18:00 (UTC), the hurricane passed by NYB and the wind blew with speeds of 22.8 and 18.6 m s^{-1} at NYB and LIS, respectively, from the southeast (169 degrees). The pressure offshore in NYB (buoy 44 065) reached a minimum of 998 hPa. At 21:00 (UTC), the recorded wind speeds at NYB and LIS were respectively 12.2 and 15.8 m s^{-1} and blew from the southwest (202 and 230 degrees). The pressure at NOAA's offshore buoy in NYB (44 065) had increased to 1008 hPa.

Fig. 2 also shows the observed water levels in panel *c* and surge in panel *d* at the four stations. Two distinct storm surges occurred, one the evening of August 4th (UTC) and the other around or soon after 00:00 (UTC) on August 5th. At the Battery, Jamaica Bay Inwood, and Hackensack stations, the observed water level plots exhibit a double peak on August 4th respectively at 17:50, 19:00, and 20:20 (UTC) with surge heights of 1.37, 1.61, and 1.83 m, respectively. These double peaks in the total water level occurred because the directly generated surge happened as the tide was retreating. At Kings Point, the directly generated surge occurred at the time of high tide on August 4th 17:30 (UTC) with a magnitude of 0.60 m, leading to a peak water level of 1.52 m. The directly generated surge at Kings Point was significantly less than the other three stations as the wind speed at Kings Point was relatively smaller than that at the other three stations and the wind was blowing southeast, i.e. across, not along, LIS. After three and a half hours, a negative surge (blowout) occurred with values ranging between −1.0 and −0.6 m at Battery, Jamaica Bay and Hackensack. Four hours later, a second surge, that we hereafter refer to as the resurgence, of 0.34 m occurred at the Battery and was slightly larger at the other two stations. The resurgence at the three stations coincided with the high tide which caused the highest observed total water level of about 1.3 m. Minor flooding occurred during the resurgence early on August 5th in Jamaica Bay and nearby areas, evidenced by an exceedance of National Weather Service "minor flood" datum at Inwood. On the other hand, at Kings Point, a larger resurgence occurred with magnitude of 0.69 m at the time of low tide, causing a second, but smaller, peak in total water level on August 5th 00:00 (UTC).

4. Performance analysis

4.1. SFAS online predictions

The online presentation of the SFAS forecast time series of water levels before and after the passage of Isaias by NYH included continuously updated time series such as the ones shown in Figs. 3(a) and 3(b). In Fig. 3(a), the time of the screen-capture is marked with the vertical dashed line, about August 3rd 8:00 EDT (12:00 UTC). The period to the left of August 3rd 2:00 EDT (6:00 UTC) is the hindcast, while the period to the right of 2:00 is the forecast. In both figures, the observation is marked by the red dots while the super-ensemble forecast is marked by the magenta line. The hindcast time series in Fig. 3(b) shows that the observations fall mostly within the forecast 90% confidence interval, marked by the gray shaded area. Although we will show below that individual ensemble members exhibited double peaks, the central forecast missed the double peak due to averaging.



(a) Forecast

(b) Hindcast

Fig. 3. SFAS online web displays of total water level forecast time series (a) prior to, and (b) after the passage of Isaias. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Ensemble time series

Fig. 4 shows the observed and simulated time series of the water levels from all ensemble members at the Battery station. The panels from top to bottom show the simulated water levels between August 2nd and August 5th. Each panel shows four and half forecasting days and one hindcast day with the initiation time in each figure set at 6:00 (UTC). As expected, the 96 meteorological forecasts yield a horizontal spread in the water levels indicating a time shift, and a vertical spread indicating variations in simulated water level values and peaks. The largest spreads occur between the afternoon of August 4th and the occurrence of the peak value right after 0:00 (UTC) on August 5th. A minority of ensemble members captured the peak water level (or higher), yet many predicted the resurgence albeit slightly later or earlier than the actual time of occurrence. Moreover, the double peak of the directly generated surge that occurred during the tidal retreat was captured by some of the ensemble members but not by the super-ensemble forecast due to averaging. Also important is the blowout, which was not captured by most of the ensemble members or super-ensemble forecast.

Fig. 5 shows plots of the COU over different forecast times at the four stations for sub-ensembles GEFS, CMC, and ECMWF-ENS and the super-ensemble consisting of all of them. The plots show that the COU range of the super-ensemble is mostly between 70% and 90%. A better overall assessment is based on the values in Table 1, which provides average COU values for the 90% confidence interval at the four stations and their mean values. Over time and across the four representative stations, the 90% confidence interval forecasted by the super-ensemble varied between 77 and 85% at the four stations, indicating reasonable estimates of uncertainty by the super-ensemble. In contrast, the forecasted 90% confidence intervals based on the sub-ensembles of the GEFS, ECMWF-ENS and CMC forecasts were respectively 67, 65, and 83%. On average, the CMC sub-ensemble yielded slightly better COU than the super-ensemble, while the super-ensemble outperformed the GEFS and ECMWF-ENS sub-ensembles. It is stressed that although one sub-ensemble may have yielded better predictions in terms of COU, considering all predictions supports using the super-ensemble. Finally, it is noted that, on average, the super-ensemble and CMC-based sub-ensemble were slightly under-dispersed, but worse under-dispersion would have occurred if only the GEFS and/or ECMWF-ENS sub-ensemble were considered. Furthermore, assessment of the entire year's storm events should provide more information about this assessment (Orton et al., 2021).

Comparison of SFAS results with P-Surge and P-ETSS is possible but these systems have 80% confidence interval. The average COU values of P-Surge model at the four stations are shown in Table 1. The P-Surge COU values ranged between 23% to 40% with a mean value of 30% compared with an ideal value of 80% for P-Surge. Also, NOAA ensemble water level forecast system P-ETSS, that runs only with NAEFS (GEFS and CMC) ensemble, showed $30-70\%$ COU values for 2018–2019 storms, compared with an ideal value of 80% for that system, which provides an 80% uncertainty range, revealing that uncertainty estimates with that system were always too small in storm events (Liu and Taylor, 2020).

Fig. 6 compares the maximum observed water levels at the four stations with the time series of maximum predicted water levels and 95th percentiles from the super- and sub-ensembles. The maximum water levels at the Battery, Jamaica Bay Inwood, and Hackensack occurred at the time of resurgence, while that at Kings Point occurred at the time of the directly generated surge. Based on these data, the super-ensemble PRE varied between -10 and -30% at Battery and Jamaica Bay, between -5 to -15% at Kings Point and between 5 to 15% at Hackensack. The positive PRE at Hackensack is likely due to the fact that the model is gridded only in waterways and not onto floodplains and, thus, only includes a converging channel in the Hackensack River, neglecting culverts and other hydraulic or hydrologic features that spread the floodwaters into the large surrounding area (The Meadowlands). At Kings Point, the range of the PRE is smaller than that at the other stations as Kings Point has a different basin and is located further away from the hurricane track. To compare, the PRE of the P-Surge central forecast (ensemble median) varied from -16% to -30% at Battery, -35% to -46% at Jamaica Bay, -25% to -46% at Hackensack, and -5% to -20% at Kings Point.

5. Discussion

The range and 95th percentile of the SFAS super-ensemble forecast provided ample warning of peak water levels during Isaias. Yet, the ensemble members and central forecasts were inaccurate on exact timing and amplitude of the peak (Figs. 3, and 6). Here, we look more closely at the uncommon storm surge, blowout and resurgence that occurred, and seek to explain the causes of these phenomena. We then assess likely sources of central forecast error. Lastly, we suggest some future improvements to be considered for mid-latitude ensemble flood forecast systems.

5.1. Directly generated surge, blowout and resurgence

One critical observation and topic for further analysis is how the peak water levels and flooding during Isaias were a consequence of a resurgence in storm surge. Isaias caused a directly generated surge, followed by a blowout and a resurgence. This sequence is not uncommon but, to the authors' knowledge, has never been reported to cause

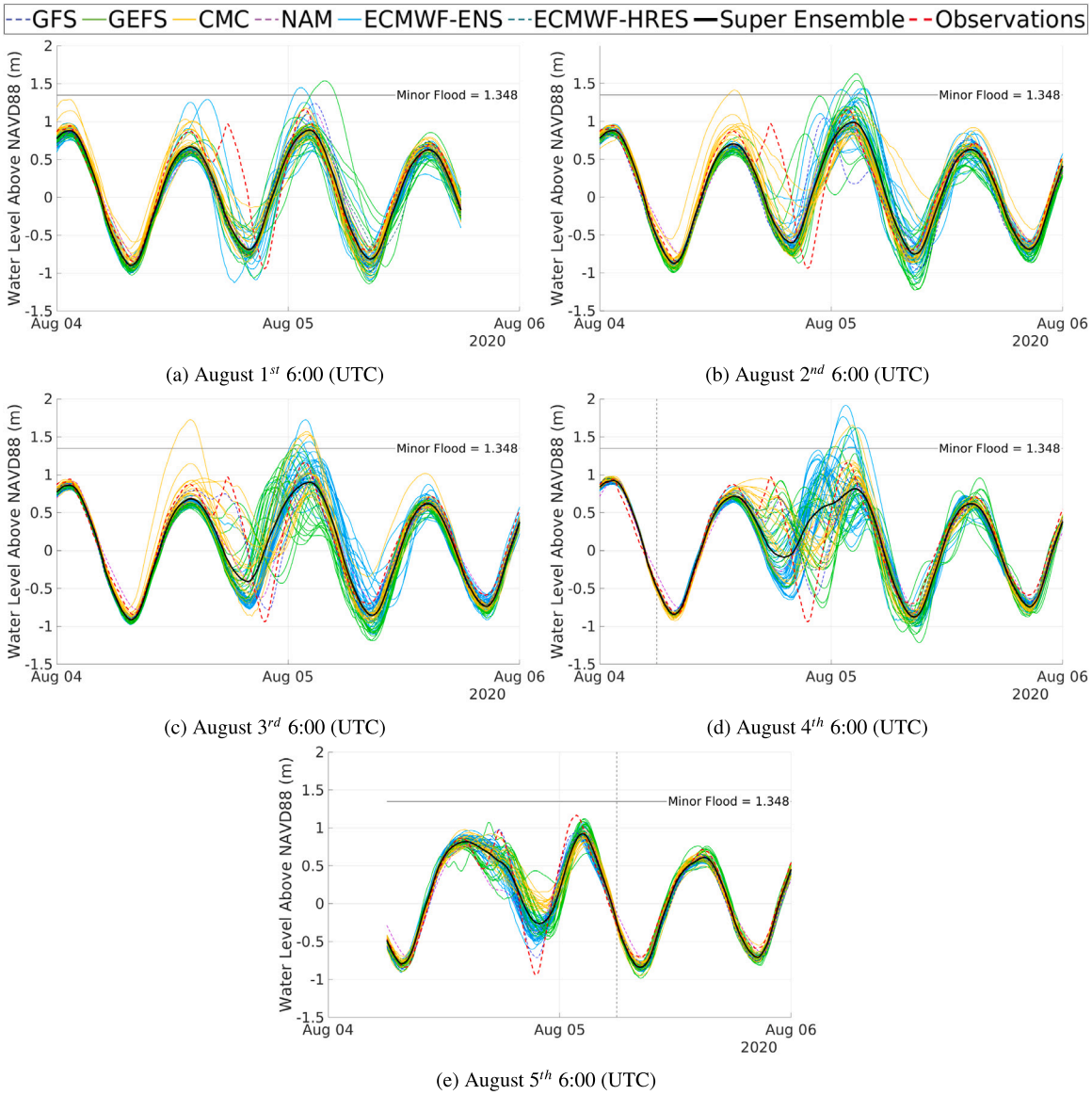


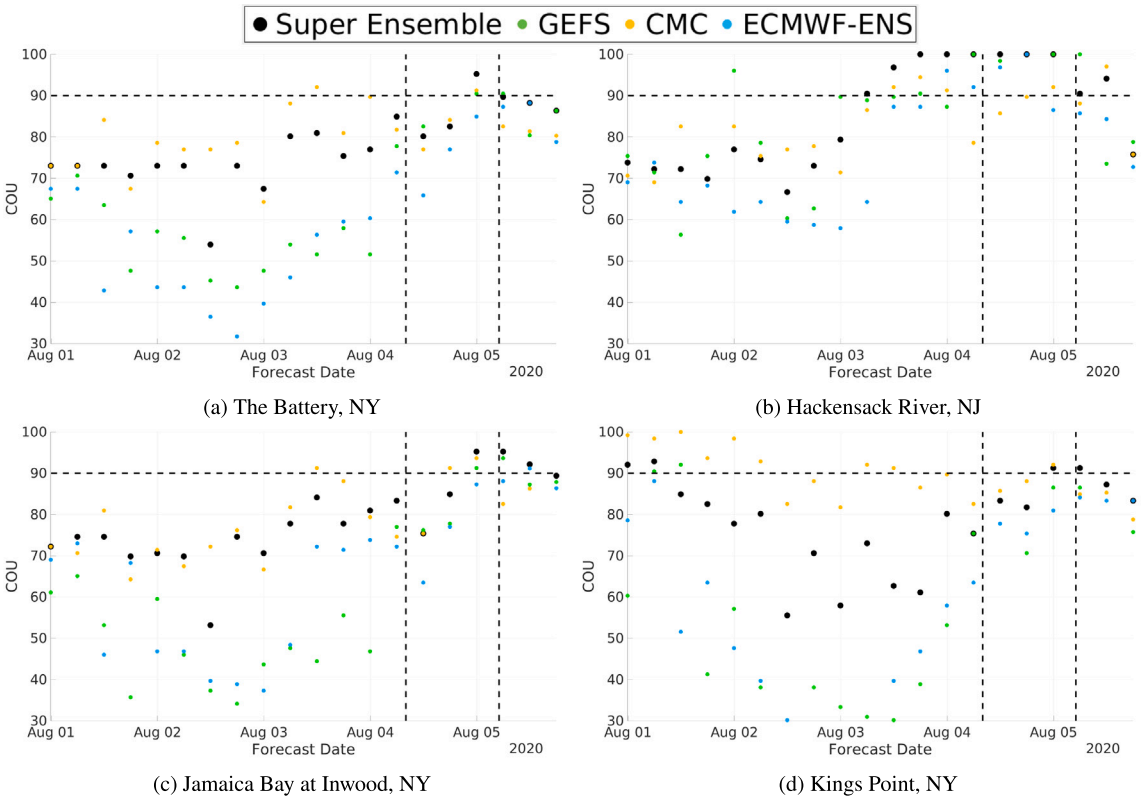
Fig. 4. The water level time series at the Battery station for all the ensemble members. Every plot shows the time series starting from a different date as shown in the captions. The day before these dates is always a hindcast while the following four and a half days are forecast.

Table 1
The average *COU* values for the mentioned confidence interval at the four stations and their mean values for SFAS and P-Surge (using advisories 21 to 29) forecasted water levels.

Ensemble name	Confidence interval (%)	The battery	Hackensack river	Jamaica Bay Inwood	Kings point	Mean COU (%)
Super-ensemble	90	77.5	85.3	78.3	78.3	79.9
GEFS	90	64.8	83.6	61.1	59.4	67.2
CMC	90	80.1	82.3	78.6	89.6	82.7
NAEFS (GEFS and CMC)	90	72.5	83	69.8	74.5	75
ECMWF-ENS	90	60.3	76.5	64.9	58.3	65
P-Surge	80	31.3	23.2	25.3	39.4	29.8

flooding or peak water levels for a storm. For instance, resurgences that occurred in NYH after the Long Island Express in 1938 and Gloria in 1985 did not cause flooding. The resurgences of these two storms at the Battery were respectively 58% and 14% of the peak surges with magnitudes of 0.55 and 0.29 m, respectively. Yet, resurgences can cause additional flooding or other water-speed related hazards (e.g. erosion), especially if they coincide with high tide or high water velocities.

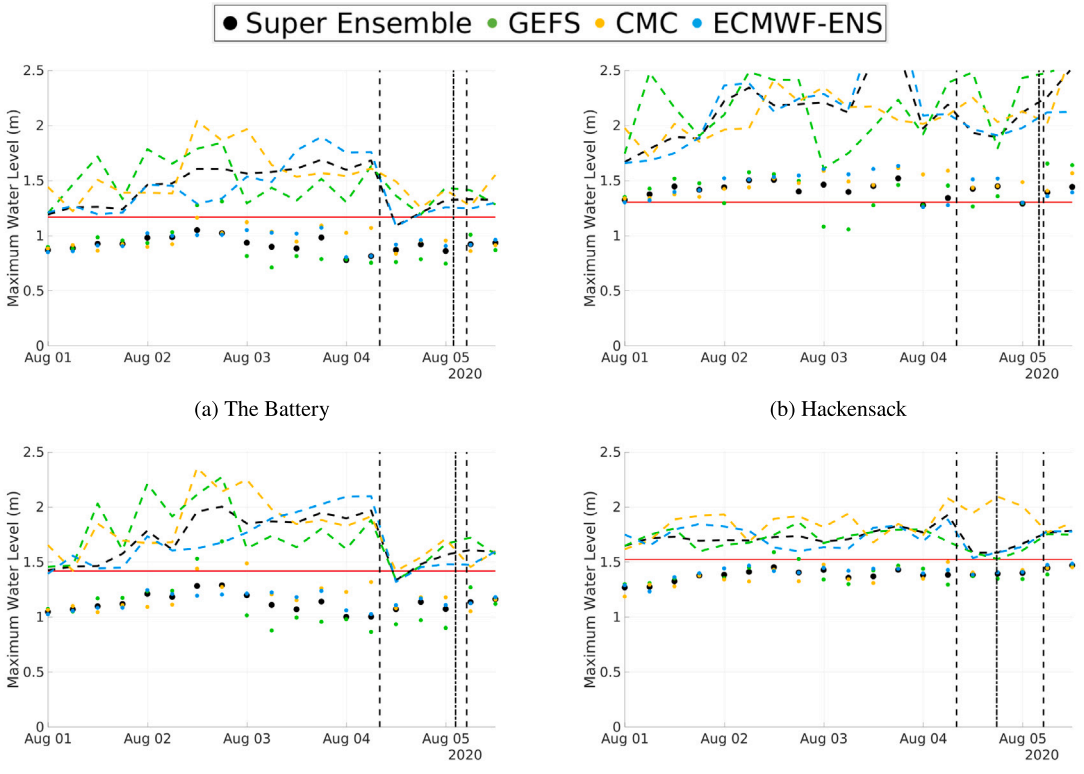
There are multiple possible mechanisms for a resurgence, including coastal trapped waves. For example, barotropic continental shelf waves and Kelvin waves can be formed when storms move across or along a mid-latitude continental shelf (e.g. Mysak and LeBlond (1978), Tang et al. (1998) and Thiebaut and Vennell (2010)). Sea level oscillations after Hurricane Sandy's passage were interpreted as being continental shelf waves, and observed to propagate from Long Island to the apex of New York Bight and down-coast to Atlantic City at 6.5 m s^{-1} (Chen et al., 2014). However, examination of tide gauge data at sites to the east along the Long Island coast and to the south along the New Jersey coast does not reveal a propagating wave traveling counter-clockwise around NY Bight. Therefore, we see no evidence of a coastal trapped wave in the case of Isaias.



(a) The Battery, NY (b) Hackensack River, NJ

(c) Jamaica Bay at Inwood, NY (d) Kings Point, NY

Fig. 5. The *COU* for the 90th percentile confidence interval for the four stations, assessed using observed data between 8AM on August 4th and 5AM on August 5th, represented by the two vertical dashed lines.



(a) The Battery (b) Hackensack

(c) Jamaica Bay Inwood (d) Kings Point

Fig. 6. Comparison of observed (red lines) and forecast storm-maximum water levels (points) and 95th percentiles (dashed curves). The storm period is bordered with Vertical dashed lines and the times of the observed maxima are marked with vertical dash–dot lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 7 shows the observed and modeled wind vector time series at buoy 44 065. We interpret the directly generated surge, blowout, and resurgence as a five-step process where storm surge at NYH has about an hour lag behind changes in NYB winds, roughly consistent with storm surge traveling as a shallow-water wave (e.g. Orton et al. (2012)):

1. **Wind-driven setup:** At buoy 44 065 (see Fig. 7 red vectors), the SSE wind blew with increasing intensity toward the Apex of NYB causing a rising wind-generated surge (setup) at NYH. SSE winds in NYB are conducive to storm surge in NYH (Lin et al., 2010; Gurumurthy et al., 2019).
2. **Relaxation of setup:** On August 4th, between 18:00 and 19:30 UTC, the wind stress forcing of wind setup at the apex of NY Bight ended as the winds turned from SSE to SSW, Fig. 7.
3. **Wind-driven setdown:** Consequently, a blowout (setdown) SSW wind forcing at 22 $m\,s^{-1}$ was initiated. These SSW winds have an offshore component along the NJ coast. SSW winds in NYB are conducive to blowout in NYH which is amplified due to inverse funneling (see below) Gurumurthy et al. (2019). SW winds along LIS (not shown) were also conducive to NYH blowout.
4. **Relaxation of setdown:** Wind at buoys 44 065 and 44 022 decreased below 15 $m\,s^{-1}$ at 20:30 UTC causing a modest relaxation of the blowout forcing.
5. **Subsequent pressure gradient forcing:** Water levels were particularly low in the harbor relative to offshore due to the inverse funneling, and as the blowout-producing winds relaxed, the sea level gradient led to the resurgence of water back into the harbor.

The sea-level see-saw response (surge-blowout-resurgence) is particularly important at NYH due to the concave coastline at the apex of NY Bight. The consecutive processes of surge and blowout described above were aided by the coastline shape, which can amplify both positive and negative surges through the coastal funneling effect (Gurumurthy et al., 2019; Mayo and Lin, 2022).

5.2. Sources of forecast error

Sources of central forecast error likely include storm forecast error (e.g. storm track, timing or intensity), inadequate meteorological product resolution, ocean model error, and ensemble averaging. The latter contributes to a tendency toward underestimating the peak water levels (e.g. the -10 to -30% PRE for Battery), as noted in Section 4.1 in reference to Fig. 4. Ensemble averaging is a challenge for capturing peaks in storm tide time series forecast but avoided in forecast systems that only provide statistics for water level temporal maxima (e.g. P-Surge). Quantifying any ocean model error would require more research, ideally using an extremely accurate meteorological reanalysis to minimize meteorological forcing errors.

Meteorological forcing provided two sources of error, through storm forecast timing and resolution bias. The wind vector time series (Fig. 7) show major timing errors that were present in many ensemble members' forecasts including that released early on August 4th (e.g. GEFS, NAM, ECMWF-ENS). These errors in timing generally resulted from inaccurate storm track forecast, with many weather model members wrongly forecasting that the TC would pass the region later than the actual time. This led to most members inaccurately predicting the timing of the storm surge as shown in Fig. 4. Also, resolution bias, arising from over-smoothing of the winds during a storm's passage, can lead to underestimation of storm surge (e.g., Jordi et al., 2019; Bloemendaal et al., 2019). This problem is increasingly obvious in Fig. 7 with the members organized in order (top to bottom) of finest to coarsest resolution. Members toward the bottom of the figure cannot capture the finer temporal variations of wind velocity in either forecast (August 4th) or hindcast (August 5th).

Resolution bias may also explain the discrepancies in forecasting the resurgence occurred as a response to the sequence of surge-then-blowout of water from the coast, as posited above. Due to the coarse meteorological forcing resolution, winds near the coast are a blurred interpolation between offshore (low surface drag) grid cells over water and inland grid cells (higher drag), whereas real-world winds likely have a more abrupt transition at or near the coastline. This can be seen in Fig. 7 where none of the meteorological products captures the same amplitude of variation in the east–west component of wind (from a SSE to SSW wind) that occurred from August 4th 15:00 to 21:00 (UTC) and some have no west wind component (e.g. CMC).

5.3. Advantages and challenges of mid-latitude flood forecast systems

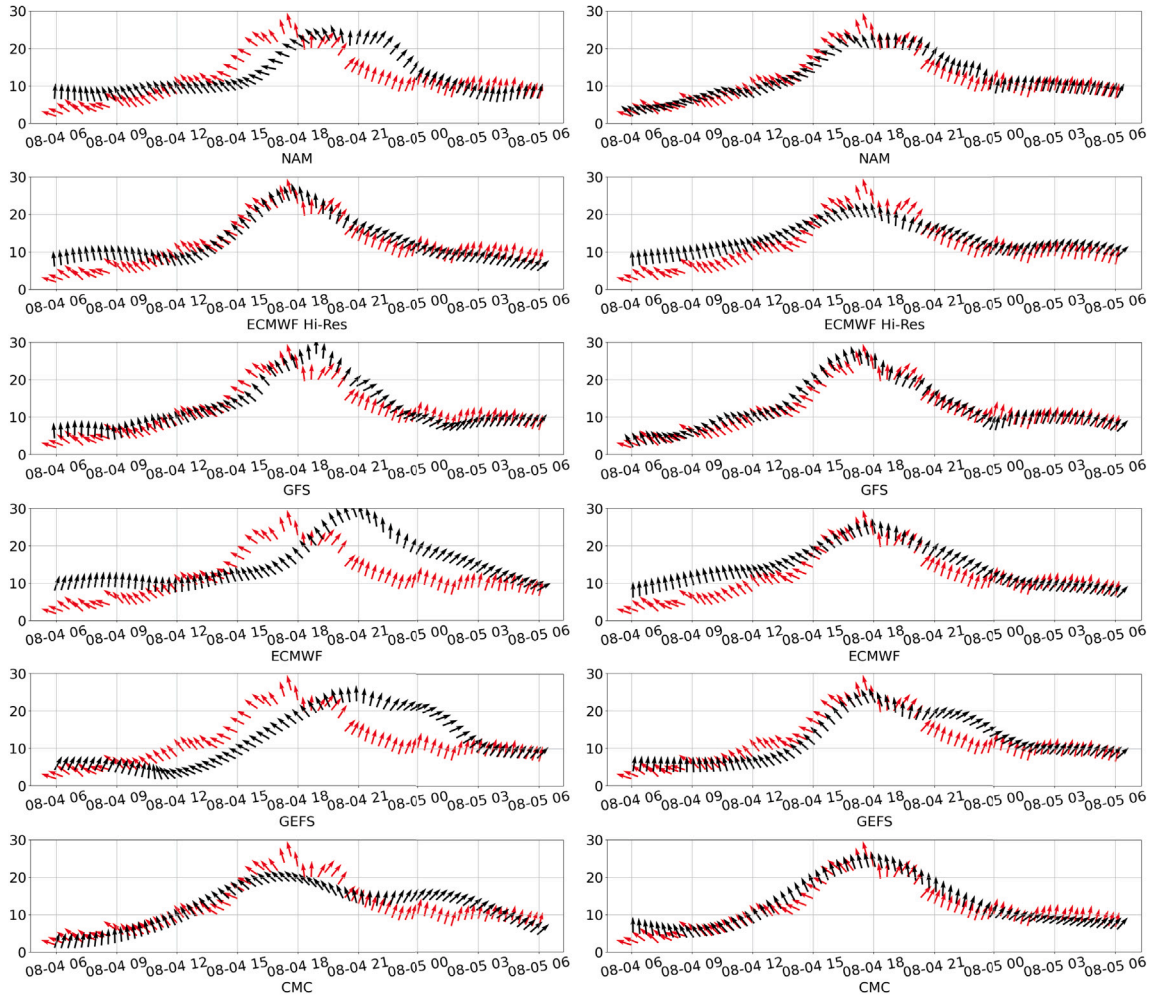
Mid-latitude ensemble coastal flood forecast systems like SFAS and P-ETSS typically utilize global or regional meteorological products and run year-round. As described in Section 1, these systems are different in many ways from National Hurricane Center P-Surge forecasts that produce forecasts only for TCs and use parametric TC meteorological forcing. Some advantages of the mid-latitude forecast systems include:

- Forecasting for ETC events, which are responsible for a majority of the top-5 ranked storm surge events across most of the US Mid-Atlantic and Northeast regions (Booth et al., 2016).
- Year-round consistent flood forecast products for events ranging from tidal floods to extreme events.
- TCs in mid-latitudes often become large and elongated in shape as they undergo transition to ETC status (Hart and Evans, 2001; Colle, 2003), leading to ambiguity in which meteorological forcing will be most accurate (realistic gridded or parametric circular TC (Taylor and Glahn, 2008; Gonzalez and Taylor, 2021)).

Mid-latitude regions generally have smaller and less frequent tropical cyclone-driven storm surge events than the lower latitudes, and it is of greater value to incorporate processes other than storm surge in coastal flood forecasting. The three-dimensional SFAS modeling integrates tides, river flows, and storm surge, and therefore leads to a more accurate forecast of water levels (Orton et al., 2012). For example, SFAS' predicted water level shows better COU, and PRE values more than that of the tropical cyclone P-Surge forecast system model for this event. However, a comprehensive comparison would require a large number of TC forecasts, which were not available at the time of publication. P-Surge only models storm surge and sums it with deterministic tide prediction (Taylor and Glahn, 2008; Gonzalez and Taylor, 2021). In areas with large tides relative to storm surge, such as NYB, LIS and the Gulf of Maine, this simplification leads to neglect of tide–surge interaction and associated forecast uncertainty. Moreover, tropical cyclone forecast models (e.g., P-ETSS, P-Surge) continue to be limited to two-dimensional modeling without stream flows nor dynamic combined modeling of tides and surges. Also, forecast systems that poorly resolve river systems (e.g. Hackensack) or neglect strong tide–surge interactions (e.g. Jamaica Bay, Kings Point) will perform more poorly on tidal or streamflow-driven events than SFAS, which explicitly accounts for these processes (Orton et al., 2012). Yet, TC forecast systems have the advantage for fast or small storms of having extremely high resolution wind and pressure forcing, whereas mid-latitude systems are limited to the available meteorological ensemble forecast products that have coarser resolution.

6. Conclusions

The SFAS ensemble prediction system for coastal total water level in the New York metropolitan area has been operating continuously for six years with a growing user base including National Weather Service Weather Forecast Offices. Evaluation of its forecasts is conducted on a continuous basis with an eye toward improving its performance. Tropical Cyclone Isaias was one of the most challenging water level



(a) Forecast (August 4th 06Z) (b) Hindcast (August 5th 06Z)

Fig. 7. Wind vectors for buoy 44 065 in New York Bight, with observations in red and forecasts in black. The wind speed is in m s^{-1}. The weather models are sorted with the finest resolution at the top panels to the coarsest at the bottom panels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

forecasts in recent years. It caused a rapidly-developing and undulating surge consisting of a positive surge, a blowout with negative surge, and a resurgence that resulted in peak water levels and flooding in certain areas of New York Harbor. In this study, we showed that SFAS outperformed the National Hurricane's Center forecast with respect to spread and accuracy. However, a comprehensive comparison should be conducted with other TC forecasts, which were not available at the time of publication.

Our evaluation of the accuracy and ensemble spread of SFAS super- and sub-ensemble forecasts shows that the range and 95th percentile of the super-ensemble forecast provided ample warning of peak water levels. However, while some ensemble members predicted the resurgence peak and its timing satisfactorily, the super-ensemble forecasts were not as accurate on the exact timing and amplitude of the peak. This arose because of inaccurate timing of the storm forecasts, ensemble averaging, and coarse-resolution atmospheric forcing. Considering wind and tide data, we determined that the resurgence traveled as a shallow-water wave rather than a coastal trapped wave. The presented analysis suggests that more accurate and higher-resolution meteorological forcing can help reduce biases in forecasting storm surge, especially those associated with tropical cyclones and challenging sequences of events as in the case of Isaias.

CRediT authorship contribution statement

Mahmoud Ayyad: Writing – original draft, Software, Methodology, Validation, Investigation, Visualization. **Philip M. Orton:** Writing – review & editing, Methodology, Validation, Investigation, Visualization. **Hoda El Safty:** Writing – review & editing, Methodology, Investigation. **Ziyu Chen:** Software. **Muhammad R. Hajj:** Writing – review & editing, Investigation, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

PMO was funded for this research by the National Oceanic and Atmospheric Administration Regional Integrated Sciences and Assessments (RISA) award NA21OAR4310313.

References

Ayyad, M., Hajj, M.R., Marsooli, R., 2021. Spatial variation in sensitivity of hurricane surge characteristics to hurricane parameters. J. Eng. Mech. 147 (10), 04021070.

Bloemendaal, N., Muis, S., Haarsma, R.J., Verlaan, M., Irazoqui Apecechea, M., de Moel, H., Ward, P.J., Aerts, J.C., 2019. Global modeling of tropical cyclone storm surges using high-resolution forecasts. Clim. Dynam. 52 (7), 5031–5044.

Booth, J.F., Rieder, H., Kushnir, Y., 2016. Comparing hurricane and extratropical storm surge for the Mid-Atlantic and Northeast Coast of the United States for 1979–2013. Environ. Res. Lett. 11 (9), 094004.

Chen, N., Han, G., Yang, J., Chen, D., 2014. Hurricane Sandy storm surges observed by HY-2A satellite altimetry and tide gauges. J. Geophys. Res. Oceans 119 (7), 4542–4548.

Cho, D., Yoo, C., Son, B., Im, J., Yoon, D., Cha, D.-H., 2022. A novel ensemble learning for post-processing of NWP Model's next-day maximum air temperature forecast in summer using deep learning and statistical approaches. Weather Clim. Extremes 100410.

Colle, B.A., 2003. Numerical simulations of the extratropical transition of Floyd (1999): Structural evolution and responsible mechanisms for the heavy rainfall over the northeast United States. Mon. Weather Rev. 131 (12), 2905–2926.

Colle, B.A., Bowman, M.J., Roberts, K.J., Bowman, M.H., Flagg, C.N., Kuang, J., Weng, Y., Munsell, E.B., Zhang, F., 2015. Exploring water level sensitivity for metropolitan New York during Sandy (2012) using ensemble storm surge simulations. J. Mar. Sci. Eng. 3 (2), 428–443.

Florescu, I., 2014. Probability and Stochastic Processes. John Wiley & Sons.

Florescu, I., Tudor, C.A., 2013. Handbook of Probability. John Wiley & Sons.

Flowerdew, J., Horsburgh, K., Wilson, C., Mylne, K., 2010. Development and evaluation of an ensemble forecasting system for coastal storm surges. Q. J. R. Meteorol. Soc. 136 (651), 1444–1456.

Flowerdew, J., Mylne, K., Jones, C., Titley, H., 2013. Extending the forecast range of the UK storm surge ensemble. Q. J. R. Meteorol. Soc. 139 (670), 184–197.

Forbes, C., Rhome, J., Mattocks, C., Taylor, A., 2014. Predicting the storm surge threat of Hurricane Sandy with the National Weather Service SLOSH model. J. Mar. Sci. Eng. 2 (2), 437–476.

Georgas, N., Blumberg, A.F., 2010. Establishing confidence in marine forecast systems: The design and skill assessment of the New York Harbor Observation and Prediction System, version 3 (NYHOPS v3). In: Estuarine and Coastal Modeling (2009). pp. 660–685.

Georgas, N., Blumberg, A., Herrington, T., Wakeman, T., Saleh, F., Runnels, D., Jordi, A., Ying, K., Yin, L., Ramaswamy, V., et al., 2016. The stevens flood advisory system: Operational H3E flood forecasts for the greater New York/New Jersey metropolitan region. Flood Risk Manag. Response 194.

Georgas, N., Orton, P., Blumberg, A., Cohen, L., Zarrilli, D., Yin, L., 2014. The impact of tidal phase on Hurricane Sandy's flooding around New York City and Long Island Sound. J. Extreme Events 1 (01), 1450006.

Gneiting, T., Raftery, A.E., Westveld III, A.H., Goldman, T., 2005. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Mon. Weather Rev. 133 (5), 1098–1118.

Gonzalez, T., Taylor, A., 2021. Development of the NWS'probabilistic tropical storm surge model. In: Proceedings of the 33rd Conference on Hurricanes and Tropical Meteorology, Ponte Vedra, FL, USA, Vol. 11.

Gurumurthy, P., Orton, P.M., Talke, S.A., Georgas, N., Booth, J.F., 2019. Mechanics and historical evolution of sea level blowouts in New York Harbor. J. Mar. Sci. Eng. 7 (5), 160.

Hart, R.E., Evans, J.L., 2001. A climatology of the extratropical transition of Atlantic tropical cyclones. J. Clim. 14 (4), 546–564.

Jelesnianski, C.P., 1992. SLOSH: Sea, Lake, and Overland Surges from Hurricanes, Vol. 48. US Department of Commerce, National Oceanic and Atmospheric Administration ....

Jordi, A., Georgas, N., Blumberg, A., Yin, L., Chen, Z., Wang, Y., Schulte, J., Ramaswamy, V., Runnels, D., Saleh, F., 2019. A next-generation coastal ocean operational system: Probabilistic flood forecasting at street scale. Bull. Am. Meteorol. Soc. 100 (1), 41–54.

Kim, S.-C., Chen, J., Shaffer, W., 1996. An operational forecast model for extratropical storm surges along the US east coast. In: Preprints, Conf. on Oceanic and Atmospheric Prediction, Atlanta, GA, Amer. Meteor. Soc. pp. 281–286.

Latto, A., Hagen, A., Berg, R., 2021. Tropical Cyclone Report: Hurricane Isaias (AL092020). Technical Report, National Hurricane Center.

Lin, N., Emanuel, K.A., Smith, J.A., Vanmarcke, E., 2010. Risk assessment of hurricane storm surge for New York City. J. Geophys. Res.: Atmos. 115 (D18).

Liu, H., Taylor, A., 2020. 4.2 Latest development in the NWS'probabilistic extra-tropical storm surge model.

Liu, H., Taylor, A., Kang, K., 2019. 3.8 Latest development in the NWS'extra-tropical storm surge model, and probabilistic extra-tropical storm surge model. In: 99th American Meteorological Society Annual Meeting. AMS.

Liu, H., Taylor, A., Spring, M., 2018. Development of the NWS'probabilistic extra tropical storm surge model and post-processing methodology. In: Proceedings of the 98th AMS Annual Meeting, Austin, TX, USA. pp. 7–11.

Mayo, T.L., Lin, N., 2022. Climate change impacts to the coastal flood hazard in the northeastern United States. Weather Clim. Extremes 100453.

Molteni, F., Buizza, R., Palmer, T.N., Petroliagis, T., 1996. The ECMWF ensemble prediction system: Methodology and validation. Q. J. R. Meteorol. Soc. 122 (529), 73–119.

Mysak, L.A., LeBlond, P.H., 1978. Waves in the Ocean. Elsevier.

Orton, P., Chen, Z., El Safty, H., Ayyad, M., Datla, R., Miller, J., Hajj, M.R., 2021. Stevens Flood Advisory System 2020 Ensemble Forecast Assessment: NY/NJ Harbor Area. Technical Report, URL: https://hudson.dl.stevens-tech.edu/sfas/reports/SFAS_Forecast2020_publicreport.pdf.

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

Orton, P.M., Hall, T., Talke, S.A., Blumberg, A.F., Georgas, N., Vinogradov, S., 2016. A validated tropical-extratropical flood hazard assessment for New York Harbor. J. Geophys. Res. Oceans 121 (12), 8904–8929.

Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. Mon. Weather Rev. 133 (5), 1155–1174.

Saleh, F., Ramaswamy, V., Georgas, N., Blumberg, A.F., Pullen, J., 2016. A retrospective streamflow ensemble forecast for an extreme hydrologic event: a case study of Hurricane Irene and on the Hudson River basin. Hydrol. Earth Syst. Sci. 20 (7), 2649–2667.

Szpilka, C., Dresback, K., Kolar, R., Feyen, J., Wang, J., 2016. Improvements for the western north atlantic, caribbean and gulf of mexico adcirc tidal database (EC2015). J. Mar. Sci. Eng. 4 (4), 72.

Tang, C., Gui, Q., DeTracey, B., 1998. Barotropic response of the Labrador/Newfoundland shelf to a moving storm. J. Phys. Oceanogr. 28 (6), 1152–1172.

Taylor, A.A., Glahn, B., 2008. Probabilistic guidance for hurricane storm surge. In: 19th Conference on Probability and Statistics, Vol. 74.

Thiebaut, S., Vennell, R., 2010. Observation of a fast continental shelf wave generated by a storm impacting Newfoundland using wavelet and cross-wavelet analyses. J. Phys. Oceanogr. 40 (2), 417–428.

de Vries, H., 2009. Probability forecasts for water levels at the coast of the Netherlands. Mar. Geod. 32 (2), 100–107.

Whan, K., Zscheischler, J., Jordan, A.I., Ziegel, J.F., 2021. Novel multivariate quantile mapping methods for ensemble post-processing of medium-range forecasts. Weather Clim. Extremes 32, 100310.