Evaluating the accuracy and uncertainty of atmospheric and wave model hindcasts during severe events using model ensembles

Ali Abdolali[1,2] ⬤ · Andre van der Westhuysen[1,2] · Zaizhong Ma[1,2] · Avichal Mehra[1] · Aron Roland[3] · Saeed Moghimi[4,5]

Abstract

Various uncertainties exist in a hindcast due to the inabilities of numerical models to resolve all the complicated atmosphere-sea interactions, and the lack of certain ground truth observations. Here, a comprehensive analysis of an atmospheric model performance in hindcast mode (Hurricane Weather and Research Forecasting model—HWRF) and its 40 ensembles during severe events is conducted, evaluating the model accuracy and uncertainty for hurricane track parameters, and wind speed collected along satellite altimeter tracks and at stationary source point observations. Subsequently, the downstream spectral wave model WAVEWATCH III is forced by two sets of wind field data, each includes 40 members. The first ones are randomly extracted from original HWRF simulations and the second ones are based on spread of best track parameters. The atmospheric model spread and wave model error along satellite altimeters tracks and at stationary source point observations are estimated. The study on Hurricane Irma reveals that wind and wave observations during this extreme event are within ensemble spreads. While both Models have wide spreads over areas with landmass, maximum uncertainty in the atmospheric model is at hurricane eye in contrast to the wave model.

1 Introduction

The frequency and destructiveness of coastal storms have required improving the accuracy of numerical weather

---

✉ Ali Abdolali
ali.abdolali@noaa.gov

1 NWS/NCEP/Environmental Modeling Center, National Oceanic and Atmospheric Administration (NOAA), College Park, MD, USA

2 I.M. Systems Group, Inc. (IMSG), Rockville, MD, 20852, USA

3 BGS IT & E, Darmstadt, Hesse, Germany

4 NOAA Coast Survey Development Laboratory, National Ocean Service, Silver Spring, MD, USA

5 University Corporation for Atmospheric Research (UCAR), Boulder, CO, USA

prediction models, incorporating three other major strategies: (1) data assimilation, (2) ensemble modeling, and (3) atmospheric-wave-surge-hydrological coupling. The coupling of atmospheric, ocean wave, surge, and hydrological models on high-resolution numerical grids has improved model accuracy by better representing nearshore/inland geometries and physics (Moghimi et al. 2020). Coupling reflects the dynamic feedbacks of model components, and improves our understanding of such a complicated system. In addition, High Performance Computing (HPC) facilitates the computational speed of the aforementioned modeling systems. Multiple sources of error still remain, from instrument and processing noise in the observational data that are used to develop the models, to the physical parameterizations that account for the unresolved physics, resolution limits, and physics simplification in the models, to the stochasticity of the natural processes themselves. Therefore, determining the damage caused by hurricanes using such numerical models requires a statistical evaluation of uncertainty.

In this study, we statistically evaluate the outputs of hindcasted deterministic and ensembles simulations from the Hurricane Weather and Research Forecasting

model hereinafter HWRF (Tallapragada et al. 2015) and WAVEWATCH III, hereinafter WW3 (WW3DG 2019), an ocean wave model forced here with HWRF 10-m surface winds. Conventionally, the performance of atmospheric models is evaluated for hurricane track parameters, available from best track parameters' tables, including but not limited to hurricane intensity, central coordinate, radius of maximum wind, central and background pressure, radii for 34, 50, and 64 knots (1 kn = 0.514444 m/s) thresholds on four quadrants (Sampson et al. 2017; Bender et al. 2017). In addition, time series of observations at fixed in situ locations such as meteorological stations, wave buoys, tide and stream gauges, and spatio-temporal along-track satellite data are used to assess the accuracy of atmospheric, wave, and surge models. However, these data are sparse, often not covering the area of interest where the damage needs to be determined and sometimes unavailable within landfall time window. To fill this gap, we first calculate the spread of ensemble model results as an estimate of our model uncertainty, and hence as a measure of the model's accuracy over the entire model domain, notably in regions away from the available observations. Then, model accuracy is evaluated against available observations in term of statistical parameters, and a paired t test with hypotheses represented in terms of p value for success or failure of model in meeting observation. The case study is Hurricane Irma, 2017, the most powerful hurricane on record in the open Atlantic region outside of the Caribbean Sea and Gulf of Mexico, until it was surpassed by Hurricane Dorian just 2 years later. The HWRF model's uncertainty is determined via analysis of 40 ensemble members, corresponding to 40 sets of initial conditions of driving variables every 6 h (Fig. 1). The high number of ensemble members allows to capture the spread of HWRF prediction errors, ensuring that the hydrodynamic models are forced with a wide enough ensemble. First, 40 members are resampled from the outputs of HWRF members randomly. Secondly and from the spread of HWRF-derived best track paramerers, relative to the National Hurricane Center (NHC) advisory (Tong et al. 2018), 40 additional ensemble members for the wind field are generated to force the downstream wave model. We analyze the statistical distributions of time series of model errors, for particular locations and important hurricane variables. We provide an exploratory method to assess the similarity between observations and HWRF/WW3 model estimates which is general enough to be useful across many geophysical variables. This is particularly important for minimizing the error propagation required by complicated coupled model systems.
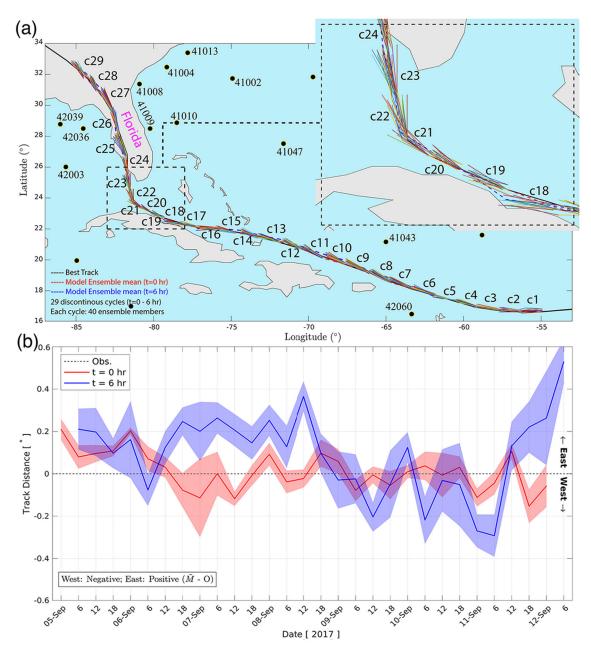
The spread of model outputs during extreme conditions of Hurricane Irma, with multiple landfall locations, allows us to evaluate the accuracy of each individual model in detail. The validation results for the investigated case show that the ensemble mean has a low bias, and that the observations fall within the ensemble spread, suggesting that it is broad enough in the context of hindcasting.

This paper is arranged as follows. A summary of potential sources of error in numerical models and observations are presented in Section 2. Section 3 provides a brief overview of the case study, Hurricane Irma and the observations for model verification (satellite and point source observations). Section 4 describes the atmospheric model, track analysis, and ensemble system. A description of WW3 and forcing ensembles is given in Section 5. Wind and wave results, extracted from the deterministic run and ensemble members, are discussed in Section 6. Description of the paired t test for time series analysis is given in Section 7. Concluding remarks are provided in Section 8.

2 Model and observation sources of error

Numerical models' inability to resolve natural processes can be either from inaccurate numerics/physics or due to the spatial and temporal limits (grid resolution and time step), where subgrid processes and short-scale changes are not considered. In addition, a numerical weather prediction (NWP) model like HWRF is often optimized to resolve dominant physics, simplifying the governing equations with bulk formulae for capturing heat and momentum transfer due to air-sea flux exchanges. Furthermore, certain physical processes (like sea spray) are often neglected because of lack of sufficient evidence to support their inclusion. Use of observations for assimilation (in HWRF) also leads to "representation error" whereby unresolved physical processes impact the observations but not the model. These errors are due to discretization errors on the coarse model grids and can also be state dependent and correlated in time (Desroziers et al. 2001; Janjic and Cohn 2006).

In principle, a phase-averaged wave model does not treat waves individually but instead uses the wave spectrum as the prognostic variable by describing the evolution of the wave density spectrum (with source and sink terms as the source of gain or loss). In other words, such a model is a deterministic description of statistical properties of the sea surface, mostly the dominant ones. It has directional and spectral resolution limits in addition to spatial and temporal resolution limits. In addition, the subgrid processes and short-scale changes are not represented in a phase-averaging

Fig. 1 Discontinuous Hurricane Irma tracks, extracted from the first 6 hrs of HWRF cycles for 40 ensemble members. The continuous best track is shown by solid black line while the mean of HWRF ensembles at $t = 0$ and $t = 6$ h is shown by dashed red and blue lines respectively. NDBC buoys equipped with meteorological and directional wave sensors are shown in entire domain. The spread of ensembles in terms of standard deviation (σ) is shown in panel (b). Seven landfalls on Barbuda (1: September 6, 05:45), St. Martin (2: September 6, 11:15), Virgin Gorda, British Virgin Islands (3: September 6, 16:30), Little Inagua, Bahamas (4: September 8, 05:00), near Cayo Romano, Cuba (5: September 9, 03:00), Cudjoe Key, Florida (6: September 10, 13:00), and near Marco Island, Florida (7: September 10, 19:30) are marked in panel (b). The gray area shows the time after final landfall

model. Within the phase averaged assumption, a prognostic tail takes into account the higher-order moments and high frequency part of spectrum with limits in model parameterizations or missing physical processes within that range.

Despite the significant progress in spectral wave models, these models lack skills in the surf zone and transition from deep water to intermediate and shallow depths where the gradients are larger and nonlinear processes often dominate. It gets more complicated in coastal zone where air-sea fluxes, wave breaking, coastal currents, reflection, refraction, wave-current interactions, fluidization and transport of sediments, bottom friction and scattering, wave-vegetation interaction, and nonlinearity become dominant (Cavaleri

2006). Most of these processes are often dealt with in an empirical way, particularly under the spectral approach, which brings uncertainty to the model. Besides, as these models are being pushed into shallower waters, the currents and water levels (tide and surges) cannot be considered independently of waves and can be identified as another source of error. Thus, when comparing the model results (in term of Representative variables of sea-state, i.e., H_s and T_p) with the observation, some confidence limits should be considered (Monbaliu 2003; WW3DG 2019; Roland and Ardhuin 2014). Decoupling of different shallow water processes, where hydrodynamics and ocean waves are modeled stand-alone or in a one-way fashion, introduces error in the models' outputs. Note that the necessity of dynamic coupling between atmospheric, wave, surge, and hydrological model is recognized; therefore, a tremendous effort across modeling groups is in place (Moghimi et al. 2020; Bakhtyar et al. 2020), leading to improvement in the accuracy of each individual model in the coupled system. On the other hand, the observations including stationary observations, along satellite tracks or radar field snapshots carry errors and uncertainties, mostly due to device accuracy, calibration, and post-processing algorithm. A few number of these variables are directly observed (i.e., wind speed at the National Data Buoy Center observatories—NDBC) and mostly are indirectly calculated from observed variables (i.e., H_s from satellite altimetry using the slope of the leading edge of the returned wave form and calibration coefficients or H_s at NDBC from spectral density data, translated from the buoy accelerations and integrate those to heave, pitch, and roll motions). It implies the fact that the observations are not 100% precise and the error embedded in such data should be considered during performance evaluation of models.

Here, our focus is on the estimation of the error generated by the atmospheric model and its ensemble members, propagated into the wave model. The source of error can be due to the aforementioned parameters of the models' inabilities and observation uncertainties.

3 Case study: Hurricane Irma, 2017

On August 30th, a Cape Verde hurricane named Irma was generated on (29.6°W, 16.1°N), swept westward over the Atlantic and reached category 5 intensity (Saffir-Simpson Hurricane Wind Scale) with four out of seven category five hurricane landfalls across the northern Caribbean Islands. Although the system weakened after the landfall in Cuba to category 2, it re-strengthened to category 4 status as it crossed the Straits of Florida, made landfall on Cudjoe Key on September 10 and later that day in Florida on Marco Island as category 3 (Cangialosi et al. 2018). The Hurricane

Irma best track with time tag (30 August–12 September 2017) is shown in Figs. 1 and 2. Irma was the ninth named storm, fourth hurricane, second major hurricane, and first category 5 hurricane of the 2017 season in Atlantic basin. Its wind speed and pressure reached $\sim$ 285 km/h and 914 mb on September 6th. In the Caribbean, the maximum observed waves reached 8 m in Cayo Romano where hurricane was category 5. The sea level in Ciego de Ávila Province rose by 3 to 3.5 m and penetrated inland more than 800 m. In Florida Keys and Southwestern Florida, the combined effect of storm surge and the tide produced maximum inundation levels of 2.5 and 3 m above ground level respectively. The NDBC network captured maximum observed waves on both sides of the Florida Peninsula with a significant height of $\sim$ 6 m at NDBC #41008 and #42036 on September 11. The peak period of observed waves was about 15 and 10 s at offshore and nearshore NDBC observations respectively. Irma was directly responsible for 52 deaths and indirectly responsible for a further 82 fatalities, with damage of 77.16 billion (2017 USD).
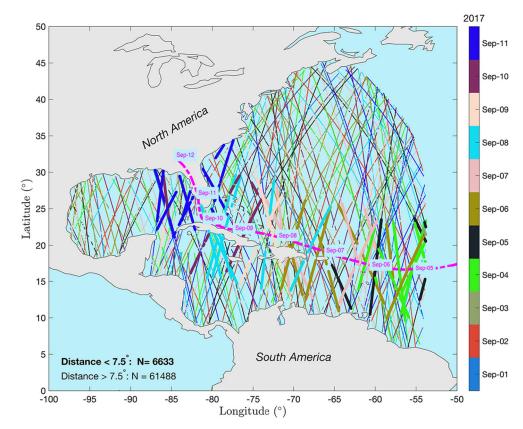
3.1 Observations

The accuracy of the atmospheric and wave models is quantified along spatiotemporal satellite observations and at stationary observations in term of time series. Besides the uncertainties, embedded in the results of numerical models, errors and uncertainties exist in in situ observations. The source of measurement errors can be due to either the instrument accuracy, calibration error, or data post-processing algorithms where the compared variables are not collected directly (i.e., significant wave height). Therefore, comparison of model results against buoy measurements and satellite altimeter data are not conclusive solely if the uncertainties in the observations are not determined. The accuracy of observations is estimated based on other independent data like hindcast model outputs and in situ measurements. For example, Abdalla et al. (2011) estimated the uncertainty in the Jason 1&2 and Envisat RA-2 satellite observation and buoy measurements using triple collocation technique for the length of 1 year and reported significant wave height absolute error of 0.13–0.19 m or Scatter Index within 5.4–7.8% range relative to the mean values for satellite and absolute error of 0.206–0.218 m or Scatter Index within 8.6–8.9% range relative to the mean values for buoy observations. For wind speed, absolute error of $\sim$ 1 m/s or 12% Scatter Index relative to the mean values for satellite and absolute error of $\sim$ 1.15 m/s or Scatter Index of 13.8 % relative to the mean values for buoy observations are reported. Note that these statistics can increase significantly during severe events as wind and wave amplitudes increase.

3.1.1 Satellite data

In this study, post-processed satellite altimeter data (wind speed and Ku-band significant wave height), collected by six altimeter missions (Sentinel-3A, Sentinel-3B, CryoSat-2, SARAL, Jason 2 and Jason 3) are used. Correction algorithms are applied for individual altimeter raw data based on its specific criteria (Queffeulou and Croiz é Fillon 2012). For wind speed, the calibrated values of normalized back-scatter from satellite altimeters (sigma0) and buoy comparison are used for correction (Abdalla 2012). For significant wave height, a linear correction is applied using buoy comparison (Queffeulou 2004). Since the buoy observations are time series at a stationary point, the projection into space is done using wave group velocity for significant wave height error estimation. The satellite footprints within our numerical domains, consisting of $\sim$ 68-k scattered data points, are shown in Fig. 2 with a temporal color bar covering August 27–September 12. The data are divided into two categories: the bold ones, within 7.5° of hurricane eye ($\sim$ 6600 samples) versus far distance ones > 7.5° ($\sim$ 61000 samples). This is done to separate the data within the active zone, where the complicated hurricane core and its inherent uncertainties are under investigation. The satellite tracks move at the speed of $\sim$ 0.05 degree/s with a sampling rate of $\sim$ 1 Hz. On the

other hand, the outputs of atmospheric and wave model are hourly on variable grid resolutions (HWRF on moving inner nested domains with resolutions of 0.099/0.033/0.011° and WW3 with variable resolutions of the unstructured grids from 110 km offshore to 200 m in nearshore regions). Therefore, proper projection and averaging are required for the validation and statistical analysis. In this regard, the model outputs are interpolated to the satellite data, where linear interpolation for time and Inverse Distance Weighting (IDW) are used to average between the three and four nearest points for unstructured and structured grids, respectively. Then model and satellite data are sorted in time for each altimeter separately. Finally, the data are averaged every $\Delta x = 0.5$ degrees in space.

3.1.2 Point source observations

In this study, the time series of meteorological and wave parameters including wind speed U_{10}, wind direction, significant wave height (H_s), peak period (T_p), and mean wave direction are compared at NDBC buoy locations. NDBC wind measurements are six 10-min average values of wind speed and direction reported each hour. Wave measurements are 20-min average value (Gilhousen 1987; Steele and Mettlach 1993). The NDBC data for this study are provided either every 10 min or hourly while the

Fig. 2 Satellite altimeters with track date within numerical domains for the east coast of the USA; The data, collected within 7.5° of the hurricane eye, are shown in bold, Hurricane Irma best track and time tags are shown by magenta

HWRF/WW3 models' outputs are hourly. For the sake of a fair comparison, we averaged the NDBC data every hour. As shown in Fig. 1, these gauges are located along the hurricane track from genesis to the landfall on both sides of the Florida peninsula where meteorological and wave parameters are collected.

4 Atmospheric model

We have used the Hurricane Weather Research and Forecasting (HWRF) model (Tallapragada et al. 2014b; Gopalakrishnan et al. 2010), which is equipped with a movable multilevel nesting technology (Zhang et al. 2016) and designed for extreme events like hurricanes. The HWRF model is a primitive-equation, non-hydrostatic, coupled atmosphere-ocean model with an atmospheric component that employs the Non-hydrostatic Mesoscale Model (NMM) dynamic core of the WRF model (WRF-NMM), with a parent and two nest domains. The parent domain covers roughly $77.2° \times 77.2°$ on a rotated latitude/longitude E-staggered grid. The location of the parent domain is determined based on the initial position of the storm and on the NHC/ Joint Typhoon Warning Center (JTWC) forecast of the 72-h position, if available. The middle nest domain, of about $17.8° \times 17.8°$, and the inner nest domain, of about $5.9° \times 5.9°$, move along with the storm using two-way interactive nesting. The stationary parent domain has an effective grid spacing of about 13.5 km, while the middle and inner nested domains have effective grid spacing of about 4.5 km and 1.5 km, respectively. The dynamic time steps are 30, 10, and 3.33 s, respectively, for the parent, middle nest, and inner nest domains. The model has 75 vertical levels with a model top at 10 hPa. The system is flexible so that different model tops and numbers of vertical levels can be used (Biswas et al. 2018). The two inner nests follow the hurricane best track, ensuring the highest resolution around the eye of the hurricane. The HWRF model is utilized in hurricane forecasting with 4 cycles per day, each projecting 120 h ahead of the hurricane in real time and adjusted itself every 6 h with assimilation of field data. In addition, the model runs on 40 ensemble members, providing an opportunity for probability and uncertainty analysis. Here, the HWRF model configurations are slightly changed to use known atmospheric conditions based on existing observations (after data quality control) and generate semi-hindcasted wind forcing. In order to save computational resources, each cycle lasts 9 h. This model provides hourly outputs which are necessary for a rapidly changing hurricane wind fields. In this study, and at each output time step, the wind field from the highest resolution nested domain is used, extracted from the HWRF model and its ensembles. We used atmospheric fields generated

by HWRF coupled to the Princeton Ocean Model (POM) (Yablonsky et al. 2015) and Fully cycled HWRF ensemble hybrid data assimilation based on hybrid Ensemble Kalman Filter (EnKF) (Zhang et al. 2009) using satellite data. The HWRF model was forced with the boundary condition (B/C) provided by the Global Forecast System (GFS) with 0.25-degree spatial grid resolution, high-resolution GFS analysis for initial condition (I/C) and best track parameters, generated by NHC guide in real time.

The HWRF ensembles are generated by initial/boundary conditions perturbations (large scale) and model physics perturbations (vortex scale) including stochastic Convective Trigger Perturbations in GFS Simplified Arakawa Shubert (SAS), Stochastic boundary layer height perturbations in Planetary Boundary Layer (PBL) scheme, Stochastic Cd perturbation and Stochastic initial wind speed and position (best track parameters) perturbations considering best track uncertainty (Zhang et al. 2014). As a unified model at NOAA, the hybrid ensemble-variational data assimilation system based on Grid-point Statistical Interpolation (GSI; Wu et al. 2002) is developed to provide the HWRF analysis through assimilating all kinds of available conventional and satellite observations. A combination of static background error covariance calculated with the National Meteorological Centre (NMC) method (Parrish and Derber 1992) and a flow-dependent background error covariance estimated from 6-h ensemble forecasts is used in this hybrid data assimilation system. In HWRF, 40-member high-resolution ensemble forecasts initialized by the Ensemble Kalman Filter (EnKF) are designed to generate the flow-dependent error covariance, which accounts for 80% of the entire background error covariance (Wang and Lei 2014; Kleist and Ide 2015; Tong et al. 2018).

An ensemble is a collection of two or more simulations running in parallel to estimate the probability density function of reconstructed fields due to the presence of inevitable uncertainties in the model and observations. The motivation for using ensemble models is to reduce the generalization error of the prediction. The source of uncertainties comes from either observational errors, poor data coverage, and errors in DA system or misrepresentation of model dynamics/physics (chaotic and nonlinear nature), impact of subgrid scale features. The more diverse and independent the members, the less error in the prediction. Here, and for this hindcasted case, the deterministic run and mean of ensembles are quite close to each other. However, it has been proven that a well-designed ensemble system will not only help represent the uncertainty (and spread) well but it will also give us better products (Alaka et al. 2019; Zhang et al. 2014) compared to deterministic simulation, especially in forecast. Although the uncertainty is small at the beginning, the errors can grow fast due to the aforementioned key

roles; therefore, the simulated results will diverge from observation. The effectiveness of ensemble modeling has been demonstrated in operational forecasting systems. Similarly, the uncertainties in hindcast modeling exist due to the presence of uncertainties in storm position, intensity, and structure, the large scale flows and Multi-scale interactions among subgrid scales.

4.1 Track analysis

The best track parameters are subjectively derived analyses of the hurricane locations, intensities, and structures. While subjective, these are real and useful estimates of the hurricane structures, intensities, and locations, which are based on the available observations (Knaff et al. 2011). A schematic view of hurricane vortex and its parameters, summarized in the best track parameters' table, is shown in Fig. 3 where a counterclockwise hurricane vortex is split into four quadrants (NW, SW, SE, NE), each one is defined by the maximum wind speed (V_{max}) with Radius of Maximum Wind (RMW) from the center ($\partial V_g / \partial r = 0$). The radius of three wind thresholds (64, 50, and 34 knots), required to define the wind profile, are shown by magenta, blue, and red lines respectively. The central pressure at Mean Sea Level (MSL) and background pressure, required for surge models are also summarized in the best track parameters' table:

– Hurricane Best Track.
– Maximum wind speed.
– Radius of Maximum Wind (RMW).
– Pressure at MSL.
– Radii for 34, 50, and 64 thresholds.

The HWRF model uses this information as an initial condition (I/C) at the beginning of each cycle. Besides, the observed best track parameters (issued by NHC), forty tables of best track parameters are generated by the model,

representing each individual member. In our analysis, the observed data are compared with the spread of ensemble members (mean and standard deviation σ) in order to reveal the uncertainties around hurricane main parameters.

4.1.1 Hurricane best track

The simulated hurricane tracks for the first 6 h of each cycle (c_i) are shown in Fig. 1a where the model is reinitialized at $t = 0$ for 40 ensemble members. The ensemble spread tends to broaden as the model steps forward, which leads to a wider spread at $t = 6$ h compared to $t = 0$ and hence discontinuity in the whole hurricane period. This discontinuity comes from the independent nature of perturbations at the initialization step. The offsets from the observed hurricane best track are shown in panel b for $t = 0$ (red) and $t = 6$ h (blue) where positive and negative values represent simulated tracks on the east and west sides of observed track, respectively. As is shown in Fig. 1, the mean of ensembles at $t = 0$ are bounded within $[-0.15 \ 0.21]° \pm 0.06°$ of best track while the mean of $t = 6$ h are within $[-0.29 \ 0.52]° \pm 0.1°$ of hurricane track.

4.1.2 Maximum wind speed (U_{10})

The time series of the maximum wind speed, 10 m above MSL, is shown in Fig. 4a, where the black line represents observed data and red and blue lines are the mean and spread of the ensemble members at $t = 0$ and $t = 6$ h respectively. The 34, 50, and 64 kn thresholds are shown in the vertical axis, determining tropical depression (< 33 kn), tropical storm ($34 - 63$ kn), and category 1 (> 64 kn), respectively, corresponding to the strength below which the shape of the vortex is no longer semi-symmetric and consequently the model uncertainty increased. As is shown in Fig. 4a, Hurricane Irma quickly reached a high wind speed (~ 155 kn—category 5) 5 days after its genesis and retained its intensity until it lost most of its energy over Cuba on

Fig. 3 A schematic view of translating wind vortex at the hurricane eye represented in terms of radii for 34, 50, and 64 knots wind thresholds, max wind speed, and its radius in 4 quadrants
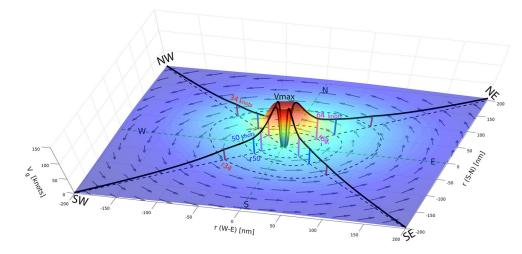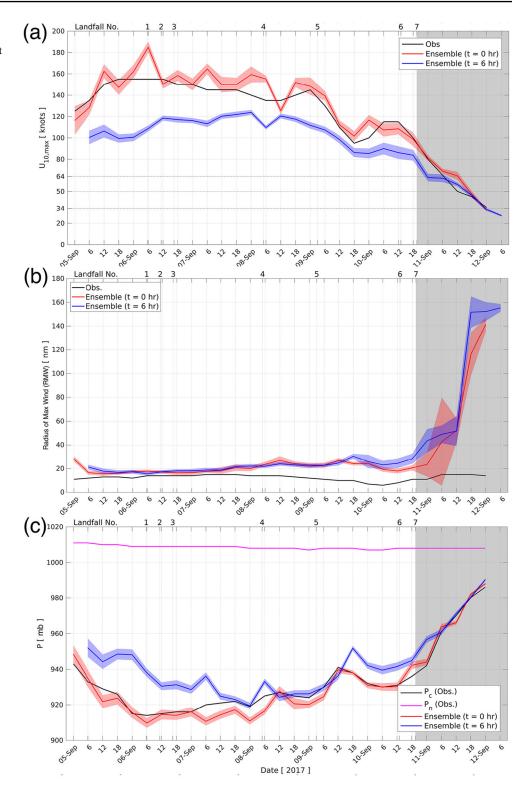
Fig. 4 The mean and spread of HWRF ensemble members in term of standard deviation (σ) at t = 0 (red) and t = 6 hrs (blue) for max wind speed (panel a); Radius of Maximum Wind (RMW) (panel b) and central pressure ,P_c and background pressure, P_n (panel c). In all subplots, seven landfalls are marked as explained in Fig. 1. The gray area shows the time after final landfall. *All model configurations and results are pre-decisional and for official use only*

September 9th. It re-intensified again to nearly 100 kn on September 10th. The wind speed dropped significantly after September 11th when it hit the main land of the USA and its structure reshaped. The modeled U_{10} is close to the observation, mostly overestimated at the beginning of each cycle within $\overline{\text{Model} - \text{Obs}} \pm \sigma = [-9.6\ 30] \pm$

5.2 kn of observed ones. The model tends to underestimate the maximum wind speed as model progresses to $t = 6$ h with a larger bias relative to the observation until the landfall in Florida on September 11th. The ensemble mean varies within $[-55.5\ 6.5] \pm 4.1$ kn of the observation at $t = 6$ h.

4.1.3 Radius of Maximum Wind

The time series of RMW is shown in Fig. 4b where it remains $\sim 20 \pm 3$ nm until the maximum wind speed dropped below 100 kn on September 9th and since then enlarged to ~ 150 nm with a wider spread of uncertainty. For the whole period of Hurricane Irma, the RMW varies between 15.5 and 141.3 with standard deviation within the range of $\pm[1 \ 38]$ nm for $t = 0$ and between 15.6 and 155.3 nm with $1.3 < \sigma < 13.25$ nm for $t = 6$ h.

4.1.4 Pressure at MSL (P_c and P_n)

The recorded background pressure (P_n) during Hurricane Irma was between 1007 and 1011 mb with average of 1008.5 mb. Figure 4c shows the pressure depression at the hurricane center P_c varying from 943 and 984 mb with the lowest value on September 6th. Similar to maximum wind speed trend, the central pressure tended to recover to the background values during each landfall and degraded once it intensified after entering the Straits of Florida. Note that the mean of ensembles tended to recover the balance in the pressure differences. As a result, the central pressure increased to get closer to the background pressure after cycle reinitializations. The variability of ensemble means is almost 2.5 mb for either initial time step and $t = 6$ h.

4.1.5 Radii for 34, 50, and 64 thresholds

As illustrated in Fig. 3 and in each quadrant, the wind profile is defined by the RMW and the distances of the 34, 50, and 64 knot thresholds, and therefore the hurricane intensity and impacted area. Although the cone of a hurricane is asymmetric by nature, the vortex retains its shape as it moves over open water, ideal for an atmospheric model like HWRF. On the contrary, the land geographical irregularities reshape the vortex structure, decrease the wind speed, and widen the impacted area. Such behavior is illustrated by the 34, 50, and 64 contours in Fig. 5a, b where the mean radii started from 102, 57, and 32 nm at the beginning and ended at 315, 102, and 52 nm as it crossed the Caribbean islands and finally hit the main land of the USA. Unlike the other parameters, the mean of ensemble members converges to the observational values from $t = 0$ to 6 h as is shown in Fig. 5c. Similarly, the spread of ensemble means decrease from 9.65, 5.36, and 4.06 nm at $t = 0$ to 6.41, 5.26, and 2.7 nm at $t = 6$ h for 34, 50, and 64 thresholds, respectively.

5 Wave model

Ensemble modeling requires a massive HPC environment and a scalable highly efficient numerical model. However, it is known that spectral wave models (i.e., WAVEWATCH III) are relatively expensive especially on large unstructured grids with very high-resolution grid cells, in which the smallest cell size governs the model time step (due to CFL constraints in the explicit solver). In these regards, substantial improvements in the WW3 model are required. Recent developments in WAVEWATCH III on unstructured grids have pushed the limits of the model in terms of minimum grid size and computational efficiency. These developments include a new parallelization based on a domain decomposition algorithm and a robust implicit solver. In this study, the WW3 model (V6.07) with the implicit scheme and domain decomposition parallelization is utilized (Abdolali et al. 2020) where a unified time step for global, spatial propagation, intra-spectral propagation, and source term is used ($\Delta t = 300$ s). In all simulations, the model resolves the source spectrum with frequencies between 0.05 and 0.9597 Hz, divided into 32 spectral bands and 36 directions with $10°$ increment. In order to include the effect of distantly generated swell, boundary conditions are imposed at the eastern open boundary nodes of the numerical domain, extracted from a global simulation on a structured grid with $0.5°$, forced by GFS wind field. In addition, Ardhuin et al. (2010) source term parameterizations (ST4), nonlinear wave-wave interaction using the discrete interaction approximation, DIA (Hasselmann et al. 1985), moving bottom friction (SHOWEX-BT4) (Ardhuin et al. 2003), depth-limited breaking based on Battjes-Janssen formulation (DB1) (Battjes and Janssen 1978), triad nonlinear interactions (Lumped Triad Interaction method LTA) (Eldeberky and Battjes 1996) and reflection by the coast (REF1) (Ardhuin and Roland 2012) have been used for the computations.

Two methods have been used to generate hourly atmospheric forcing from 40 sets of HWRF model outputs. The first one is taken from the HWRF outputs directly on the entire numerical domain. The same number of ensemble members is resampled randomly. To generate the forcing, u and v components of U_{10} are interpolated on a grid with the resolution of inner domain (1.5 km). Then, the parent domain data is used outside the middle domain. Similarly, the middle domain data is used outside of the inner domain. As a result, the highest quality data is kept. This procedure is done iteratively for all members, resulted in a four-dimensional array of atmospheric data (i, j, n, k) where i and j are the coordinates of the grid node, n is the number of original HWRF member, and k is the time. Finally, the mth forcing fields $u(x, y, t)$ and $v(x, y, t)$ are filled on grid node $x = i, y = j$ at time $t = k$ and from random values between 1 and 40. Since this fields are randomly extracted from the original HWRF members, the statistics of the original data and the generated ones are the same (green and light blue clouds for satellite and buoy data respectively in Fig. 6).
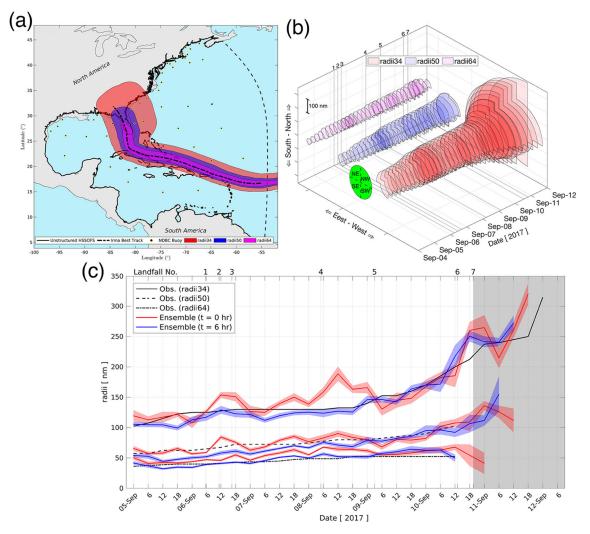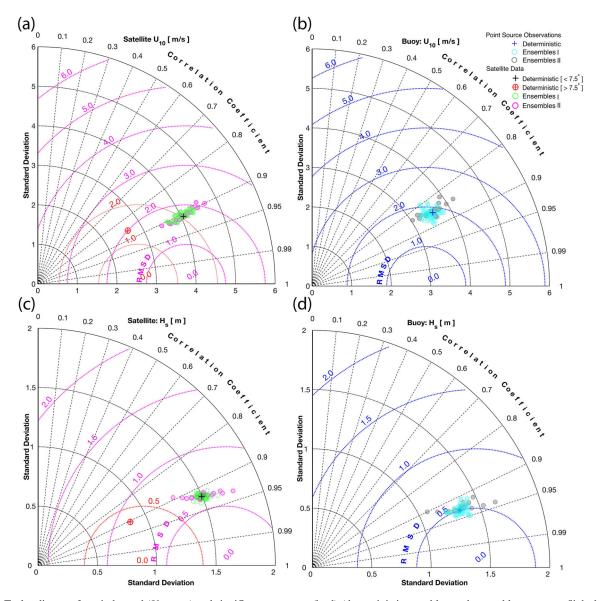
Fig. 5 Hurricane Irma 34, 50, and 64-knot wind threshold contours (panel a); temporal evolution over quadrants (panel b), and model ensembles' means and their spreads versus observation (panel c). Seven landfalls are marked as explained in Fig. 1. The gray area shows the time after final landfall. *All model configurations and results are pre-decisional and for official use only*

The second method is based on the spread of hurricane track information (summarized Section 4.1) and the mean of HWRF members (this time the mean of all members generated with the same methodology as method one). In this method, the mean of HWRF ensembles is perturbed 40 times to generate a smooth evolution of the winds forcing the wave model, rather than the potentially jumpy atmosphere tracks. Each of them represents either the cross-track error (on the eastern or western sides of best track), along-track error (moving ahead or behind the best track), intensity error (larger or smaller wind speed), or size error (wider or narrower RMW and radii 34, 50, and 64 kn) at $t = 0$ and $t = 6$ h or random combinations of aforementioned parameters (see supplementary video

as an example of HWRF perturbation for the hurricane size: Wider (left), Mean (center), and Narrower (right)). The spread of each parameters is taken from analysis of observed best track parameters and HWRF ensembles. Note that the HWRF model is focusing on hurricane vortex and the moving high resolution nested grids are always centered the hurricane eye. However, the outer structure of the wind field in hurricanes is important and can make a noticeable difference in the wave field and cannot be neglected. Therefore, the mean of HWRF ensembles are used in all forcing data. It should be pointed out that that the statistical analysis is performed on the atmospheric and wave data, generated from the aforementioned methods (as shown in Fig. 6). However, and for the sake of visibility, we

Fig. 6 Taylor diagram for wind speed (U_{10}: a, c) and significant wave height (H_s: c, d), representing modeled and collected data along satellite track (a, c) (deterministic run: black (near track $<$ 7.5°)/red (far field $>$ 7.5°) and ensemble runs: magenta/green) and at buoy locations (b, d) (deterministic run: blue and ensemble runs: gray/light blue) in terms of the Pearson correlation coefficient, the root mean square deviation (RMSD), and the standard deviation σ. *All model configurations and results are pre-decisional and for official use only*

show time series and scatter plots for data, taken from the forcing dataset used by the wave model rather than direct from HWRF in Section 6 (method 2).

6 Results and discussion

We first compared the atmospheric and wave model outputs along satellite altimeters track inside our numerical domain for the period of Hurricane Irma (1–12 September, 2017) as shown in Fig. 2. Our analysis is done for the U_{10} and H_s observations in the vicinity of hurricane cone and far field observations separately. Figure 6a and b show Taylor

diagrams of these results, combining standard deviation (σ), the root mean square deviation (RMSD), and correlation coefficient (CC) for the observation, model ensembles, and deterministic run ($\sim$mean of ensembles). For the near hurricane data ($<$ 7.5°), the deterministic runs have standard deviations of 4.05 m/s and 1.5 m corresponding to observational values of 3.75 m/s and 1.58 m within the range of 3.46–4.68 m/s and 1.27–1.76 m for wind and wave parameters, respectively. The RMSD of 1.71 within range of 1.57–2.09 m/s for HWRF ensembles and 0.62 within range of 0.59–0.73 m for WW3 ensembles are estimated. In addition, the cloud of ensembles have CC in the range of 0.89–0.91 and mean of 1.71 m/s for wind speed while the

WW3 model has correlation coefficient within 0.89–0.93 range with mean of 0.92 m for the significant wave height. The biases of deterministic runs are 0.09 m/s and −0.2 m for wind and wave respectively, while the ensembles' biases are within the range of [−1.08 1.26] m/s and [−0.65 0.28] m. On the contrary and for far distance observations (> 7.5°), the HWRF and WW3 models have σ = 2.63 m/s and 0.86 m relative to the value of 2.54 m/s and 0.89 m for the observations, RMSD = 1.36 m/s and 0.38 m, and CC = 0.86 and 0.9. Note that the apparent better performance at far distance observations is due to smaller values of wind and wave activities, away from the hurricane.

For the sake of visibility, we could not show 40 clouds of scatters plots and corresponding linear regression plot for the ensemble runs; however, we have shown the lower and upper limit of all 40 linear regression plots with magenta. Those lines are within two dashed magenta lines.

In addition, from linear regression analysis, a slight overestimation of HWRF model is observed with skill of 1.01 for the deterministic run (red line) within the range of 0.95–1.05 for 40 ensemble members (lower and upper limits are shown by magenta lines) while WW3 underestimates the significant wave height with skills of 0.91 for the deterministic run within range of 0.85–0.98 for 40 ensemble members (Fig. 7a and b).
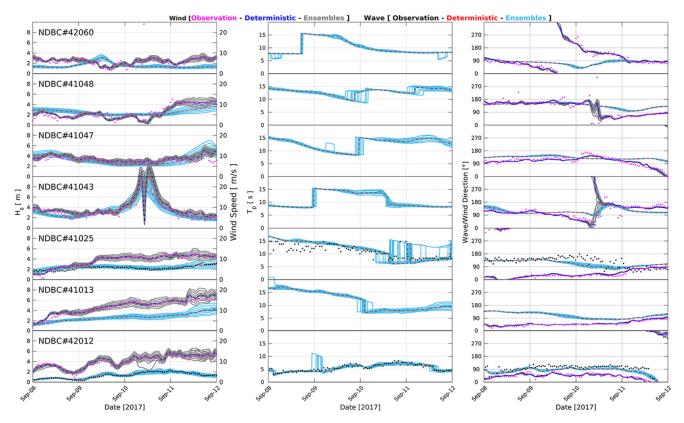
Second, we analyzed the time series of atmospheric and wave model outputs at NDBC point source observations (Fig. 1). The results are shown in Figs. 8 and 9 as time series of wind speed and wind direction for HWRF model and significant wave height (H_s), peak period (T_p), and mean

Fig. 7 Linear regression comparison between satellite altimeter data (a, b) and at buoy observations (c, d) versus HWRF and WW3 models, for wind speed (U_{10}: a, c) and significant wave height (H_s: b, d) . The linear regression (red) for deterministic run and ensemble upper and lower limits (magenta) are shown in each subplot. *All model configurations and results are pre-decisional and for official use only*
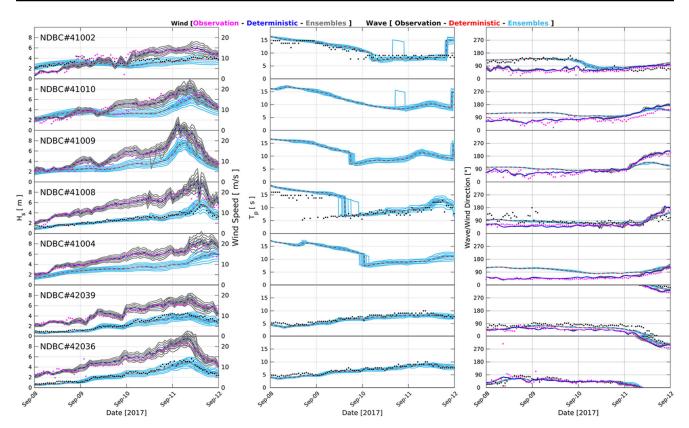
Fig. 8 Atmospheric and wave model validation at the NDBC buoy locations (offshore), deterministic HWRF model (blue) versus observations (magenta), and deterministic WW3 model (red) versus observations (black). The spread of ensemble members (σ) is shown in gray and light blue: (left) Significant wave height (H_s) and wind speed U_{10}; (middle) peak period (T_p) and (right) mean wave direction and wind direction. *All model configurations and results are pre-decisional and for official use only*

wave direction for WW3 model. The time series of model ensembles spread around mean values are shown for wind speed and significant wave height with gray and light blue, respectively, where the observations fall within ensemble spread. Overall performance at buoy locations is shown in the Taylor diagrams presented in Fig. 6. For wind speed, the standard deviation (σ) for the ensemble varies between 3.04 and 4.19 with mean of 3.57 m/s relative to 2.89 m/s for the observations while the root mean square deviation (RMSD) range is 1.7–2.3 with mean value of 2 m/s. The correlation coefficient (CC) range is 0.82–0.87 with mean value of 0.85. A better correlation coefficient is observed for the significant wave height time series at buoy location with range of 0.9–0.94 and mean value of 0.93. The standard deviation of the deterministic run is 1.33 m with range of 1.07–1.63 relative to 1.38 m for the observation. The RMSD for the mean of members is 0.6 within 0.52–0.74 range. The biases of deterministic runs are 0.31 m/s and −0.08 m for wind and wave respectively, while the ensembles' biases are within the range of [−0.85 1.47] m/s and [−0.48 0.36] m. The skill of 0.99 for the deterministic run (red line) within range of 0.94–1.04 for the ensemble members (lower and upper limits are shown by magenta lines) for wind speed

compared to 0.96 for the deterministic run within range of 0.88–1.03 for the ensemble members for H_s are obtained, as shown in linear regression plots (Fig. 7c and d).

In addition to the analysis on the spatiotemporal data collected by satellite altimeters and time series at stationary NDBC buoys, snapshots of wind and wave fields, and the spread of ensemble members in the vicinity of hurricane cone are shown in Figs. 10 and 11 for HWRF and WW3 models respectively. The interval between snapshots is 24 h and the spatial span is 10°, centered at hurricane eye where the uncertainty is concentrated. It is clear that the larger uncertainty in the wind leads to larger uncertainty in the wave products. The comparison between best track analysis in Section 4, and wind field plots indicate that the ensemble spread is smaller when the hurricane is blowing over open water. As the hurricane gets closer to geographical irregularities, the misrepresentation of small islands due to model resolution, or more importantly, uneven destruction of vortex structure lead to larger uncertainly in the atmospheric model and subsequently wave model outputs. Furthermore, as hurricane gets wider on its way, the uncertainty in the radius increases and therefore spreads of wind and wave get wider. A closer

Fig. 9 Atmospheric and wave model validation at the NDBC buoy locations (nearshore), deterministic HWRF model (blue) versus observations (magenta), and deterministic WW3 model (red) versus observations (black). The spread of ensemble members (σ) is shown in gray and light blue: (left) Significant wave height (H_s) and wind speed U_{10}; (middle) peak period (T_p) and (right) mean wave direction and wind direction. *All model configurations and results are pre-decisional and for official use only*

look also shows that the maximum uncertainty in the wind field occurs at the hurricane eye while there is a depression in the uncertainty in wave fields at hurricane center. This behavior is due to the fact that the wave field integrates the momentum transferred from the wind field in space and time, and hence leads to less variation in the wave height.

7 Accuracy assessment with paired *t* test

For locations where in situ observations are available, we can perform a statistical test to evaluate model accuracy. For a given ensemble model run, or the ensemble mean, we need a test to compare the modeled time series to the observed time series. Here we aim to test whether a given level of accuracy (here 90%) is reached, which is equivalent to an error level of 10%. We consider this requirement to be stricter than achieving a mean bias of <10%, since a significant number of individual model data points could still differ by more than 10% from the corresponding observation. At the same time, it is considered unreasonably strict to require that every model data point has an error of less than 10%, considering the natural variability in the

observed phenomenon (e.g., wind U_{10} wave height H_s) and observational error. As a result, the accuracy assessment will focus on the mean relative difference between the modeled and observed time series, and test whether this mean difference is below 10%. The paired *t* test accounts for observational uncertainty as well as the expected differences due to model errors.

Since the model and observation both describe the same process (e.g., wind speed or wave height), there is a dependence between the modeled and observed time series variables. In this setting, the paired *t* test hypothesis test is appropriate. To test whether the mean difference between these two time series is less than 10%, we set the following null hypothesis H_0 and alternative hypothesis H_a:

$$H_0 : \mu_d < 0.1$$
$$H_a : \mu_d > 0.1 \tag{1}$$

where the mean relative difference is defined as $d_i = (X_{i,\mathrm{mod}} - X_{i,\mathrm{obs}})/X_{i,\mathrm{obs}}$, and X_i is the model variable at time i being tested. Since the alternative hypothesis states that the relative difference is greater than 0.1, this constitutes an upper-tailed test. This test has the following assumptions (Ott and Longnecker 2015):
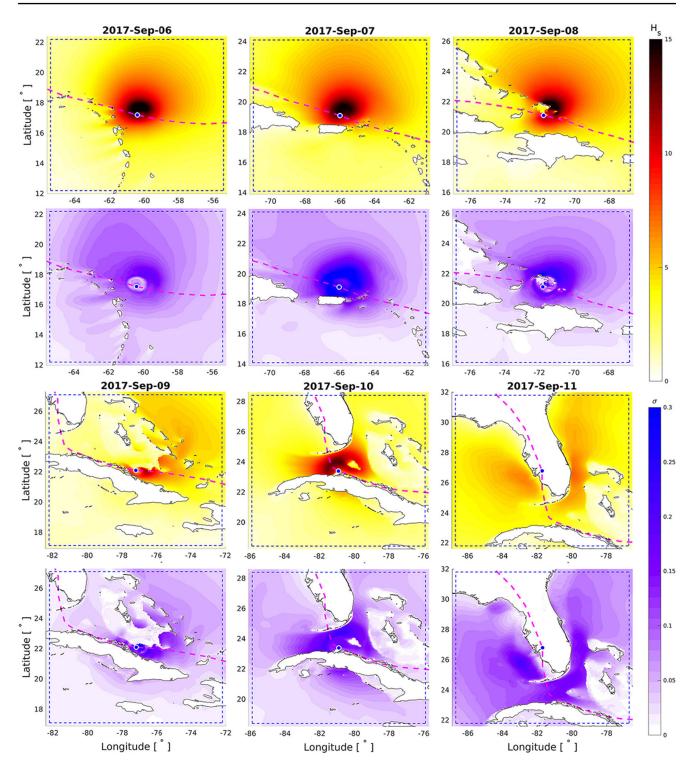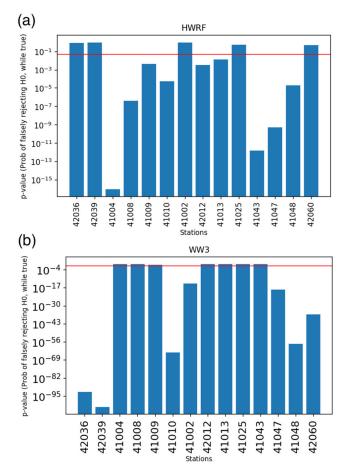
Fig. 10 Snapshots of ensemble mean for wind speed U_{10} (m/s) between 6 and 11 September (rows 1 and 3), and corresponding standard deviation σ (rows 2 and 4) close to the hurricane eye. The variations are indicated with reference to the color bar where white corresponds to 0. *All model configurations and results are pre-decisional and for official use only*

– That the sampling distribution of d_i is a normal distribution.
– That the d_i samples are independent.

The hypothesis test is conducted at the standard level of significance of $\alpha = 0.05$. This means that the null hypothesis that the mean difference between two time series at a given

Fig. 11 Snapshots of ensemble mean for significant wave height H_s (*m*) between 6 and 11 September (rows 1 and 3), and corresponding standard deviation σ (rows 2 and 4) close to the hurricane eye. The variations are indicated with reference to the color bar where white corresponds to 0. *All model configurations and results are pre-decisional and for official use only*

station is less than 0.1 (or 10%) should be rejected if the *p* value of this statistical test is < 0.05. In practical terms, this means that the probability of erroneously rejecting the null hypothesis (that 90% accuracy is met), given that it is true, is less than 5%.

Figure 12 shows the results of the paired *t* tests at the NDBC stations for the ensemble means of HWRF and WW3. Indicated in the two subplots is the critical *p* value = 0.05. In the upper panel of Fig. 12, we see that for HWRF five stations (42036, 42039, 41002, etc.) have *p* values exceeding 0.05, so that for these, the null hypothesis that the mean relative difference between the modeled and observed time series is less than 0.10 cannot be rejected. These stations are therefore considered 90% accurate. By contrast, at the remaining nine stations (e.g., 41004 and 41043), *p* values are close to zero, indicating that the null hypothesis can be rejected at a significance level of $\alpha = 0.05$. The lower panel in Fig. 12 shows the corresponding results for WW3. Here we see that a number of stations that did not reach 90% accuracy in terms of wind speed actually reach that accuracy level in terms of significant wave height (e.g., 41004, 41008, and 41009). The poorest

performers (lowest *p* values) are found to be the stations 42036 and 42039 offshore of Tampa in the Gulf of Mexico, despite having accurate winds. We speculate that this is due to the complex land-sea transition at these stations, due to the offshore winds from the vortex at this location.

8 Conclusion

In this study, we have performed a comprehensive statistical analysis on the atmospheric and spectral wave models performance in a hindcast setting for Hurricane Irma (2017), a category 5 hurricane which made its final landfall in southwest Florida. A well-known atmospheric model, designed for hurricane modeling (HWRF), is used to drive WW3, which utilize the latest updates of the operational HWRF at the National Centers for Environmental Prediction (NCEP), incorporate the post-processed data for data assimilation, a high-resolution topography, and a high resolution land-sea mask (Tallapragada et al. 2014a; Tong et al. 2018; Ma et al. 2020). The spread of the atmospheric model is evaluated using 40 ensemble members from semi-hindcasted HWRF model simulations, for which each member is initiated independently and accounts for the unresolved physics and stochastic distribution of error. Forty sets of continuous wind fields were generated around the mean of HWRF ensembles, which represent cross-track error, along-track error, intensity error, or size error. These forcing were used to drive a WW3 model. The recent advances in the WW3 model on unstructured triangular meshes, including the new parallelization algorithm and implicit numerical solver, have made the model more efficient and accurate, bypassing numerical restrictions and CFL constraints (Abdolali et al. 2020). These new capabilities allowed us to run WW3 on a suite of ensemble members on an unstructured grids with ~200-m resolutions near the US East Coast and adequate eastward extent, allowing for appropriate generation of hurricane waves from winds over a large basin. Hence, the error propagation from atmospheric model to the wave model was tracked and analyzed.

We have performed a validation study that compares the atmospheric and wave models' results with satellite altimeter data for wind speed and significant wave height, with hurricane track information against observed and interpreted ones by NHC at the beginning and *t* = 6 h of each cycle, and with point source observations from the NDBC network for meteorological and wave parameters. The wave model forced by the available HWRF ensemble winds reveals the uncertainties and errors embedded in the upstream atmospheric model that are propagated downstream to the wave model. The HWRF and WW3 models' performances were evaluated at stationary NDBC buoys and along satellite altimeter footprints revealing a

Fig. 12 Paired *t* tests results at the NDBC stations for the ensemble means of HWRF (a) and WW3 (b). *All model configurations and results are pre-decisional and for official use only*

good agreement between model outputs and observations. As shown in Figs. 8 and 9, wind and wave observations are within ensemble spreads, implying that the spread of HWRF and WW3 outputs is adequate to cover observations. In addition, the estimated hurricane track from the HWRF model was compared to observations. A detailed illustration of hurricane track parameters and statistical analysis is discussed in this paper. It is shown that HWRF ensemble has a wider spread over areas with landmass due to the absence of subgrid scale features in the model, misrepresentation of the complex land-air-sea interaction in the model equations, and uneven destruction of vortex structure. The migration of the errors, introduced by the atmospheric models and manifested in the wave model, shows similar and different model behaviors. For example, wider spread of atmospheric model, specially near landmasses, led to a wider spread of wave model outputs. On the other hand, maximum uncertainty in the atmospheric model is at hurricane eye while wave model uncertainty is small at the center due to less variation of momentum transfer from wind to wave model. Along with uncertainty analysis of atmospheric and wave model across the entire domain, a paired *t* test is used at observation locations to evaluate whether the mean relative difference between the modeled and observed time series is below a certain value or not. The two aforementioned methods for model uncertainty evaluation (for places with no observation) and the paired *t* test (for locations where observations are available), considering the observational error shed light on the importance of ensemble modeling for uncertainty evaluation of model, and provides metrics for model accuracy evaluation during severe events.

References

Abdalla S (2012) Ku-band radar altimeter surface wind speed algorithm. Mar Geodesy 35:276–298. https://doi.org/10.1080/01490419.2012.718676

Abdalla S, Janssen PA, Bidlot J-R (2011) Altimeter near real time wind and wave products: random error estimation. Mar Geodesy 34:393–406

Abdolali A, Roland A, van der Westhuysen A, Meixner J, Chawla A, Hesser TJ, Smith JM, Sikiric MD (2020) Large-scale hurricane modeling using domain decomposition parallelization and implicit scheme implemented in wave- watch iii wave model. Coast Eng 157:103656. https://doi.org/10.1016/j.coastaleng.2020.103656

Alaka GJ, Zhang X, Gopalakrishnan SG, Zhang Z, Marks FD, Atlas R (2019) Track uncertainty in high-resolution HWRF ensemble forecasts of Hurricane Joaquin. Weather Forecast 34:1889–1908. https://doi.org/10.1175/WAF-D-19-0028.1

Ardhuin F, O'reilly W, Herbers T, Jessen P (2003) Swell transformation across the continental shelf. Part I: attenuation and directional broadening. J Phys Oceanogr 33:1921–1939. https://doi.org/10.1175/1520-0485(2003)033<1921:STATCS>2.0.CO;2

Ardhuin F, Rogers E, Babanin AV, Filipot J-F, Magne R, Roland A, van der Westhuysen A, Queffeulou P, Lefevre J-M, Aouf L, Collard F (2010) Semiempirical dissipation source functions for ocean waves. Part I: definition, calibration, and validation. J Phys Oceanogr 40:1917–1941. https://doi.org/10.1175/2010JPO4324.1

Ardhuin F, Roland A (2012) Coastal wave reflection, directional spread, and seismoacoustic noise sources. J Geophys Res Oceans 117. https://doi.org/10.1029/2011JC007832

Bakhtyar R, Maitaria K, Velissariou P, Trimble B, Mashriqui H, Moghimi S, Abdolali A, Van der Westhuysen AJ, Ma Z, Clark EP, Flowers T (2020) A new 1d/2d coupled modeling approach for a riverine-estuarine system under storm events: application to Delaware river basin. J Geophys Res Oceans 125:e2019JC015822. https://doi.org/10.1029/2019JC015822

Battjes JA, Janssen J (1978) Energy loss and set-up due to breaking of random waves. In: Coastal engineering 1978, pp 569–587. https://doi.org/10.1061/ 9780872621909.034

Bender MA, Marchok TP, Sampson CR, Knaff JA, Morin MJ (2017) Impact of storm size on prediction of storm track and intensity using the 2016 operational GFDL hurricane model. Weather Forecast 32:1491–1508. https://doi.org/10.1175/WAF-D-16-0220.1

Biswas MK, Abarca S, Bernardet L, Ginis I, Grell E, Iacono M, Kalina E, Liu B, Liu T, Marchok Q, Mehra A, Newman K, Sippel J, Tallapragada V, Thomas B, Wang W, Winterbottom H, Zhang Z (2018) Hurricane weather research an forecasting (HWRF) model:

2018 scientific documentation. Available at https://dtcenter.org/HurrWRF/users/docs/index.php

Cangialosi LAS, John P, Berg R (2018) Tropical cyclone report: hurricane irma (al112017), 30 August–12 September 2017

Cavaleri L (2006) Wave modeling: where to go in the future. Bull Am Meteorol Soc 87:207–214

Desroziers G, Brachemi O, Hamadache B (2001) Estimation of the representativeness error caused by the incremental formulation of variational data assimilation. Q J R Meteorol Soc 127:1775–1794. https://doi.org/10.1002/qj.49712757516

Eldeberky Y, Battjes JA (1996) Spectral modeling of wave breaking: application to Boussinesq equations. J Geophys Res Oceans 101:1253–1264

Gilhousen DB (1987) A field evaluation of NDBC moored buoy winds. J Atmos Ocean Technol 4:94–104

Gopalakrishnan S, Liu Q, Marchok T, Sheinin D, Surgi N, Tuleya R, Yablon-sky R, Zhang X (2010) Hurricane weather research and forecasting (HWRF) model scientific documentation. Development Testbed Center

Hasselmann S, Hasselmann K, Allender J, Barnett T (1985) Computations and parameterizations of the nonlinear energy transfer in a gravity-wave specturm. Part II: parameterizations of the nonlinear energy transfer for application in wave models. J Phys Oceanogr 15:1378–1391. https://doi.org/10.1175/1520-0485(1985)015<1378:CAPOTN>2.0.CO;2

Janjic T, Cohn SE (2006) Treatment of observation error due to unresolved scales in atmospheric data assimilation. Mon Weather Rev 134:2900–2915. https://doi.org/10.1175/MWR3229.1

Kleist DT, Ide K (2015) An osse-based evaluation of hybrid variational– ensemble data assimilation for the NCEP GFS. Part ii: 4denvar and hybrid variants. Mon Weather Rev 143:452–470

Knaff JA, Sampson CR, Fitzpatrick PJ, Jin Y, Hill CM (2011) Simple diagnosis of tropical cyclone structure via pressure gradients. Weather forecast 26:1020–1031. https://doi.org/10.1175/WAF-D-11-00013.1

Ma Z, Liu B, Mehra A, Abdolali A, van der Westhuysen A, Moghimi S, Vinogradov S, Zhang Z, Zhu L, Wu K, Shrestha R, Kumar A, Tallapragada V, Kurkowski N (2020) Investigating the impact of high-resolution land-sea masks on hurricane forecasts in HWRF. Atmosphere 11. https://doi.org/10.3390/atmos11090888

Moghimi S, Van der Westhuysen A, Abdolali A, Myers E, Vinogradov S, Ma Z, Liu F, Mehra A, Kurkowski N (2020) Development of an esmf based flexible coupling application of ADCIRC and WAVEWATCH III for high fidelity coastal inundation studies. J Marine Sci Eng 8. https://doi.org/10.3390/jmse8050308

Monbaliu J (2003) Chapter 5: spectral wave models in coastal areas. In: Lakhan VC (ed) Advances in coastal modeling. vol 67 of Elsevier Oceanography Series, pp 133–158. Elsevier. https://doi.org/10.1016/S0422-9894(03)80122-8

Ott RL, Longnecker MT (2015) An introduction to statistical methods and data analysis. Nelson Education

Parrish DF, Derber JC (1992) The national meteorological centers spectral statistical-interpolation analysis system. Mon Weather Rev 120:1747–1763

Queffeulou P (2004) Long-term validation of wave height measurements from altimeters. Mar Geodesy 27:495–510. https://doi.org/10.1080/01490410490883478

Queffeulou P, Croiz é Fillon D (2012) Global altimeter SWH data set-April 2012. Spatial Anography Laboratory, IFREMER

Roland A, Ardhuin F (2014) On the developments of spectral wave models: numerics and parameterizations for the coastal ocean. Ocean Dyn 64:833–846

Sampson CR, Fukada EM, Knaff JA, Strahl BR, Brennan MJ, Marchok T (2017) Tropical cyclone gale wind radii estimates for the western north pacific. Weather Forecast 32:1029–1040. https://doi.org/10.1175/WAF-D-16-0196.1

Steele KE, Mettlach T (1993) NDBC wave data—current and planned. In: Ocean wave measurement and analysis. ASCE, pp 198–207

Tallapragada V, Bernardet L, Biswas MK, Gopalakrishnan S, Kwon Y, Liu Q, Marchok T, Sheinin D, Tong M, Trahan S et al (2014a) Hurricane weather research and forecasting (HWRF) model: 2013 scientific documentation. HWRF Development Testbed Center Tech. Rep, 99

Tallapragada V, Kieu C, Kwon Y, Trahan S, Liu Q, Zhang Z, Kwon I-H (2014b) Evaluation of storm structure from the operational HWRF during 2012 implementation. Mon Weather Rev 142:4308–4325. https://doi.org/10.1175/MWR-D-13-00010.1

Tallapragada V, Kieu C, Trahan S, Zhang Z, Liu Q, Wang W, Tong M, Zhang B, Strahl B (2015) Forecasting tropical cyclones in the western north pacific basin using the CNEP operational HWRF: real-time implementation in 2012. Weather Forecast 30:1355–1373. https://doi.org/10.1175/WAF-D-14-00138.1

Tong M, Sippel JA, Tallapragada V, Liu E, Kieu C, Kwon I-H, Wang W, Liu Q, Ling Y, Zhang B (2018) Impact of assimilating aircraft reconnaissance observations on tropical cyclone initialization and prediction using operational HWRF and GSI Ensemble Variational hybrid data assimilation. Mon Weather Rev 146:4155–4177. https://doi.org/10.1175/MWR-D-17-0380.1

Wang X, Lei T (2014) GSI-based four-dimensional ensemble-variational (4DEnsVar) data assimilation: formulation and single-resolution experiments with real data for NCEP global forecast system. Mon Weather Rev 142:3303–3325

Wu W-S, Purser RJ, Parrish DF (2002) Three-dimensional variational analysis with spatially inhomogeneous covariances. Mon Weather Rev 130:2905–2916

WW3DG (2019) User Manual and System Documentation of WAVEWATCH III version 6.07, The WAVEWATCH III Development Group. Tech. Note 333, NOAA/NWS/NCEP/MMAB,College Park, MD, USA, 465 pp

Yablonsky RM, Ginis I, Thomas B, Tallapragada V, Sheinin D, Bernardet L (2015) Description and analysis of the ocean component of NOAA's operational hurricane weather research and forecasting model (HWRF). J Atmos Ocean Technol 32:144–163. https://doi.org/10.1175/JTECH-D-14-00063.1

Zhang F, Weng Y, Sippel JA, Meng Z, Bishop CH (2009) Cloud-resolving hurricane initialization and prediction through assimilation of doppler radar observations with an ensemble Kalman filter. Mon Weather Rev 137:2105–2125. https://doi.org/10.1175/

Zhang Z, Tallapragada V, Kieu C, Trahan S, Wang W (2014) HWRF based ensemble prediction system using perturbations from GEFS and stochastic convective trigger function. Trop Cycl Res Rev 3:145–161

Zhang X, Gopalakrishnan SG, Trahan S, Quirino TS, Liu Q, Zhang Z, Alaka G, Tallapragada V (2016) Representing multiple scales in the hurricane weather research and forecasting modeling system: design of multiple sets of movable multilevel nesting and the basin-scale HWRF forecast application. Weather Forecast 31:2019–2034. https://doi.org/10.1175/WAF-D-16-0087.1