Using a WRF-ADCIRC Ensemble and Track Clustering to Investigate Storm Surge Hazards and Inundation Scenarios Associated with Hurricane Irma

ALEX M. KOWALESKI^a

Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, Pennsylvania

REBECCA E. MORSS, DAVID AHIJEVYCH, AND KATHRYN R. FOSSELL

National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 14 August 2019, in final form 28 April 2020)

ABSTRACT

This article investigates combining a WRF-ADCIRC ensemble with track clustering to evaluate how uncertainties in tropical cyclone–induced storm tide (surge + tide) predictions vary in space and time and to explore whether this method can help elucidate inundation hazard scenarios. The method is demonstrated for simulations of Hurricane Irma (2017) initialized at 1200 UTC 5 September, approximately 5 days before Irma's Florida landfalls, and 1200 UTC 8 September. Mixture models are used to partition the WRF ensemble tracks from 5 and 8 September into six and five clusters, respectively. Inundation is evaluated in two affected regions: southwest (south and west Florida) and northeast (northeast Florida through South Carolina). For the 5 September simulations, inundation in the southwest region varies significantly across the ensemble, indicating low forecast confidence. However, clustering highlights the areas of inundation risk in south and west Florida associated with different storm tracks. In the northeast region, every cluster has high inundation probabilities along a similar coastal stretch, indicating high confidence at a ~5-day lead time that this area will experience inundation. For the 8 September simulations, track and inundation in both regions vary less across the ensemble, but clustering remains useful for distinguishing among flooding scenarios. These results demonstrate the potential of dynamical TC–surge ensembles to illuminate important aspects of storm surge risk, including highlighting regions of high forecast confidence where preparations can reliably be initiated early. The analysis also shows how clustering can augment probabilistic hazard forecasts by elucidating inundation scenarios and variability across a surge ensemble.

1. Introduction

Track forecasts of North Atlantic tropical cyclones (TCs) have become markedly more skillful during the last two decades (Landsea and Cangialosi 2018), and intensity forecasts have shown improvement during the last 10 years (NHC 2019). However, predicting which locations will experience TC-related hazards (e.g., damaging winds, rain-induced flooding, storm surge) at longer lead times remains challenging. Currently, operational NHC track and intensity forecasts are issued out to 120 h, while hurricane wind and storm surge watches and warnings are only issued 48 and 36 h, respectively, before the expected arrival of the storm (NWS 2019). Longer lead-time hazard forecasts could help public officials, businesses, and

members of the public plan and implement protective actions appropriate to the threats that are most likely to be experienced in different geographical areas. Thus, developing techniques for translating increased track and intensity forecast skill into more skillful hazard forecast information at lead times beyond 48 h offers significant potential for protecting lives and property.

This article seeks to advance the science and practice of longer lead-time TC hazard prediction by using dynamic atmosphere–surge ensemble modeling to explore probabilistic and scenario-based forecasting of storm tide¹

^aCurrent affiliation: Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado.

Corresponding author: Alex Kowaleski, alexmkowaleski@gmail.com

¹ Hazardous conditions from coastal flooding are often produced by a combination of storm-induced surge and tide, which is technically called “storm tide.” Even when the tidal component is included, however, the hazard is often referred to as “storm surge” or “surge” (e.g., in NHC storm surge watches and warnings). Thus, when discussing the predictability of surge-induced hazards from TCs in general, we use NHC's surge terminology, consistent with the high predictability of the tidal component in the absence of storm-induced surge.

hazards in different regions at risk from an approaching TC. We focus on storm surge hazards here because they are leading threats to lives and property from landfalling TCs (Rappaport 2014) and a primary motivation for coastal evacuation when a TC approaches. Although this article examines only storm surge hazards, we anticipate that similar ensemble modeling and analytic techniques could be used to investigate other TC hazards such as high wind, excessive rainfall, and combined surge–wave–rainfall-induced flooding.

As a starting point for investigating these issues, we examine Hurricane Irma (2017) as it approached the mainland United States. Irma is a case of particular interest for storm surge forecasting because it threatened to produce coastal flooding in much of south and west Florida, as well in northeast Florida, Georgia, and South Carolina (Cangialosi et al. 2018). Given the large coastal population in these regions and the geography of the Florida peninsula, significant lead time was required for making evacuation decisions, especially in south Florida. Improved coastal inundation forecasts several days before landfall could have given threatened areas additional time to prepare, while reducing evacuation and other disruptions in areas that ended up experiencing relatively minor impacts. By studying Irma, we aim to explore techniques for surge hazard prediction that may be extended to other storms, retrospectively or in real time.

We approach storm surge hazard prediction from an ensemble perspective because of the uncertainty inherent in predicting TC-induced coastal flooding, especially at specific locations (Morrow et al. 2015; Fossell et al. 2017). Over the last few decades, numerical ensemble simulations have become widely used in operational weather forecasting and in predictability and prediction research (Buizza 2018; Benjamin et al. 2019). For TCs, today's dynamical ensemble forecasts typically represent different possible evolutions of the storm's track and intensity, along with its structure and surrounding environment. The changes in each TC property's ensemble spread with lead time indicates how that property's anticipated forecast uncertainty evolves. Here, we extend dynamical ensemble TC forecasting into the hazard space to explore the anticipated uncertainty and potential information content in different aspects of storm surge forecasts. Although several previous studies have used dynamical atmosphere–surge ensembles to explore coastal flood prediction (e.g., Flowerdew et al. 2010, 2013; Di Liberto et al. 2011; Colle et al. 2015; Georgas et al. 2016; Dietrich et al. 2018), limited work has focused on understanding the characteristics and potential utility of dynamical TC–surge ensemble forecasts.

A common format for presenting forecast uncertainty is a map of the likelihood of a variable (e.g., inundation

depth) exceeding a threshold at each location within a region, as in the NHC's P-Surge forecast products for storm surge risk (Morrow et al. 2015). However, it can be difficult to effectively communicate forecast uncertainty information to decision makers, especially for the low hazard probabilities that tend to be prevalent at longer lead times (NASEM 2006, 2018; Morss et al. 2008; Morrow et al. 2015; Fundel et al. 2019). To overcome this difficulty, one way that emergency managers incorporate predictive uncertainties into their decisions is by using scenarios (Demuth et al. 2012; Bostrom et al. 2016). Thus, we also explore whether partitioning the TC–surge ensemble forecasts into smaller groupings can support interpretation of the predictive data.

Previous work with atmospheric ensembles has found that clustering can elucidate natural groupings within the ensemble by partitioning ensemble members to maximize intergroup spread and minimize intragroup spread, (e.g., Keller et al. 2011, 2014; ECMWF 2019). For TCs, Kuruppumullage Don et al. (2016) and Kowaleski and Evans (2016, 2018) used regression mixture-model clustering (Gaffney et al. 2007) to partition ensemble TC tracks based on their time-dependent trajectories. In this study, we utilize this TC track clustering methodology on an atmospheric ensemble and extend it to explore the resulting clusters in surge hazard space. Although many TC properties (e.g., size, intensity, structure) affect storm surge hazard, we focus here on TC track clusters because track is a leading influence on surge prediction at lead times beyond 24 h (Fossell et al. 2017). As with several other aspects of the methodology, the specific clustering technique used here is an initial exploration of the potential for applying clustering to TC–surge ensembles, and we anticipate that further development of surge clustering techniques may improve their utility.

The research presented here has two main goals. First, we explore what, if anything, a storm surge ensemble generated using an ensemble of dynamical TC forecasts can reveal about surge hazard prediction, beyond the information provided by the TC ensemble itself. Second, we explore what additional information, if any, is provided by clustering of the TC–surge ensemble. We investigate both topics in different regions of the southeastern United States threatened by Irma, for two lead times corresponding to approximately 5 and 2 days before Irma's Florida landfalls. The results elucidate the sensitivity of Irma's surge-induced inundation in different regions to large-scale (intercluster) and small-scale (intracluster) track variations. In doing so, they provide new understanding about how uncertainties in TC surge predictions vary in space and time, and they

illustrate the potential for using dynamical ensembles and clustering to extend the lead time of useful surge prediction, at least in some situations and regions.

To investigate these topics, we use the Weather Research and Forecasting (WRF) Model to generate ensembles of convection-permitting atmospheric simulations of Hurricane Irma at two initialization times. We then use output from these atmospheric simulations to drive simulations of storm tide (surge + tide) and associated coastal inundation using the Advanced Circulation (ADCIRC) model. At each lead time, the ensemble of inundation predictions is examined; then the ensemble of TC tracks is clustered using regression mixture models and the resulting inundation clusters are analyzed. We analyze the results using two metrics: 1) probability of inundation at specific locations in different regions; and 2) magnitude and timing of regionally integrated inundation volume. Given our particular interest in longer lead-time surge predictions, the latter metric is motivated by results in [Fossell et al. \(2017\)](#) indicating that using volume-integrated metrics may help extend the lead time of skillful surge prediction. Results will indicate whether the demonstrated methods help reveal associations between TC track and distinct surge inundation scenarios that could be useful for real-time TC hazard forecasting and risk communication.

Although the WRF-ADCIRC modeling conducted for this study produces a probabilistic storm surge ensemble for Irma, our purpose in conducting these simulations is not to evaluate the skill of probabilistic storm surge forecasts generated from this ensemble or to understand how different aspects of Irma's evolution influenced coastal flooding. Rather, we seek to demonstrate the dynamical ensemble and clustering methodology presented here, and to explore the utility of these techniques for developing deeper understanding of uncertainties in surge predictions and for describing storm surge hazard scenarios.

The remainder of the paper proceeds as follows: an overview of Irma's evolution and its impacts is provided in [section 2](#); the methodology for generating and clustering the ensemble and analyzing the storm tide and inundation output are described in [section 3](#). [Section 4](#) presents results, and [section 5](#) summarizes and discusses key findings and implications.

2. Synoptic evolution of Irma and storm surge

The synoptic history of Irma described here is based on [Cangialosi et al. \(2018\)](#). After cyclogenesis in the eastern Atlantic on 30 August, Irma rapidly intensified, becoming a major hurricane 48 h later. By 1200 UTC

5 September, Irma had attained category-5 intensity while located about 400 km east of the eastern Leeward Islands.

Irma maintained category-5 intensity for 60 h as it ravaged Barbuda and the British Virgin Islands before passing north of Puerto Rico and Hispaniola. It weakened to a category 4 while moving through the southern Bahamas, but regained category-5 status as it turned westward toward Cuba. After landfalling in Cuba early on 9 September, Irma weakened to category-2 intensity. However, when Irma moved northwestward over the Florida Straits early on 10 September, it reintensified into a category-4 hurricane.

Irma made its first of two U.S. landfalls near Cudjoe Key, Florida, at 1300 UTC 10 September, at category-4 intensity (115 kt; 59 m s^{-1}). As Irma moved northward, it weakened due to increasing southwesterly vertical wind shear; however, its wind field expanded substantially, especially east of the center. Irma made its second and final U.S. landfall in southwest Florida, near Marco Island, at 1930 UTC 10 September, with 100-kt (51 m s^{-1}) maximum sustained winds. Over Florida, Irma continued to weaken, falling to tropical-storm strength by 1200 UTC 11 September. It weakened into a remnant low by 0600 UTC 12 September.

Irma's slight rightward jog prior to its final landfall greatly reduced storm surge inundation in much of inhabited southwest Florida ([Cangialosi et al. 2018](#); their Fig. 23); maximum values of approximately 3 m occurred along the uninhabited Monroe County coast. However, Irma's expansive wind field in its northeast quadrant generated a substantial storm surge along the northeast Florida, Georgia, and South Carolina coasts. A storm tide sensor south of St. Augustine, Florida, recorded a water level of approximately 1.46 m above mean higher high water (MHHW; the historical mean of water levels at each day's higher high tide, used as a measure of inundation above normally dry ground), while sensors at Fort Pulaski, Georgia, and Charleston, South Carolina, recorded maximum water levels of 1.44 and 1.27 m above MHHW, respectively ([NOAA Tides & Currents 2020](#)). Water levels at least 1 m above MHHW were also recorded at Ormond Beach, Florida, Jacksonville Beach, Florida, and Oyster Landing, South Carolina. In addition to the primarily surge-induced flooding, Irma's surge combined with rainfall runoff to produce widespread flooding along the St. Johns River in Jacksonville, Florida ([Cangialosi et al. 2018](#)). Here we examine only the surge-induced component of the flooding, but the methods could readily be extended to investigate total (tide + surge + wave + rain) flooding with an integrated flood hazard model.

TABLE 1. Model configuration used in WRF version 3.8 simulations of Hurricane Irma. Model domains are depicted in Fig. 1. Settings used in all three domains are only listed in the Domain 1 column. Further information on the selected settings can be found in the WRF V3.8 user's guide (UCAR 2019).

Model parameter	Domain 1	Domain 2	Domain 3
Horizontal grid spacing (km)	27	9	3
Domain size (grid points)	350×310	325×325	295×295
Vertical levels	38		
Model top (hPa)	20		
Time step (s)	120	40	13.333
Cumulus convection	New Tiedtke (option 16)	New Tiedtke	Explicit
Boundary layer	Yonsei University (YSU)		
Ocean mixed layer temperature and depth	Hybrid Coordinate Ocean Model (HYCOM) analysis		
Ocean model	One-dimensional mixed layer		
Ocean–air flux	Donelan C_D + Garratt C_K (option 2)		
Microphysics	WRF single-moment 6-class microphysics scheme (WSM6)		
Land surface	Noah land surface model (NLSM)		
Turbulence	Smagorinsky 2D		
Diffusion	Second-order diffusion		
Scalar advection	Positive definite		
Radiation (longwave and shortwave)	Rapid Radiative Transfer Model for GCMs (RRTMG)		

3. Methodology

a. Generation of the WRF-ADCIRC ensemble

Ensemble forecasts from the European Centre for Medium-Range Weather Forecasts Integrated Forecasting System (ECMWF-IFS) initialized at 1200 UTC 5 September and 1200 UTC 8 September provide initial and lateral boundary conditions for the two sets of WRF atmospheric simulations examined here. Output from each WRF simulation is then used to drive a corresponding ADCIRC simulation, generating an ensemble of storm tide simulations for each initialization time. The ECMWF-IFS data used in this study were obtained from retrospective simulations that are identical to the operational ECMWF-IFS ensembles (L. Magnusson 2018, personal communication). The initial conditions for the ECMWF-IFS ensemble are generated using perturbations from a combination of an ensemble of data assimilations and the initial leading 50 singular vectors in each hemisphere, along with the initial 5 leading singular vectors associated with up to 6 TCs (ECMWF 2017). ECMWF-IFS ensemble data were obtained every 6 h at 0.15° horizontal resolution on 91 vertical model levels and the surface.

Each ECMWF ensemble member (50 perturbed members + control) provides initial conditions and updated boundary conditions for a convection-permitting regional simulation using WRF version 3.8 (Skamarock et al. 2008), with the model configuration shown in Table 1. The WRF simulations are two-way triple-nested with two vortex-following nests centered on Irma (Fig. 1) and 3-km horizontal resolution in the innermost nest. The moving nests remain static during the first 3 h of model integration to allow the vortex to balance. WRF

simulations initialized at 1200 UTC 5 September are terminated at 1200 UTC 14 September. Simulations initialized at 1200 UTC 8 September are terminated at 1200 UTC 13 September.

To evaluate whether the WRF ensembles are suitable for generating storm surge ensembles, we evaluate their performance in simulating Irma's track and intensity evolution. We compare the WRF ensembles to the ECMWF ensembles that provided initial and boundary conditions; the ECMWF is the most skillful global ensemble for North Atlantic TC track prediction

FIG. 1. WRF Model domains used in this study. The two inner nests remain centered on Irma throughout the simulations. The inner nests are at their position at the initialization time of the 1200 UTC 5 Sep simulations.

FIG. 2. Tracks of Hurricane Irma from the WRF ensemble initialized at (a) 1200 UTC 5 Sep and (b) 1200 UTC 8 Sep along with the best track (red). Colored symbols indicate the position of Irma at 1200 UTC 8, 10, and 12 Sep in each ensemble member and the best track. Tracks from the ECMWF ensemble forecasts initialized at (c) 1200 UTC 5 Sep and (d) 1200 UTC 8 Sep are shown for comparison. WRF and ECMWF tracks are shown at 3- and 6-h time steps, respectively. Tracks are terminated at 1200 UTC 14 Sep in (a) and (b) and 1200 UTC 13 Sep in (c) and (d).

(Leonardo and Colle 2017). Although the 5 September WRF ensemble has a slightly smaller track spread than the ECMWF ensemble beyond 72 h (Figs. 2a,c; Table 2), the 51 WRF members vary between landfalling on the Florida panhandle and recurving out to sea. Irma's observed track remains well within the ensemble spread through landfall in Florida, though most simulations do not capture Irma's north-northwestward motion after 10 September. The 8 September WRF ensemble has a slightly smaller track spread than the ECMWF ensemble through 84 h (Figs. 2b,d; Table 3). Irma's observed track falls near the western edge of the WRF track distribution, indicating that the WRF ensemble produces more eastward tracks than the ECMWF ensemble.

For the 5 September simulations, both WRF and ECMWF have a low intensity bias at landfall, but WRF more accurately simulates Irma's intensity evolution

during the 2 days leading up to its observed landfalls (Figs. 3a,c; Table 2). Among 8 September simulations, the WRF ensemble produces smaller intensity errors at landfall compared to the ECMWF ensemble (Figs. 3b,d). Overall, the WRF ensemble generates track and intensity errors that are competitive with the ECMWF ensemble (Tables 2 and 3), and its higher resolution allows it to better represent the intense winds of Irma's inner core. At both initialization times the WRF ensemble produces sufficiently realistic track and intensity evolutions to use the simulations to generate a storm surge ensemble for investigating coastal inundation and the utility of track clustering to describe storm surge hazards.

The storm tide simulations in this study employ ADCIRC version 52 in two-dimensional mode, using the model configuration and procedures similar to those described in Fossell et al. (2017). ADCIRC solves the

TABLE 2. Performance of the WRF and ECMWF ensembles initialized at 1200 UTC 5 Sep. Positive and negative intensity errors indicate ensemble-mean maximum sustained winds that are stronger and weaker, respectively, than the best track value. Ensemble spread is calculated as the average distance between ensemble members and the ensemble mean at each time step. ECMWF track and intensity values are obtained from the UCAR TIGGE database (NCAR 2020).

Forecast hour	Time	WRF ensemble-mean track error (km)	ECMWF ensemble-mean track error (km)	WRF ensemble-mean intensity error (m s^{-1})	ECMWF ensemble-mean intensity error (m s^{-1})	WRF ensemble track spread (km)	ECMWF ensemble track spread (km)	WRF ensemble intensity spread (m s^{-1})	ECMWF ensemble intensity spread (m s^{-1})
24	1200 UTC 6 Sep	30	55	-27.2	-39.8	52	51	3.0	2.5
48	1200 UTC 7 Sep	32	80	-16.5	-32.7	87	90	2.6	2.6
72	1200 UTC 8 Sep	81	145	-10.1	-29.3	115	113	4.1	2.5
96	1200 UTC 9 Sep	132	162	-4.6	-15.7	141	150	5.4	4.3
120	1200 UTC 10 Sep	165	134	-14.4	-17.8	212	240	3.8	3.9
144	1200 UTC 11 Sep	258	309	16.9	15.2	310	350	4.3	2.7
168	1200 UTC 12 Sep	643	640	19.3	22.1	374	409	7.4	6.6

TABLE 3. As in Table 2, but for WRF and ECMWF simulations initialized at 1200 UTC 8 Sep.

Forecast hour	Time	WRF ensemble-mean track error (km)	ECMWF ensemble-mean track error (km)	WRF ensemble-mean intensity error (m s^{-1})	ECMWF ensemble-mean intensity error (m s^{-1})	WRF ensemble track spread (km)	ECMWF ensemble track spread (km)	WRF ensemble intensity spread (m s^{-1})	ECMWF ensemble intensity spread (m s^{-1})
24	1200 UTC 9 Sep	9	32	-9.5	-16.4	28	31	4.6	2.0
48	1200 UTC 10 Sep	84	9	-13.4	-16.6	60	68	3.2	2.3
72	1200 UTC 11 Sep	112	79	13.8	8.4	88	111	2.5	2.1
96	1200 UTC 12 Sep	214	232	5.9	3.8	144	114	1.7	3.1

FIG. 3. Intensity evolution of Hurricane Irma from the WRF ensemble initialized at (a) 1200 UTC 5 Sep and (b) 1200 UTC 8 Sep. The best track intensity evolution of Irma is in red. Intensity evolution from ECMWF ensemble forecasts initialized at (c) 1200 UTC 5 Sep and (d) 1200 UTC 8 Sep are shown for comparison. WRF and ECMWF data are shown at 3- and 6-h time steps, respectively. Pink symbols indicate the intensity of Irma at the time step before its first U.S. landfall. Crossing the Florida Keys is considered a landfall even if Irma's center did not pass over an island.

shallow-water equations for a depth-integrated barotropic hydrodynamic circulation, producing domain-wide water levels at each 2-s model time step. ADCIRC uses a variable-resolution, finite-element mesh. The mesh used in this study, as in [Fossell et al. \(2017\)](#), was built and validated by Riverside and AECOM ([Riverside Technology and AECOM 2015](#)) and provided by the National Ocean Service ([Feyen et al. 2015](#)). The validation of this ADCIRC configuration included evaluation of hindcast performance for 7 major tropical cyclones in the southern and southeast United States, for use by NOAA for tropical cyclones near and after landfall ([Riverside Technology and AECOM 2015](#); [Feyen et al. 2015](#)).

The ADCIRC grid contains nearly three million nodes spanning from Texas to Maine and inland to the 10-m elevation contour. Grid resolution is coarse far from land and as fine as 200 m in coastal regions; we selected this mesh to balance resolving smaller-scale inundation features caused by heterogeneities in topography and bathymetry with the ability to simulate surge inundation from TCs making landfall along a wide swath of the U.S. coastline. The mesh employed in this study uses spatially varying nodal attributes that allow for individual nodes to be assigned attribute values

(e.g., bottom friction, wind drag, etc.). The Manning's N values used for bottom friction range from 0.02 (open water) to 0.20 (Palustrine forested wetland), and account for the built environment by classifying intensity of development, resulting in Manning's N values of 0.07 (low intensity developed) to 0.120 (medium and high intensity developed). Further details can be found in [Riverside Technology and AECOM \(2015\)](#).

Tides in ADCIRC are generated by applying tidal forcing at Atlantic Ocean grid boundaries using 13 tidal constituents from the TPXO7.2 tidal atlas ([Riverside Technology and AECOM 2015](#) and references within) beginning on 21 August. Tidal forcing is increased during the first 10 days of ADCIRC integration using a hyperbolic tangent ramping function to ensure smooth spinup. As will be demonstrated in [section 4](#), tides contribute substantially to the inundation experienced from northeast Florida north along the Atlantic Coast.

Meteorological forcing from each WRF simulation is applied by interpolating hourly 10-m U , V , and mean sea level pressure (MSLP) from the 27- and 3-km WRF nests onto regular latitude-longitude grids; these data are then interpolated to each ADCIRC time step and used to drive water levels in the corresponding ADCIRC simulation. Wind and MSLP forcing are applied to

FIG. 4. Locations corresponding to ADCIRC nodes where maximum water levels and inundation probabilities are evaluated (Tables 6 and 7). The division between the southwest (SW) and northeast (NE) regions is indicated. The SW and NE region are not limited to the area shown; the ADCIRC grid covers the entire U.S. coastline from Texas to Maine. Colors indicate ocean depth and topographical height.

ADCIRC from the beginning of each WRF simulation, using a hyperbolic tangent ramping function during the first 24 h to ensure that water levels adjust smoothly. These procedures generate an ADCIRC ensemble for each WRF ensemble initialization time.² Water levels output by these ADCIRC simulations do not include the effects of wave setup. Although wave setup would ideally be included in an operational model, it is not currently employed in the NHC's P-Surge forecasts, and for simplicity we exclude it here.

The performance of the WRF-ADCIRC ensemble is evaluated by comparing the simulated and observed maximum water levels at eight ADCIRC nodes corresponding to tide gauge locations (Fig. 4). For the 5 September ensemble (Table 4), the observed maximum water level falls within the range of simulated water levels at all locations. For the 8 September ensemble (Table 5), the ensemble spread encompasses the observed maximum water level at 6 of 8 locations; the large spread also indicates the substantial forecast uncertainty even at this time range.

At Vaca Key and Fernandina Beach, where the observed maximum water level falls below the 8 September simulated range, the eastward bias of the WRF ensemble

may contribute to the higher-than-observed water levels. Most ensemble members track Irma's core closer to Fernandina Beach than its observed track (Fig. 2b), and likely generate a stronger onshore flow. At Vaca Key, the WRF ensemble's eastward bias likely causes a stronger-than-observed northerly flow, enhancing water levels at the bayside station. Additionally, the relatively coarse resolution of the ADCIRC grid may cause bathymetry and topography to differ between the observation locations and the ADCIRC nodes used to compare water levels. A finer ADCIRC mesh, along with bottom and surface friction refinements, may improve the ADCIRC component of this ensemble.

b. Regression mixture-model clustering of the WRF ensembles

Next, the tracks of Irma among the WRF ensembles are clustered using regression mixture models. For both initialization times, clustering is performed using the 72-h track segments between 1200 UTC 9 September and 1200 UTC 12 September to focus on variations in Irma's track near landfall. Irma's position in each track is provided at 3-h intervals. In regression mixture-model clustering, each ensemble track is probabilistically assigned to each of K models (clusters) by its fit to the latitude and longitude central trajectories, error covariance matrix, and weight (population) defining each model. This assignment is performed through an expectation-maximization (E-M) algorithm [see Gaffney et al. (2007) and Camargo et al. (2007) for details] that (i) calculates the parameters of each cluster (polynomial coefficients, error covariance matrix, and cluster size), and (ii) calculates each track's cluster membership probabilities to maximize the likelihood (fit) of the mixture model. Because the E-M algorithm may converge on a local, rather than global likelihood maximum in parameter space, this process is repeated 500 times using random initial cluster membership probabilities. The solution with the highest likelihood (best fit) is then selected.

Regression mixture-model clustering requires specifying the polynomial order and number of clusters prior to clustering. The optimal polynomial order and number of clusters are determined here by comparing results produced by combinations of polynomial order first-sixth and number of clusters 2–9. Bayesian information criterion (BIC) and root mean squared displacement (RMSD), along with visual inspection, provide guidance for selecting the optimal cluster characteristics among the candidate partitions. BIC is calculated from the log-likelihood, with a larger penalty incurred for a more complex mixture-model (higher polynomial order and/or more clusters); it favors solutions that

² For the 5 September ensemble, ADCIRC became unstable simulating members 33 and 39. Although these members were included in track clustering, and both assigned to the green (F) cluster, ADCIRC output from these members are not included in the results.

TABLE 4. For the WRF-ADCIRC ensemble initialized at 1200 UTC 5 Sep: comparison between maximum observed and simulated water levels at the ADCIRC node corresponding to each station in Fig. 4.

Location	Station ID	Region	Observed maximum water level (m)	Ensemble mean maximum water level (m)	Range of maximum water levels in ensemble (m)	Percent of ensemble members ≥ 1 m above MHHW
McKay Bay Entrance, FL	8726667	SW	0.53	0.15	−0.17–1.90	6
Naples, FL	8725110	SW	1.29	0.51	0.13–1.46	14
Key West, FL	8724580	SW	0.82	0.54	0.39–1.13	6
Vaca Key, FL	8723970	SW	0.78	1.02	0.38–2.56	41
Virginia Key, FL	8723214	SW	1.11	0.95	0.26–2.77	35
Fernandina Beach, FL	8720030	NE	1.10	1.37	0.66–2.49	80
Fort Pulaski, GA	8670870	NE	1.44	0.80	0.37–1.94	27
Charleston, SC	8665530	NE	1.27	0.77	0.33–1.93	27

balance goodness-of-fit with parsimony. RMSD is calculated from the displacement between the TC position in each simulation and the position given by the central polynomial trajectories of the cluster containing that simulation. For RMSD calculations and throughout this paper, each ensemble member is deterministically assigned to the cluster with its highest membership probability. For both BIC and RMSD, lower values indicate that a clustering solution has a better fit to the set of ensemble tracks.

Results from Kuruppumullage Don et al. (2016) and Kowaleski and Evans (2016) indicate that selecting the optimal polynomial order is more straightforward than selecting the optimal number of clusters, so we select polynomial order first. For the 5 September ensemble, BIC values for solutions with 2–9 clusters undergo a mean increase (degradation) beyond second order, while RMSD values decrease (improve) until third order and increase for higher orders (Fig. 5a). For the 8 September ensemble, BIC decreases through third order, while RMSD improvement drops off dramatically beyond third order (Fig. 5b). Therefore, a third-order polynomial appears reasonable for both 5 and 8 September ensemble simulations.

For both sets of ensemble simulations, neither BIC nor RMSD gives a clear indication of the optimal number of clusters (Figs. 5c,d). For the ensemble initialized at 1200 UTC 5 September, the increase in RMSD between the 7-cluster and 8-cluster solutions suggests a 7-cluster solution. However, a comparison between the 6- and 7-cluster solutions (not shown) shows that cluster memberships among the two solutions are nearly identical, except for the splitting of one of the eastern clusters in the 7-cluster solution. Thus, a 6-cluster solution is selected. For the ensemble initialized at 1200 UTC 8 September, visual inspection of the clustering results indicates that the 5-cluster solution produces a reasonable partition.

c. Analysis of the WRF-ADCIRC ensemble output

Results from the ADCIRC simulations are examined in terms of inundation, defined here as the water depth that occurs over normally dry ground (grid points above MHHW). Not including points below MHHW in the metrics allows us to interpret results in terms of flooding that occurs in areas that are not normally inundated by tide alone. An additional threshold of 1-m inundation is applied to focus on significant flooding that would warrant a watch or warning in a real-world setting.³

The inundation defined above is analyzed using two regionally oriented metrics. The first metric is inundation volume, computed as in Fossell et al. (2017) by multiplying inundation depth at each node by the cell area and summing over nodes with at least 1-m inundation, excluding the leftmost and rightmost 2.5% to filter out inundation unrelated to Irma. We selected the inundation volume metric because it provides an integrated metric of the magnitude of flooding in an affected region, including the area inundated and the inundation depth; both are relevant for decision making. We examine inundation volume in time series. The second metric is the probability of the maximum inundation reaching the 1-m threshold at different locations, computed as the percentage of members within the full ensemble or a cluster that attain that threshold at each ADCIRC node at any time. These probabilities are depicted on maps similar to the NWS's probabilistic storm surge exceedance maps, but only for inundation over normally dry land.

Irma produced significant inundation in two regions: 1) south Florida and 2) from northeast Florida through South Carolina. To evaluate inundation in each of these

³ The current National Hurricane Center operational threshold for storm surge watches and warnings is 3-ft (0.91 m) inundation above MHHW.

TABLE 5. As in Table 4, but for simulations initialized at 1200 UTC 8 Sep.

Location	Station ID	Region	Observed maximum water level (m)	Ensemble mean maximum water level (m)	Range of maximum water levels in ensemble (m)	Percent of ensemble members ≥ 1 m above MHHW
McKay Bay Entrance, FL	8726667	SW	0.53	0.06	-0.14–0.57	0
Naples, FL	8725110	SW	1.29	0.90	0.20–2.04	35
Key West, FL	8724580	SW	0.82	0.67	0.43–1.30	10
Vaca Key, FL	8723970	SW	0.78	1.46	0.90–2.37	90
Virginia Key, FL	8723214	SW	1.11	2.08	0.70–2.92	96
Fernandina Beach, FL	8720030	NE	1.10	2.06	1.53–2.75	100
Fort Pulaski, GA	8670870	NE	1.44	1.82	1.06–2.61	100
Charleston, SC	8665530	NE	1.27	1.39	0.84–1.89	94

regions separately, we partition the volume-integrated inundation metric into “southwest” and “northeast” regions (Fig. 4). The dividing point between regions is selected near Fort Lauderdale, Florida, because this location does not experience inundation exceeding 1 m from any of the simulations initialized on 5 September. We also examine inundation probability maps within each inundation region.

To complement the two metrics described above, we also examine inundation probabilities across the entire ensemble and within each cluster at select locations (Fig. 4). These point-based results provide a more granular

perspective, demonstrating how large- and small-scale track variations influence simulated inundation at specific locations.

4. Results

a. Simulations initialized at 1200 UTC 5 September

First, to provide context for interpreting the surge ensemble results, we provide an overview of the characteristics of the 51-member WRF ensemble initialized at 1200 UTC 5 September. The WRF ensemble tracks and evolutions of 34- and 64-kt ($1 \text{ kt} \approx 0.51 \text{ m s}^{-1}$) wind

FIG. 5. Mean decrease in Bayesian information criterion (BIC) and root mean squared displacement (RMSD) with each additional polynomial order for solutions using 2–9 clusters from the WRF ensemble initialized at (a) 1200 UTC 5 Sep and (b) 1200 UTC 8 Sep; decrease in BIC and RMSD with each additional cluster using third-order polynomials from simulations initialized at (c) 1200 UTC 5 Sep and (d) 1200 UTC 8 Sep.

FIG. 6. For the WRF ensemble initialized at 1200 UTC 5 Sep: (a) tracks of the members composing each cluster, and (b) cluster-mean tracks. The individual members in (a) and mean positions of Irma in (b) at 1200 UTC 10 Sep (the time of its landfall in the Florida Keys) are indicated by colored dots. Clusters are labeled A–F for reference in the text.

radii are depicted in Figs. 6 and 7, respectively, color coded by cluster. As shown in Fig. 6a, 5 days before landfall, the WRF ensemble indicates that Irma will likely affect areas along the southeast U.S. coast, but

where in the southeast United States is highly uncertain. The clusters represent possible track scenarios that vary from Irma landfalling along the northern Gulf Coast to tracking north toward the Carolinas (Fig. 6b).

FIG. 7. For the WRF ensemble initialized at 1200 UTC 5 Sep: evolution of quadrant-averaged 34- and 64-kt wind radii for each ensemble member, grouped by cluster. Black dots and triangles indicate the average 34- and 64-kt wind radii, respectively, at the time of Irma's first U.S. landfall in each simulation. Crossing the Florida Keys is considered a landfall, even if Irma's center did not pass over an island. The mean and standard deviation of landfall intensity (kt) among members of each cluster are shown in the top left of each panel.

FIG. 8. For the WRF-ADCIRC ensemble initialized at 1200 UTC 5 Sep: full-ensemble inundation volume time series in the (a) southwest and (b) northeast regions, along with full-ensemble probabilities of inundation of 1 m or greater above MHHW for coastal areas within the (c) southwest and (d) northeast regions. Black stars mark the time and volume of maximum inundation produced by each ensemble member.

Associated with these track variations, Fig. 7 shows that Irma's 34- and 64-kt wind radii vary significantly around the time that Irma makes landfall in different simulations. However, before Irma approaches the mainland United States, the storm's size is relatively similar within each cluster, and as the Irma nears the southeastern United States, each of the WRF ensemble members has a 34-kt wind radius exceeding 150 n mi (278 km), when averaged across Irma's four quadrants. Although Irma attains a maximum intensity of at least category 4 (≥ 113 kt) in nearly all simulations, its intensity at first U.S. landfall averages between 71.7 and 78.1 kt (36.9 and 40.2 m s^{-1}) in each cluster, except for the magenta cluster, which averages 89.3 kt (45.9 m s^{-1} ; Fig. 6). Magenta members likely have higher landfall intensities because Irma interacts less with Cuba, and landfalls in south Florida, rather than farther north. Despite Irma's relatively low landfall intensities across

the ensemble, the storm has an expansive circulation capable of generating substantial storm surge inundation along the U.S. coast.

Next, we examine the characteristics of the surge inundation generated by the full WRF-ADCIRC ensemble initialized on 5 September (Fig. 8). In the southwest region, Figs. 8a and 8c indicate that there is significant uncertainty in how much inundation this region will experience, along with when and where the inundation will occur. This is consistent with what one would expect given the WRF ensemble's large track spread. There is a small region in far-southeast Miami-Dade County and the upper Florida Keys where the WRF-ADCIRC ensemble predicts a 45%–65% probability of 1-m inundation, but otherwise the ensemble indicates a 5%–35% chance of 1-m inundation over most of coastal south Florida (Fig. 8c). The WRF-ADCIRC ensemble also indicates a < 10% chance of inundation greater than 1 m

in west Florida, near Tampa (not shown). In other words, at this lead time the ensemble indicates a widespread, but relatively low, risk of coastal flooding across much of south and west Florida. This uncertainty in which locations will experience significant surge at this lead time is consistent with prior research by Fossell et al. (2017), but it creates challenges for initiating evacuation planning in this highly populated region.

In the northeast region, Fig. 8b illustrates that inundation is closely connected to the tidal cycle, with many ensemble members generating substantial inundation across multiple high tides. As in the southwest region, the WRF-ADCIRC ensemble indicates that there is significant uncertainty in inundation volume and timing in this region. However, far more ensemble members generate substantial inundation in the northeast region compared to the southwest; for example, 29 ensemble members produce maximum inundation volume greater than 1 km^3 in the northeast region, whereas only 9 produce maximum inundation volume greater than 0.5 km^3 in the southwest. The ensemble's prediction that the northeast region has a greater risk of significant surge-induced flooding is reinforced by the inundation probability map in Fig. 8d, which shows that many locations have inundation probabilities exceeding 95%. Thus, 5 days before landfall, the WRF-ADCIRC ensemble suggests high confidence that there will be surge-induced flooding along a stretch of the northeast Florida and Georgia coasts. It cannot yet be predicted when this flooding will occur (Fig. 8b), or how far inland or to the north or south it will extend. However, given that operational surge forecasts at this lead time were not available in real time, this type of information (if reliable) has potential to provide substantially more time for evacuating vulnerable populations.

Finally, we analyze the WRF-ADCIRC results by cluster, to explore whether clustering has potential to provide additional predictive information beyond the full ensemble. In the southwest region, Fig. 9 depicts inundation volume time series for each of the six track clusters, and Fig. 10 depicts the corresponding maps of 1-m inundation probability. To facilitate quantitative comparison, Table 6 presents all-ensemble and cluster inundation probabilities from Fig. 10 at select ADCIRC nodes containing tide gauges.

Whereas Fig. 8a indicates a wide range of inundation volume and timing in the southwest region, Fig. 9 helps distinguish between inundation situations within the ensemble: the red cluster is associated with potential for significant inundation on 12 September (Fig. 9a), the blue and magenta clusters with potential for significant inundation on 10 or 11 September (Figs. 9b,c), and the cyan, green, and orange clusters with lower inundation potential (except for one orange member; Figs. 9d–f).

Figure 10 and Table 6 indicate that these different situations are associated with greater risk in different areas, within the broad coastal region at risk depicted in Fig. 8c. Red cluster members produce little inundation threat to south Florida (Fig. 10a); the hazard is concentrated farther north in the Florida panhandle (not shown), and along the west Florida coast, near Tampa (McKay Bay Entrance in Table 6). In the blue cluster, inundation probabilities are highest in southwest Florida (Fig. 10b; Naples in Table 6), peaking from Cape Romano through Everglades City and the unpopulated Monroe County coast. In the magenta cluster, inundation is most likely in the middle–upper Keys and southeast Miami–Dade County (Fig. 10c; Vaca Key and Virginia Key in Table 6). In the cyan cluster, inundation is most likely in areas similar to the magenta cluster (Fig. 10d), but less widespread, consistent with the lower inundation volumes in Fig. 9d. In the orange and green clusters, inundation is most likely in similar areas of southeast Florida as magenta and cyan, but with lower probabilities (Figs. 10e,f).

Some aspects of these results are consistent with inundation differences that would be expected from the different storm tracks in different clusters (Fig. 6). For example, in most of the green and orange cluster members (E, F; Fig. 9), Irma remains east of the southwest region and produces a primarily offshore flow, producing less inundation than many of the red, blue and magenta cluster members (A, B, C), in which Irma landfalls within the southwest region and produces a substantial onshore flow. However, other aspects of the results suggest that the WRF-ADCIRC ensemble, combined with clustering, may help to distinguish surge inundation scenarios beyond what a qualitative interpretation of anticipated surge inundation based on WRF ensemble storm tracks would suggest. For example, although the magenta (C) and cyan (D) clusters have similar cluster-mean landfall locations (Fig. 6b), and comparable wind radii (Figs. 7c,d), the magenta cluster is associated with greater surge risk (Figs. 9c,d and 10c,d). In addition, the McKay Bay Entrance probabilities in Table 6 suggest that Irma poses a greater flooding threat to the Tampa region when it tracks well to the west, as in red cluster members, rather than closer to the region, as in blue members (Fig. 6a). Inundation probabilities from the orange and green clusters at the bayside Vaca Key station suggest the risk of bayside flooding in the Keys if Irma tracks to the east.

These results indicate that clustering the TC–surge ensemble has potential to provide information about the hurricane scenarios likely to generate inundation in different areas of south and west Florida, at different times. Such information may be useful, especially compared to the limited information available in Figs. 8a,c.

FIG. 9. Time series of inundation volume in the southwest region for members of each cluster, from the WRF-ADCIRC ensemble initialized at 1200 UTC 5 Sep. Black stars mark the time and volume of maximum inundation in each ensemble member.

FIG. 10. Probabilities of inundation of at least 1 m above MHHW in south Florida, calculated within each cluster of the WRF-ADCIRC ensemble initialized at 1200 UTC 5 Sep. The colored dots in the bottom right of each panel correspond to the track clusters shown in Fig. 6.

For example, these results tell forecasters that if Irma's track uncertainty could be reduced to the spread of one of the clusters, information about the inundation risk in different areas would improve markedly. Alternatively, if subsequent forecasts begin to shift toward one of the track clusters, the likelihood of that surge scenario would increase. However, intracluster spreads are substantial, with different ensemble members within a single cluster yielding quite different inundation outcomes.

This is especially evident in the red (A), blue (B), and magenta (C) clusters (Figs. 9a–c). Although an improved clustering technique, more focused on surge, may be able to elucidate more distinct scenarios, at this lead time there are still significant uncertainties in surge predictions for the southwest region, even within clusters.

In the northeast region, Figs. 11 and 12 and Table 6 indicate that clustering again helps distinguish different inundation situations within the ensemble. Here, however,

TABLE 6. For the WRF-ADCIRC ensemble initialized at 1200 UTC 5 Sep: probability of 1-m inundation or greater at select ADCIRC nodes containing tide gauges across the full ensemble and among members of each cluster. Maximum inundation values for the full ensemble and each cluster are in parentheses. Station locations are shown in Fig. 4. The two 5 Sep green cluster members for which ADCIRC became unstable are not included.

Location	Station ID	Region	Percent of members with inundation $\geq$ 1m (maximum inundation; m)						
			Full ensemble	Cluster A red	Cluster B blue	Cluster C magenta	Cluster D cyan	Cluster E orange	Cluster F green
McKay Bay Entrance, FL	8726667	SW	6 (1.90)	75 (1.90)	0 (0.83)	0 (0.01)	0 (−0.06)	0 (0.21)	0 (0.11)
Naples, FL	8725110	SW	14 (1.46)	25 (1.17)	67 (1.46)	0 (0.57)	0 (0.70)	0 (0.46)	0 (0.34)
Key West, FL	8724580	SW	6 (1.13)	0 (0.45)	33 (1.13)	0 (0.53)	0 (0.59)	0 (0.63)	0 (0.50)
Vaca Key, FL	8723970	SW	41 (2.56)	0 (0.48)	11 (1.06)	100 (2.56)	57 (1.67)	36 (2.22)	13 (1.43)
Virginia Key, FL	8723214	SW	35 (2.77)	0 (0.49)	56 (1.34)	70 (2.67)	57 (2.3)	9 (2.77)	0 (0.56)
Fernandina Beach, FL	8720030	NE	80 (2.49)	0 (0.95)	100 (2.14)	100 (2.49)	71 (1.4)	91 (2.2)	63 (1.22)
Fort Pulaski, GA	8670870	NE	27 (1.94)	25 (1.57)	56 (1.38)	30 (1.42)	0 (0.48)	36 (1.94)	0 (0.62)
Charleston, SC	8665530	NE	27 (1.93)	0 (0.67)	22 (1.12)	40 (1.35)	0 (0.88)	64 (1.93)	0 (0.57)

unlike in the southwest region, all of the clusters generate flooding along a similar stretch of the northeast Florida–Georgia–South Carolina coastline. The red (A) and orange (E) clusters are associated with high potential for inundation in this area on 12 September and 11 September, respectively. The blue (B) and magenta (C) clusters predict inundation over the most extensive area of the northeast Florida–Georgia coast, on 12 and 11 September, respectively. The cyan (D) and green (F) clusters indicate a lower risk of widespread inundation, but still a high probability that certain areas along the northeast Florida–Georgia coast will be flooded. Inundation probabilities of 100% are most widespread for the blue (B), magenta (C), and orange (E) clusters, in which Irma tracks closest to this stretch of coastline. However, inundation is likely in this region even for the red, cyan, and green clusters (Figs. 12a,d,f), in which most ensemble members track Irma’s center far from this area (Fig. 6a). This demonstrates the importance of Irma’s outer circulation, and the storm’s interaction with the synoptic flow, in generating water level rises in this region.

While the full ADCIRC ensemble predicts a substantial risk of the northeast region flooding 5 days before Irma’s landfall, results from ADCIRC nodes collocated with northeast region gauges (Table 6), show how clustering is helpful in identifying the track scenarios that generate an inundation hazard at individual locations. At Fernandina Beach, near the Florida–Georgia border, all members of the magenta and blue clusters, and a majority of cyan, orange, and green, generate at least a 1-m inundation. In contrast, at Fort Pulaski and Charleston, no members of the cyan or green clusters generate inundation reaching 1 m, and probabilities do not exceed 56% and 64%, respectively, among any other clusters. These results suggest that inundation near the Georgia–South Carolina border and along the South Carolina coast requires a closer approach from Irma than inundation on

the northeast Florida coast. Thus, beyond the widespread regional inundation predicted by the dynamical ensemble 5 days before landfall, clustering shows potential for helping identify the large-scale inundation scenarios that may occur within this region.

b. Simulations initialized at 1200 UTC 8 September

The five track clusters generated from WRF simulations initialized at 1200 UTC 8 September (Fig. 13) are much more tightly grouped near landfall than those produced from simulations initialized 3 days earlier, indicating a marked increase in track forecast confidence. Two days before Irma’s landfall, all 51 members predict a landfall somewhere in south Florida. The decrease in ensemble spread near landfall between 5 and 8 September is also reflected in wind radii time series (Fig. 14); every 8 September member shows Irma landfalling with quadrant-averaged 34- and 64-kt wind radii of at least 191 n mi (354 km) and 45 n mi (83 km), respectively. Irma’s intensity at its first U.S. landfall in each cluster averages between 90.5 and 103.7 kt (46.6 and 53.3 m s^{−1}).

Consistent with this smaller atmospheric ensemble spread, the 8 September WRF-ADCIRC ensemble also shows less spread in surge-induced inundation than the 5 September ensemble, in both the southwest and northeast regions (Figs. 15a,b). In the southwest region, Fig. 15a indicates a higher likelihood of significant inundation at this lead time (27 members produce inundation volumes $\geq$ 0.5 km³, compared to 9 from 5 September), with the time range narrowed to 10–11 September. Ensemble-wide inundation probabilities exceed 65% along much of Florida’s southeast and southwest coasts (Fig. 15c), with a lower level of risk extending north to Naples (Table 7). As indicated by the McKay Bay Entrance probabilities in Table 7, the ensemble now indicates little to no inundation risk farther north along the west Florida coast.

FIG. 11. As in Fig. 9, but for inundation volume in the northeast region.

In the northeast region, the WRF-ADCIRC ensemble indicates high confidence in a widespread inundation event along the northeast Florida, Georgia, and South Carolina coasts on 11 September. All 51 ensemble members produce a maximum inundation volume

exceeding 1.5 km^3 (Fig. 15b), and inundation probabilities exceed 95% over much of the northeast coastal region (Fig. 15d, Table 7).

The WRF-ADCIRC ensemble clustering results for the southwest region are shown in Figs. 16 and 17,

FIG. 12. As in Fig. 10, but for the northeast Florida, Georgia, and South Carolina coasts.

along with Table 7. Mean maximum inundation volume is largest (1.02 km^3) in the red cluster (Fig. 16a), which has the westernmost tracks, and smallest (0.34 km^3) in the orange cluster (Fig. 16e), which has the easternmost tracks. Consistent with this, the inundation risk is greatest in southwest Florida, including Naples, in the red cluster (Fig. 17a; Table 7), while southeast Florida shows a similar risk across all clusters. The blue, magenta, and cyan clusters, which make landfall in a similar area near the tip of Florida, do not seem to represent meaningfully distinct surge inundation scenarios. Although the

dynamical surge ensemble indicates that some inundation will likely occur in southeast Florida regardless of storm track, at this 2-day lead time the clustering technique used here does not provide as much additional information in the southwest region as it did at a 5-day lead time, and the coastal extent and inland reach of the inundation still depends on the details of Irma's track and evolution.

In the northeast region, the red (A) and cyan (D) clusters produce a single unambiguous inundation maximum around 1800 UTC 11 September, with a much lower maximum at the high tide 12 h earlier

FIG. 13. As in Fig. 6, but for the WRF ensemble initialized at 1200 UTC 8 Sep. Clusters are labeled A–E for reference in the text.

(Figs. 18a,d). In contrast, the blue (B), magenta (C), and orange (E) clusters generate two comparable inundation maxima around 0600 and 1800 UTC 11 September, with both maxima generally lower than the red and cyan maxima (Figs. 18b,c,e). The magenta and blue clusters also show persistently high inundation between the two maxima, indicating that substantial flooding persists for more than a full tidal cycle. In the red and cyan clusters, Irma's peak wind forcing occurs near high tide late on 11 September, whereas it occurs

between high tides in the magenta and blue clusters. The orange cluster includes a mix of the two scenarios. All of the clusters indicate a high likelihood of inundation across much of the coastal region, with minor differences in the likely coastal and inland extent of flooding (Fig. 19; Table 7). Together, the results from the 5 and 8 September simulations show how the WRF-ADCIRC ensemble, combined with track clustering, can help to elucidate areas where surge is anomalously predictable at longer lead times, as well as potentially

FIG. 14. As in Fig. 7, but for the WRF ensemble initialized at 1200 UTC 8 Sep.

FIG. 15. As in Fig. 8, but for the WRF-ADCIRC ensemble initialized at 1200 UTC 8 Sep.

highlight different scenarios of inundation timing and duration.

5. Summary and conclusions

Storm surge inundation from landfalling tropical cyclones presents a major threat to lives and property.

Improving the surge risk information that forecasters can provide at longer lead times can give decision makers greater time for planning and implementing protective actions. Dynamical ensembles have the potential to capture variations in TC evolution, including track, that are relevant for storm surge, but that may not be fully represented by current operational surge

TABLE 7. As in Table 6, for the WRF-ADCIRC ensemble initialized at 1200 UTC 8 Sep.

Location	Station ID	Region	Percent of members with inundation ≥ 1 m (maximum inundation; m)					
			Full ensemble	Cluster A red	Cluster B blue	Cluster C magenta	Cluster D cyan	Cluster E orange
McKay Bay Entrance, FL	8726667	SW	0 (0.57)	0 (0.42)	0 (0.57)	0 (0.18)	0 (0.04)	0 (0.10)
Naples, FL	8725110	SW	35 (2.04)	82 (2.04)	57 (1.74)	20 (1.22)	18 (1.14)	8 (1.05)
Key West, FL	8724580	SW	10 (1.30)	27 (1.30)	14 (1.01)	0 (0.61)	9 (1.08)	0 (0.68)
Vaca Key, FL	8723970	SW	90 (2.37)	82 (1.69)	100 (1.70)	80 (2.22)	100 (2.0)	92 (2.37)
Virginia Key, FL	8723214	SW	96 (2.92)	100 (2.3)	100 (2.92)	100 (2.73)	100 (2.73)	83 (2.84)
Fernandina Beach, FL	8720030	NE	100 (2.75)	100 (2.31)	100 (2.75)	100 (2.58)	100 (2.18)	100 (2.63)
Fort Pulaski, GA	8670870	NE	100 (2.61)	100 (2.61)	100 (2.16)	100 (1.87)	100 (2.31)	100 (2.44)
Charleston, SC	8665530	NE	94 (1.89)	73 (1.48)	100 (1.45)	100 (1.53)	100 (1.81)	100 (1.89)

FIG. 16. As in Fig. 9, but for the WRF-ADCIRC ensemble initialized at 1200 UTC 8 Sep.

modeling strategies. Here, we employ WRF-ADCIRC simulations of Hurricane Irma to explore whether this type of joint ensemble can provide useful information about TC-related inundation hazards, and whether clustering can elucidate distinct inundation scenarios. The goal is not to demonstrate the predictive skill of this

particular ensemble implementation, but rather to investigate the utility of methodologies that, with further development, could benefit TC hazard forecasting.

Ensembles of 51 WRF simulations of Irma are initialized at 1200 UTC 5 September and 1200 UTC 8 September, approximately 5 and 2 days before Irma's

FIG. 17. As in Fig. 10, but for the WRF-ADCIRC ensemble initialized at 1200 UTC 8 Sep.

observed Florida landfalls. ECMWF ensemble forecasts provide initial and lateral boundary conditions for the WRF simulations. Wind and sea level pressure output from each WRF simulation drive a corresponding ADCIRC simulation that computes storm tide, generating an ensemble of inundation simulations.

Regression mixture-model clustering of Irma's track is used to partition each WRF-ADCIRC ensemble into clusters. Within the whole ensemble, and across individual clusters, integrated inundation volume, inundation

probability maps, and inundation probabilities at specific locations are analyzed. Because Irma generated significant inundation along two distinct coastal stretches, results are examined separately for south Florida and northeast Florida through South Carolina.

For 5 September simulations, regional-scale forecast confidence is low in the southwest region, with large variation in maximum inundation volume and inundation timing across the WRF-ADCIRC ensemble. Clustering helps extract additional information from the ensemble

FIG. 18. As in Fig. 11, but for the WRF-ADCIRC ensemble initialized at 1200 UTC 8 Sep.

by highlighting different scenarios in inundation timing and areas likely to be inundated. The inundation probability maps and table identify specific areas in the southwest region, such as the middle–upper Florida Keys and southeast Miami–Dade County, that have near-100% inundation probabilities in specific clusters.

Cluster analysis also highlights that the greatest inundation threat to the Tampa Bay region comes from scenarios in which Irma passes well to the west, rather than when it tracks near or over the bay, illustrating the potential for the dynamical WRF–surge ensemble at this lead time to provide additional information

FIG. 19. As in Fig. 12, but for the WRF-ADCIRC ensemble initialized at 1200 UTC 8 Sep.

beyond that provided by the TC evolutions in the WRF ensemble alone.

In the northeast region, the 5 September simulations show that substantial regional-scale inundation is highly likely. Much of the Georgia coast has near-100% inundation probabilities across the full ensemble. Thus, the WRF-ADCIRC ensemble highlights the high risk to this coastal region a full 5 days before landfall, 36 h earlier than the first operational surge forecasts for Irma (NHC 2020), and well beyond the 1–2-day predictability limit for location-specific surge inundation identified by

Fossell et al. (2017). Track clustering reveals that every cluster has $\geq 75\%$ inundation probabilities along a lengthy coastal stretch, emphasizing that the high surge risk exists across widely disparate track scenarios. Even if Irma's core misses the northeast region, the ensemble indicates that the storm will likely generate substantial inundation there. This type of information, if available in real time, could provide additional time for protective actions.

Across the 8 September ensemble, far more simulations produce a large southwest region inundation

volume compared to 5 September, as every simulation landfalls in south Florida. Clustering distinguishes different southwest Florida inundation risk scenarios based on Irma's track, while showing that the hazard in Miami–Dade County and the middle–upper Keys is high across Irma's potential tracks. Thus, despite the smaller track spread on 8 September, track clustering is still able to provide some additional information about surge risk in different areas of south and west Florida, though not as clearly as with the 5 September ensemble.

In the northeast region, every 8 September simulation generates a large maximum inundation volume, and every cluster produces an extensive area of 100% inundation probabilities. Thus, the WRF-ADCIRC ensemble predicts a widespread regional flooding event with high confidence 2 days before landfall. In addition, clustering highlights how differences in landfall timing produce different scenarios of maximum inundation volume and inundation duration. Two clusters generate a maximum surge forcing near high tide, yielding a single unambiguous inundation maximum. Members of the other three clusters generate their maximum forcing between high tides, yielding a lower maximum inundation, but a longer period of flooding. For future storms, this type of information could tell forecasters when predicting a hurricane's forward speed with greater precision would yield tangible benefits to forecasts of inundation volume and duration, and it could help tell emergency managers what types of flooding scenarios their jurisdictions may experience.

Results from this paper demonstrate that a suitably initialized WRF-ADCIRC ensemble can generate an ensemble of TC storm tide simulations that are useful for investigating uncertainties in surge hazard prediction. Although the storm surge inundation that occurs at different locations also depends on TC intensity and structure, our results illustrate that clustering by TC track can help to highlight distinct flooding risk scenarios within the ensemble, while showing how large- and small-scale track variations affect surge inundation. Examining results across and within individual clusters also helps to highlight how hazard forecast confidence may increase as TC track uncertainty decreases. Revisions to the clustering technique presented here may help clustering add additional value to the TC–surge ensemble.

Due to the high computational cost of ADCIRC, running a large ADCIRC ensemble may not yet be practical for real-time forecasting. However, we anticipate that the method demonstrated here does not depend on the surge model used. For example, this clustering method could be employed with a surge ensemble generated using the more efficient Sea Lake and Overland Surge from Hurricanes (SLOSH) model, which NHC

currently uses for its surge forecasts. The TC tracks could be obtained from dynamical atmospheric models, and the wind fields parameterized from model output for use in SLOSH.

The findings of this paper indicate that dynamical atmosphere–surge ensembles, along with track clustering, show promise for evaluating TC-induced storm surge hazards. To rely on such ensembles, however, the dynamic atmospheric ensemble must be able to adequately represent true forecast uncertainty across a range of cases. In this work, most 8 September simulations track right of Irma's observed track, especially after landfall. Although we cannot definitively say that this ensemble failed to capture the true forecast uncertainty, a right-biased ensemble would likely underestimate the coastal flooding threat on the Florida west coast and overestimate the threat on the Georgia and South Carolina coasts. Therefore, our study suggests that further work is needed to investigate whether atmospheric ensemble forecasts approximate the true forecast uncertainty sufficiently well to drive probabilistic storm surge forecasts. Other areas for future work include applying similar methods to other TCs and other types of TC hazards such as high winds, tornadoes, rainfall, and combined surge–wave–rainfall flooding, as well as evaluating how different methods of representing TC intensity and structure within an ensemble affect storm surge forecasts.

Acknowledgments. We thank Linus Magnusson at ECMWF, who provided the ECMWF ensemble forecasts used to initialize the WRF simulations in this study. We are also grateful to Chris Snyder and Chris Davis for their advice on generating the WRF ensemble. Portions of this research were supported by National Science Foundation Award 1331490. We would like to acknowledge high-performance computing support from the Penn State Institute for Computational and Data Sciences–Advanced CyberInfrastructure, led by Jenni Evans, as well as from Cheyenne ([doi:10.5065/D6RX99HX](https://doi.org/10.5065/D6RX99HX)) provided by NCAR's Computational and Information Systems Laboratory. The National Center for Atmospheric Research is sponsored by the National Science Foundation. We appreciate the three anonymous reviewers whose comments substantially improved this manuscript.

Data availability statement: ECMWF forecasts used to initialize the WRF simulations are not publicly available, but were provided by Linus Magnusson at ECMWF upon request. If you desire to use these data for your research, please contact the corresponding author and they will be made available.

REFERENCES

- Benjamin, S. G., J. M. Brown, G. Brunet, P. Lynch, K. Saito, and T. Schlatter, 2019: 100 years of progress in forecasting and NWP applications. *A Century of Progress in Atmospheric and Related Sciences: Celebrating the American Meteorological Society Centennial*, Meteor. Monogr., No. 59, Amer. Meteor. Soc., <https://doi.org/10.1175/AMSMONOGRAPHIS-D-18-0020.1>.
- Bostrom, A., R. E. Morss, J. K. Lazo, J. L. Demuth, H. Lazrus, and R. Hudson, 2016: A mental models study of hurricane forecast and warning production, communication, and decision-making. *Wea. Climate Soc.*, **8**, 111–129, <https://doi.org/10.1175/WCAS-D-15-0033.1>.
- Buizza, R., 2018: Introduction to the special issue on “25 years of ensemble forecasting”. *Quart. J. Roy. Meteor. Soc.*, **145** (S1), 1–11, <https://doi.org/10.1002/qj.3370>.
- Camargo, S. J., A. W. Robertson, S. J. Gaffney, P. Smyth, and M. Ghil, 2007: Cluster analysis of typhoon tracks: Part I: General properties. *J. Climate*, **20**, 3635–3653, <https://doi.org/10.1175/JCLI4188.1>.
- Cangialosi, J. P., A. S. Latta, and R. Berg, 2018: National Hurricane Center tropical cyclone report: Hurricane Irma (30 August–12 September 2017). NOAA/NWS Rep. AL112017, 111 pp., https://www.nhc.noaa.gov/data/tcr/AL112017_Irma.pdf.
- Colle, B. A., and Coauthors, 2015: Exploring water level sensitivity for metropolitan New York during Sandy (2012) using ensemble storm surge simulations. *J. Mar. Sci. Eng.*, **3**, 428–443, <https://doi.org/10.3390/jmse3020428>.
- Demuth, J. L., R. E. Morss, B. H. Morrow, and J. K. Lazo, 2012: Creation and communication of hurricane risk information. *Bull. Amer. Meteor. Soc.*, **93**, 1133–1145, <https://doi.org/10.1175/BAMS-D-11-00150.1>.
- Di Liberto, T., B. Colle, N. Georgas, A. Blumberg, and B. Taylor, 2011: Verification of a multimodel storm surge ensemble around New York City and Long Island for the cool season. *Wea. Forecasting*, **26**, 922–939, <https://doi.org/10.1175/WAF-D-10-05055.1>.
- Dietrich, J. C., A. Muhammad, M. Curcic, A. Fathi, C. N. Dawson, S. S. Chen, and R. A. Luettich Jr., 2018: Sensitivity of storm surge predictions to atmospheric forcing during Hurricane Isaac. *J. Waterw. Port Coastal Ocean Eng.*, **144**, 04017035, [https://doi.org/10.1061/\(ASCE\)WW.1943-5460.0000419](https://doi.org/10.1061/(ASCE)WW.1943-5460.0000419).
- ECMWF, 2017: IFS documentation CY43R3. ECMWF Tech. Rep., accessed 25 July 2019, https://www.ecmwf.int/en/publications/search/?solrsort=sort_label%20asc&secondary_title=%22IFS%20Documentation%20CY43R3%22.
- , 2019: ENS cluster products. Accessed 1 April 2019, <https://www.ecmwf.int/en/forecasts/documentation-and-support/ens-cluster-products>.
- Feyen, J., S. Vinogradov, T. G. Asher, J. Halgren, and Y. Funakoshi, 2015: Development of operational storm surge guidance to support total water predictions. *14th Int. Workshop on Wave Hindcasting and Forecasting/Fifth Coastal Hazard Symp.*, Key West, FL, JCOMM, X3, http://www.waveworkshop.org/14thWaves/Presentations/X3_Feyen_Waves_OpsSurgeGuidance_Nov2015-FINAL.pdf.
- Flowerdew, J., K. Horsburgh, C. Wilson, and K. Mylne, 2010: Development and evaluation of an ensemble forecasting system for coastal storm surges. *Quart. J. Roy. Meteor. Soc.*, **136**, 1444–1456, <https://doi.org/10.1002/qj.648>.
- , K. Mylne, C. Jones, and H. Tittle, 2013: Extending the forecast range of the UK storm surge ensemble. *Quart. J. Roy. Meteor. Soc.*, **139**, 184–197, <https://doi.org/10.1002/qj.1950>.
- Fossell, K. R., D. Ahijevych, R. E. Morss, C. Snyder, and C. Davis, 2017: The practical predictability of storm tide from tropical cyclones in the Gulf of Mexico. *Mon. Wea. Rev.*, **145**, 5103–5121, <https://doi.org/10.1175/MWR-D-17-0051.1>.
- Fundel, V. J., N. Fleischhut, S. M. Herzog, M. Göber, and R. Hagedorn, 2019: Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers and end-users. *Quart. J. Roy. Meteor. Soc.*, **145**, 210–231, <https://doi.org/10.1002/qj.3482>.
- Gaffney, S. J., A. W. Robertson, P. Smith, S. J. Camargo, and M. Ghil, 2007: Probabilistic clustering of extratropical cyclones using regression mixture models. *Climate Dyn.*, **29**, 423–440, <https://doi.org/10.1007/s00382-007-0235-z>.
- Georgas, N., and Coauthors, 2016: The Stevens Flood Advisory System: Operational H3E flood forecasts for the Greater New York/New Jersey Metropolitan region. *Int. J. Saf. Secur. Eng.*, **6**, 648–662, <https://doi.org/10.2495/SAFE-V6-N3-648-662>.
- Keller, J. H., S. C. Jones, J. L. Evans, and P. A. Harr, 2011: Characteristics of the TIGGE multimodel ensemble prediction system in representing forecast variability associated with extratropical transition. *Geophys. Res. Lett.*, **38**, L12802, <https://doi.org/10.1029/2011GL047275>.
- , —, and P. A. Harr, 2014: An eddy kinetic energy view of physical and dynamical processes in distinct forecast scenarios for the extratropical transition of two tropical cyclones. *Mon. Wea. Rev.*, **142**, 2751–2771, <https://doi.org/10.1175/MWR-D-13-00219.1>.
- Kowaleski, A. M., and J. L. Evans, 2016: Regression mixture model clustering of multimodel ensemble forecasts of hurricane Sandy: Partition characteristics. *Mon. Wea. Rev.*, **144**, 3825–3846, <https://doi.org/10.1175/MWR-D-16-0099.1>.
- , and —, 2018: Relationship between the track and structural evolution of Hurricane Sandy (2012) using a regional ensemble. *Mon. Wea. Rev.*, **146**, 4279–4302, <https://doi.org/10.1175/MWR-D-18-0121.1>.
- Kuruppumullage Don, P., J. L. Evans, F. Chiaromonte, and A. M. Kowaleski, 2016: Mixture-based path clustering for synthesis of ECMWF ensemble forecasts of tropical cyclone evolution. *Mon. Wea. Rev.*, **144**, 3301–3320, <https://doi.org/10.1175/MWR-D-15-0214.1>.
- Landsea, C. W., and J. P. Cangialosi, 2018: Have we reached the limits of predictability for tropical cyclone track forecasting? *Bull. Amer. Meteor. Soc.*, **99**, 2237–2243, <https://doi.org/10.1175/BAMS-D-17-0136.1>.
- Leonardo, N. M., and B. A. Colle, 2017: Verification of multimodel ensemble forecasts of North Atlantic tropical cyclones. *Wea. Forecasting*, **32**, 2083–2101, <https://doi.org/10.1175/WAF-D-17-0058.1>.
- Morrow, B. H., J. K. Lazo, J. Rhome, and J. Feyen, 2015: Improving storm surge risk communication: Stakeholder perspectives. *Bull. Amer. Meteor. Soc.*, **96**, 35–48, <https://doi.org/10.1175/BAMS-D-13-00197.1>.
- Morss, R. E., J. L. Demuth, and J. K. Lazo, 2008: Communicating uncertainty in weather forecasts: A survey of the U.S. public. *Wea. Forecasting*, **23**, 974–991, <https://doi.org/10.1175/2008WAF0007088.1>.
- NASEM, 2006: *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts*. The National Academies Press, 124 pp.
- , 2018: *Integrating Social and Behavioral Sciences Within the Weather Enterprise*. The National Academies Press, <https://doi.org/10.17226/24865>.

- NCAR, 2020: THORPEX Interactive Grand Global Ensemble (TIGGE) model tropical cyclone track data. Research Data Archive, National Center for Atmospheric Research, Computational and Information Systems Laboratory, accessed 3 January 2020, <https://doi.org/10.5065/D6GH9GSZ>.
- NHC, 2019: National Hurricane Center forecast verification. National Hurricane Center, accessed 1 April 2019, <https://www.nhc.noaa.gov/verification/>.
- , 2020: GUIDANCE from the Probabilistic Tropical Storm Surge (P-Surge) model. National Hurricane Center, accessed 10 February 2020, <https://slosh.nws.noaa.gov/psurge2.0/>.
- NOAA Tides & Currents, 2020: Water levels-station selection. NOAA, accessed 10 April 2020, <https://tidesandcurrents.noaa.gov/stations.html?type=Water+Levels>.
- NWS, 2019: Hurricane and tropical storm watches, warnings, advisories and outlooks. NOAA/NWS, accessed 1 April 2019, <https://www.weather.gov/safety/hurricane-ww>.
- Rappaport, E. N., 2014: Fatalities in the United States from Atlantic tropical cyclones: New data and interpretation. *Bull. Amer. Meteor. Soc.*, **95**, 341–346, <https://doi.org/10.1175/BAMS-D-12-00074.1>.
- Riverside Technology and AECOM, 2015: Mesh development, tidal validation, and hindcast skill assessment of an ADCIRC model for the hurricane storm surge operational forecast system on the U.S. Gulf-Atlantic Coast. NOAA/NOS Tech. Rep., 180 pp., https://www.earthsystemcog.org/site_media/projects/umac_model_advisory/HSSOFS_Development_Evaluation_20150410.pdf.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <https://doi.org/10.5065/D68S4MVH>.
- UCAR, 2019: User's guides for the Advanced Research WRF (ARW) modeling system, version 3.8. UCAR, accessed 12 August 2019, http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3.8/contents.html.