NOAA'S COASTAL OCEAN REANALYSIS: GULF OF MEXICO, ATLANTIC, AND CARIBBEAN

Silver Spring, Maryland
January 2025

noaa **National Oceanic and Atmospheric Administration**

U.S. DEPARTMENT OF COMMERCE
National Ocean Service
Center for Operational Oceanographic Products and Services

Center for Operational Oceanographic Products and Services
National Ocean Service
National Oceanic and Atmospheric Administration
U.S. Department of Commerce

The National Ocean Service (NOS) Center for Operational Oceanographic Products and Services (CO-OPS) provides the national infrastructure, science, and technical expertise to collect and distribute observations and predictions of water levels and currents to ensure safe, efficient and environmentally sound maritime commerce. The Center provides the set of water level and tidal current products required to support NOS' Strategic Plan mission requirements, and to assist in providing operational oceanographic data/products required by NOAA's other Strategic Plan themes. For example, CO-OPS provides data and products required by the National Weather Service to meet its flood and tsunami warning responsibilities. The Center manages the National Water Level Observation Network (NWLON), a national network of Physical Oceanographic Real-Time Systems (PORTS®) in major U. S. harbors, and the National Current Observation Program consisting of current surveys in near shore and coastal areas utilizing bottom mounted platforms, subsurface buoys, horizontal sensors and quick response real time buoys. The Center: establishes standards for the collection and processing of water level and current data; collects and documents user requirements, which serve as the foundation for all resulting program activities; designs new and/or improved oceanographic observing systems; designs software to improve CO-OPS' data processing capabilities; maintains and operates oceanographic observing systems; performs operational data analysis/quality control; and produces/disseminates oceanographic products.

NOAA'S COASTAL OCEAN REANALYSIS: GULF OF MEXICO, ATLANTIC, AND CARIBBEAN

Analise Keeney, Gregory Dusek, John Callahan, John Ratcliff, Tigist Jima
National Ocean Service
Center for Operational Oceanographic Products and Services
Silver Spring, Maryland, USA

William Brooks, Doug Marcy
National Ocean Service
Office for Coastal Management
Charleston, South Carolina, USA

Brian Blanton, Jeffrey Tilson
University of North Carolina
Renaissance Computing Institute
Chapel Hill, North Carolina, USA

Taylor G. Asher, Richard A. Leuttich Jr.
University of North Carolina
Coastal Resilience Center
Chapel Hill, North Carolina, USA

Matthew J. Widlansky, Linta Rose
University of Hawai'i at Mānoa
Cooperative Institute for Marine and Atmospheric Research
Honolulu, Hawaii, USA

U.S. DEPARTMENT OF COMMERCE
Gina M. Raimondo, Secretary

National Oceanic and Atmospheric Administration
Dr. Richard W. Spinrad, Under Secretary of Commerce for Oceans and Atmosphere

National Ocean Service
Nicole LeBoeuf, Assistant Administrator

Center for Operational Oceanographic Products and Services
Marian Westley, PhD, Director

NOAA'S COASTAL OCEAN REANALYSIS: GULF OF MEXICO, ATLANTIC, AND CARIBBEAN

Cheryl Morse
RPS Group, A Tetra Tech Company
Abington, England, UK

Jana Haddad, John Callahan, John Ratcliff
Ocean Associates, Inc.
Arlington, VA, USA

Blake Waring
Consolidated Safety Systems, Inc.
Fairfax, VA, USA

January 2025

U.S. DEPARTMENT OF COMMERCE
Gina M. Raimondo, Secretary

National Oceanic and Atmospheric Administration
Dr. Richard W. Spinrad, Under Secretary of Commerce for Oceans and Atmosphere

National Ocean Service
Nicole LeBoeuf, Assistant Administrator

Center for Operational Oceanographic Products and Services
Marian Westley, PhD, Director

NOTICE

Mention of a commercial company or product does not constitute an endorsement by NOAA. Use of information from this publication for publicity or advertising purposes concerning proprietary products or the tests of such products is not authorize.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

NOAA's Center for Operational Oceanographic Products and Services (CO-OPS) is the authoritative source for accurate, reliable, and timely data on tides, water levels, and currents, as well as other coastal oceanographic and meteorological information. CO-OPS maintains and operates NOAA's National Water Level Observation Network (NWLON), which provides real-time and historic water level observations at over 200 locations across U.S. coastlines. These observations enable monitoring and prediction of long-term sea level change, extreme events, and coastal flood frequency, which is essential information for preparing and planning for coastal hazards. While very useful, NWLON-based information is directly applicable to and representative of the conditions in the immediate vicinity of each water level station, creating extensive gaps in coverage in both densely populated areas and along remote or rural coastlines where distance between station locations can exceed 200 miles. These data are only collected in locations with continuous inundation (i.e., open water areas) and do not provide information over land, within many estuaries, or along most streams and rivers.

To bridge gaps in service and more equitably serve the Nation's coastal communities, NOAA's Coastal Ocean Reanalysis (CORA) couples long-term water level observations with hydrodynamic modeling to create historical information between tide stations. CORA water levels are simulated with ADvanced CIRCulation (ADCIRC) for ocean circulation modeling and coupled with a phase-averaging model called Simulating WAves Nearshore (SWAN) to produce surface gravity wave spectra and account for time-averaged wave contributions. Coastal water level observations from NOAA's NWLON are low-pass filtered and assimilated into the model to account for long-term sea level variability and to reduce model errors. The domain of this reanalysis spans the Gulf of Mexico, Atlantic (East), and Caribbean coastlines (or CORA-GEC). Modeling and statistical analyses were conducted for a period of 44 years, 1979-2022, with each year handled independently.

Skill assessment performed by the University of Hawaii's Sea Level Center found that CORA performed well when compared to NWLON observations. Modeled hourly water level errors across all 112 NWLON locations for the assimilated model run have a mean absolute error (MAE) of only 9.1 cm and a root mean squared error (RMSE) of only 11.4 cm. These errors are an improvement of roughly 40% compared to the initial unassimilated run. Hourly errors at stations used for the assimilation (MAE of 7.6 cm and RMSE of 9.4 cm) are unsurprisingly improved, and errors at stations withheld from the assimilation are still quite low (MAE of 10.6 cm and RMSE of 13.3 cm). Importantly, mean bias is acceptable regardless of assimilation status and is under 1 cm across all stations in the assimilation run.

Model error is dependent on geography and time. Regional performance differences are presumably related to the density of water level stations available for assimilation, with areas having fewer stations showing lesser performance across most metrics (e.g., along the western Gulf Coast). CORA performance was consistently weakest at water level stations located up coastal rivers, most likely due to inadequate resolution of these water bodies in the ADCIRC grid, a lack of riverine inflow, and fewer stations for assimilation. With a typical coastal resolution of 400-500 m, the mesh is too coarse to capture some narrow inlets into bays, and errors can be relatively large. Temporal dependence to errors is also suspected due to the increasing availability of stations (more stations in later decades) for assimilation or validation, as well as dependence on seasonality and various water level forcings that may not be captured well in the model representation. Model resolution is key to improving accuracy under all

conditions and in all areas, but it is especially important for improving accuracy behind barrier islands and up river.

CORA adeptly captures storm surges for many extratropical and tropical events, like hurricanes Sandy and Irma. However, CORA underperforms for some tropical systems, especially those that might be represented poorly by the relatively coarse weather forcing resolution (~30 km). This typically results in peak surge being lower than observed, especially in cases with storms that have strong, concentrated wind fields.

Considering both longer-term water level variability and extreme events, CORA closely tracks observed water level variations. Looking at several example stations, spectral errors in water level are several orders of magnitude less than the modeled water level variability at time scales between 1 and 180 days. Similarly, a cursory extreme analysis shows that CORA simulated extremes at 5- and 10-year return intervals with errors within 10 cm of the observed water levels for the majority of comparison stations.

Waves are an important aspect of CORA, as they are included to account for time-averaged wave contributions to storm surge, known as "wave setup." In addition to the coupled contributions of waves, SWAN spectral wave output is included in CORA, even though this was not a primary focus of CORA version 1.1. A limited skill assessment of waves demonstrates acceptable performance at a minimal number of coastal wave observation sites available in the study region, though a more comprehensive assessment of wave performance will be completed in the future.

Derived coastal flood products are also an important aspect of CORA development. To enable continued development and enhancement of these products, the ADCIRC output is interpolated onto a 500 m grid to create continuous geospatial resolution. Decoupling the model mesh from the grid for applied products enables future updates to the mesh, or even the entire model structure, without needing to rederive the spatial delivery of downstream products. Geoprocessing tools are used to create a 500 m grid to include all ADCIRC nodes within 3 km of shore, with onshore grid points capturing the spatial extent of nodes that experience at least 1 flood event per year.

Future enhancements to CORA-GEC include improved accuracy of tide simulations, meteorological forcing that includes low-frequency events like tropical cyclones, and increased model resolution. Increasing model resolution may be especially relevant to improve performance in complex coastal regions like those behind barrier islands and up rivers. Reanalysis of the Pacific domain began in 2024 and will be sectioned into 3 sub-regions: the Pacific Coast and Hawaiian Islands, Alaska, and the Pacific Islands. The Great Lakes are not yet a part of CORA datasets, though an approach has been drafted and the project team acknowledges that a reanalysis of the lakes will improve the resolution of 6-month Lake Level Predictions and add to the value of reanalysis datasets in the future.

Data associated with this project and report are considered CORA-GEC version 1.1. Versioning is in place to identify the datasets analyzed here from previous and future versions of CORA, including version 0.9, which was assessed in a validation performed by the University of Hawaii's Sea Level Center (Rose et al. 2024).

1. INTRODUCTION & BACKGROUND

NOAA's Center for Operational Oceanographic Products and Services (CO-OPS) has a robust history of collecting, storing, archiving, and distributing authoritative oceanographic data from across the coastal United States. These authoritative datasets include coastal water level observations, which provide the foundation for research and products focused on many topics including sea level monitoring (Zervas 2009), mean sea level (MSL) projections (Sweet et al. 2022), extreme and minor high tide flooding (HTF; Sweet et al. 2018; Sweet et al. 2022), and subseasonal to seasonal flood prediction (Dusek et al. 2023). These research efforts and resulting products are made possible with NOAA's National Water Level Observation Network (NWLON), a system of over 200 real-time water level stations with extensive, high-quality long-term data records that in some cases extend over 100 years. Though these water level products are critical for monitoring, safety, and preparedness, as well as improving scientific understanding to make predictions about future coastal conditions, dependence on NWLON stations means these variables are almost exclusively provided at specific point locations, sometimes hundreds of miles apart. We are repeatedly reminded that researchers, partners, users, and stakeholders require water level data everywhere, not just where there happens to be a station. This raises the question: Can we use foundational NWLON observation and state-of-the-art numerical modeling to produce long-term water level data everywhere along the U.S. ocean coastline?

NOAA answers this question through the development of the Coastal Ocean Reanalysis (CORA). CORA is a coastal hydrodynamic model reanalysis of hourly water levels and wind-driven waves at a spatial resolution of approximately 200-500 m along the entire Gulf of Mexico, Atlantic (East), and Caribbean coastlines (CORA-GEC) from 1979-2022. Recent advancements in atmospheric (European Centre for Medium-Range Weather Forecasts [ECMWF] Reanalysis version 5 [ERA5][1]) and hydrodynamic modeling (Luettich et al. 1992; Westerink et al. 2008; Booij et al. 1999; Zijlema 2010) enable improvements in accuracy and spatial resolution of depth and extent of coastal inundation modeling. NOAA's robust, long-term NWLON observations are then leveraged for data assimilation to account for long-term sea level variations, resulting in further improved model accuracy for hourly water level errors typically less than 10 cm. Applying reanalyzed modeled atmospheric forcings from the ERA5[2] coastal hydrodynamic model allows for the development of long-term, multi-decadal, historical coastal water level and waves datasets for a variety of uses. Half of the NWLON observations used to produce CORA were assimilated into the modeling to increase accuracy by reducing error; the other half were intentionally set aside to enable independent model validation.

CORA water levels are simulated with the ADvanced CIRCulation[3] model (ADCIRC). ADCRIC's 2-dimensional barotropic mode accounts for tides, wind, and atmospheric pressure-driven surge and wind-wave characteristics via ADCIRC's coupling with the phase-averaging Simulating WAves Nearshore model (SWAN; Dietrich et al. 2011a) that simulates surface

[1] ECMWF Reanalysis: https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/qj.3803

[2] ERA5 Documentation:
   a. Characterizing ERA-Interim and ERA5 surface wind biases using ASCAT: https://os.copernicus.org/articles/15/831/2019/
   b. Improvements in storm surge representation: https://link.springer.com/article/10.1007/s00382-019-05044-0

[3] ADCIRC Publications: https://adcirc.org/home/documentation/adcirc-related-publications/

gravity wave spectra and accounts for time-averaged wave contributions (wave setup) to coastal water levels[4]. Coastal water level observations are low-pass filtered and assimilated into the model following the method described in Asher et al. (2019). There are 112 NWLON stations within the Gulf, East, and Caribbean model domain, with 53 used for data assimilation and 59 used for an independent assessment of the assimilated model results. Validation of preliminary version 0.9 of CORA for the Gulf and East coasts (Rose et al. 2024) found the reanalysis compares closely with most non-assimilated water level observations. CORA version 0.9 demonstrates improved performance when compared to the state-of-the-art eddy-resolving (1/12°) global ocean reanalysis (GLORYS12; Jean-Michel et al. 2021).

CORA datasets are available through NOAA's Open Data Dissemination Platform (NODD)[5] via Amazon Web Services (AWS). Datasets have been optimized to enable rapid access and data retrieval. Python-based Jupyter Notebooks[6] are available to guide users through accessing, analyzing, and visualizing CORA datasets through common use cases. In addition to CORA's native ADCIRC datasets, gridded datasets are interpolated to a 500 m coastal grid to address the need for reduced file sizes, to provide a consistent framework for service delivery, and to streamline the development of downstream products and applications. Gridded data are being used to develop visualizations and statistics like daily maximums, event-driven water levels, seasonal variations, long-term trends, and spatial extent of flood exposure, among many others. Gridded datasets provide continuous geospatial resolution, a necessity for enhancing sea level trends, extreme water level prediction, and valuable coastal hazard planning and mitigation products like the NOAA Monthly HTF Outlook[7] (Dusek et al. 2022). National gridded datasets with continuous spatial resolution will help enhance the accuracy of the national flood frequency estimates and projections provided in the Interagency Sea Level Scenarios (Sweet et al. 2022) and the 5th National Climate Assessment (IPCC 2014).

This report details the approach, results, and performance of CORA-GEC version 1.1. Please note that preliminary versions of CORA, like version 0.9 analyzed in Rose et al. (2024) and version 1.0 used for prototyping, are enhanced and superseded by CORA-GEC version 1.1. Section 2 provides a project description, including details for different components of the project, naming conventions and future coverage expansion. Section 3 describes overall methods, including ADCIRC configuration, mesh, forcing, datums, and units of the model output. Section 4 includes the data acquisition, processing, and assimilation scheme, as well as how the error assessment was completed. Results are covered in Section 5 and include a comparison with NWLON observations, wave observations, an extremes analysis, and a summary of model limitations. Section 6 addresses grid development and interpolation and describes some initial downstream product development. Lastly, Section 7 provides acknowledgements, references, and appendices.

[4] Simulation WAves Nearshore (SWAN): https://swanmodel.sourceforge.io/
[5] NOAA Open Data Dissemination (NODD) Platform via Amazon Web Services: https://registry.opendata.aws/noaa-nos-cora/
[6] GitHub Repository for CORA's Python Jupyter Notebooks: https://github.com/NOAA-CO-OPS/CORA-Coastal-Ocean-Reanalysis
[7] NOAA Monthly High Tide Flooding Outlook: https://tidesandcurrents.noaa.gov/high-tide-flooding/monthly-outlook.html

2. PROJECT DESCRIPTION

NOAA's CORA is the culmination of collaborative efforts across the National Ocean Service (NOS), the Coastal Resilience Center and the Renaissance Computing Institute (RENCI) at the University of North Carolina (UNC) at Chapel Hill, and the University of Hawaii Sea Level Center (UHSLC) at the Cooperative Institute for Marine and Atmospheric Research (CIMAR). Though global atmosphere and ocean reanalyses are common, they have not been frequently utilized for high-resolution coastal applications. CORA-derived datasets mark a shift in the way NOAA provides impactful coastal oceanographic products and services. Most importantly, the increased resolution of CORA's modeled historical water levels allows NOAA to more equitably provide valuable information to coastal communities located far from water level stations.

Requirements for this collaborative project were drafted in 2020 and signed into agreement in 2021, establishing roles and responsibilities for each participant. CO-OPS leads and manages the project, provides verified hourly water levels for stations with a record from 1979-2022, and supports model skill assessment, data access, and product development. NOAA's Office of Coast Management (OCM) is responsible for developing the methodology for translating reanalyzed data from an unstructured triangular ADCIRC mesh to a uniformly gridded product with data points every 500 m along the coast, including riverine areas, estuaries, and bays. The Integrated Ocean Observing System (IOOS) funded collaboration with the RPS Group, who optimize data for cloud-based analysis and interpolate ADCIRC datasets to a grid, developing a JupyterHub environment to develop Jupyter Notebooks for analyses and product development, and brokering transition of data to the NODD.

Outside of NOAA, partnership with the UNC RENCI propels the entire project by developing and testing methods, configuring numerical meshes needed for each reanalysis, performing several iterations of model runs, providing the high-performance computing (HPC) needed to run the models, and ensuring quality results through skill assessment. Additional academic collaboration with UHSLC CIMAR helped assess the skill of a preliminary version of CORA compared to water level observations and coastal sea levels from a global reanalysis (Rose et al. 2024).

CORA will be a national dataset, including the U.S. Pacific Coast, Hawaii, affiliated Pacific Islands, Alaska, and U.S. Territories. Datasets are being developed in 3 major sections with reanalysis of the Gulf of Mexico, Atlantic Ocean, Puerto Rico, and U.S. Virgin Islands (i.e., GEC) occurring first. This approach is helpful for managing computation time and overall data management but is ultimately used because ADCIRC grids for this region already exist. ADCIRC grids for the Pacific reanalysis are distinct, disjointed, and occurring in 3 parts: the Pacific Coast and Hawaiian Islands, Alaska, and U.S. Territories.

Data reviewed, analyzed, presented in this technical report, and available for public consumption are labeled "version 1.1." The initial version of the CORA-GEC was labeled "version 0.9" and used for the published validation study (Rose et al. 2024). Several improvements were made from version 0.9 to version 1.0, however the update introduced some shortcomings with properly accounting for long-term sea level variability during periods of limited observations. Thus, some portions of the data assimilation process were further improved (as detailed in section 4) to establish version 1.1, marking the first full public release for operational use and application.

The Great Lakes are not included in CORA datasets, though an approach has been drafted and the project team acknowledges that a reanalysis of the lakes will improve the

resolution of 6-month Lake Level Predictions and add to the value of reanalysis datasets in the future.

3. METHODS

Overall Reanalysis Workflow

The following is an overview of the approach for generating this reanalysis. Modeling and statistical analyses are conducted over the 44-year period of 1979-2022, with each year handled independently. The general workflow for a year is shown in Figure 1. In step 1, an initial (prior) year-long prediction of historical water levels is computed with the ADCIRC model using only tides and meteorological forcing. From this, a prior error series is generated (step 2) by comparing the prior predicted water levels with observed water levels at specified locations. This error is analyzed and processed to generate additional forcing terms in ADCIRC that assimilate the errors into the model (step 3), resulting in a posterior prediction (step 4). The effectiveness or prediction skill of the assimilation is determined by comparing observed water levels to the posterior prediction (step 5).

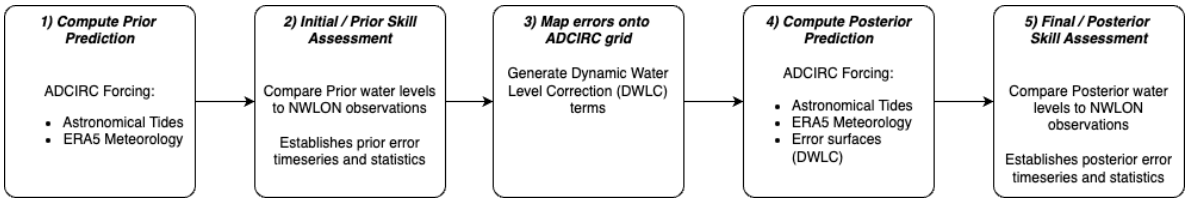

Figure 1. Sequence of steps in the reanalysis workflow, performed for each simulated year.

Each year of the reanalysis is run independently to simplify overall simulation, staging, processing, and management of resources on HPC clusters. This also allows for concurrent simulation of multiple years. Each year's simulation begins on 1 December at 00Z Coordinated Universal Time (UTC)[8] of the previous year, followed by a 15-day spin-up period during which tidal and meteorological forcings are ramped up to full strength. Tide nodal factors and equilibrium arguments for each tidal harmonic are specified at values for the middle of each simulation year. Steps 1 and 4 are the computationally intensive steps, as they involve running the ADCIRC model for each year on a large spatial grid. Steps 2, 3, and 5 are largely statistical analysis of the errors determined in steps 1 and 4.

ADCIRC, Grids, & Configuration

This reanalysis uses the depth-integrated implementation of the storm surge and tide model ADCIRC[9] (version 56) and its coupling to the SWAN wind wave model for all predictions and simulations. Specifically, version 56 was used to leverage an improved wetting/drying algorithm critical to capturing overland inundation. ADCIRC solves the shallow water equations using continuous Galerkin, linear finite elements (Luettich et al 1992; Westerink et al. 2008). It is coupled with the 3rd generation phase-averaged spectral wave model SWAN[10] that solves the conservation of wave action equation in time and space (Booij

[8] Coordinated Universal Time (Z-time): https://www.noaa.gov/jetstream/time
[9] https://adcirc.org/
[10] https://swanmodel.sourceforge.io/features/features.htm

et al. 1999; Zijlema 2010). Conveniently, SWAN operates on the same linear triangular mesh as ADCIRC. The 2 models are coupled through the exchange of total water depths and currents passed from ADCIRC to SWAN and by wave radiation stress gradients passed from SWAN to ADCIRC at each shared time step (Dietrich et al. 2011a,b). The coupled model is efficiently parallelized for execution on HPC systems (Tanaka et al. 2011; Dietrich et al. 2011a; Kerr et al. 2013), making it ideal for numerous coastal modeling applications, including tidal dynamics and prediction (e.g., Blanton et al. 2004; Hill et al. 2011), real-time and operational predictions (e.g., Fleming et al. 2008; Blanton et al. 2012; Dresback et al. 2013), short-term event-based studies (e.g., Kerr et al. 2013; Hope et al. 2013; Cyriac et al. 2018), and long-term hindcasts like this reanalysis.

NOAA's Hurricane Surge On-Demand Forecast System[11] (HSOFS) grid (Riverside Technologies, Inc. and AECOM 2015) is used for the CORA simulations because of its intermediate coastal spatial resolution and its long history within NOAA for operational coastal water level and storm surge guidance programs and the Extratropical Surge and Tide Operational Forecast Systems (ESTOFS; Riverside Technologies, Inc. and AECOM 2015). The grid has 1.8 million nodes and 3.6 million elements (Figure 2), and it covers the western North Atlantic Ocean and all overland areas of the neighboring U.S. coast up to an elevation of approximately 10 m above global MSL (more detailed datum information is provided in a subsequent section). The mesh provides a detailed representation of inlets, rivers, barrier islands, roadways, and other key features. Elevation data is based on bathymetry and lidar-derived topography available at the time the mesh was generated. Mesh resolution in coastal areas is typically 400-500 m and as high as 200 m in some areas, making it an intermediate-resolution grid developed specifically for national-scale real-time storm surge applications (as opposed to a high-resolution [50-100 m] grid one might develop for regionally or locally focused applications). The offshore/deep ocean resolution ranges from about 10-30 km. The grid's only open boundary extends from Nova Scotia out to the 55°W longitude down to the Suriname coast in northern South America (the easternmost curved boundary line). Details of the HSOFS grid development are available in Riverside Technologies, Inc. and AECOM (2015).

ADCIRC is configured using standard physics and numerical parameters, with several key parameters set using ADCIRC's nodal attribute functionality. Bottom friction is specified as a node-dependent Manning's function. Due to increased frictional effects from canopy cover and upwind directional surface roughness lengths, wind stress reduction over land is specified for each ADCIRC node on land using land cover data from the 2006 NOAA Coastal Change Analysis Program (C-CAP) dataset. The HSOFS grid requires a 2-second timestep for stability. Lateral eddy viscosity is specified using the flow- and grid-scale dependent Smagorinsky mixing scheme (Smagorinsky 1963). Details of nodal attribute implementation in ADCIRC and data sources used to set frictional parameters are described in Westerink et al. (2008).

[11] NOAA Hurricane Surge on Demand Forecast: https://www.weather.gov/sti/coastalact_surgewg

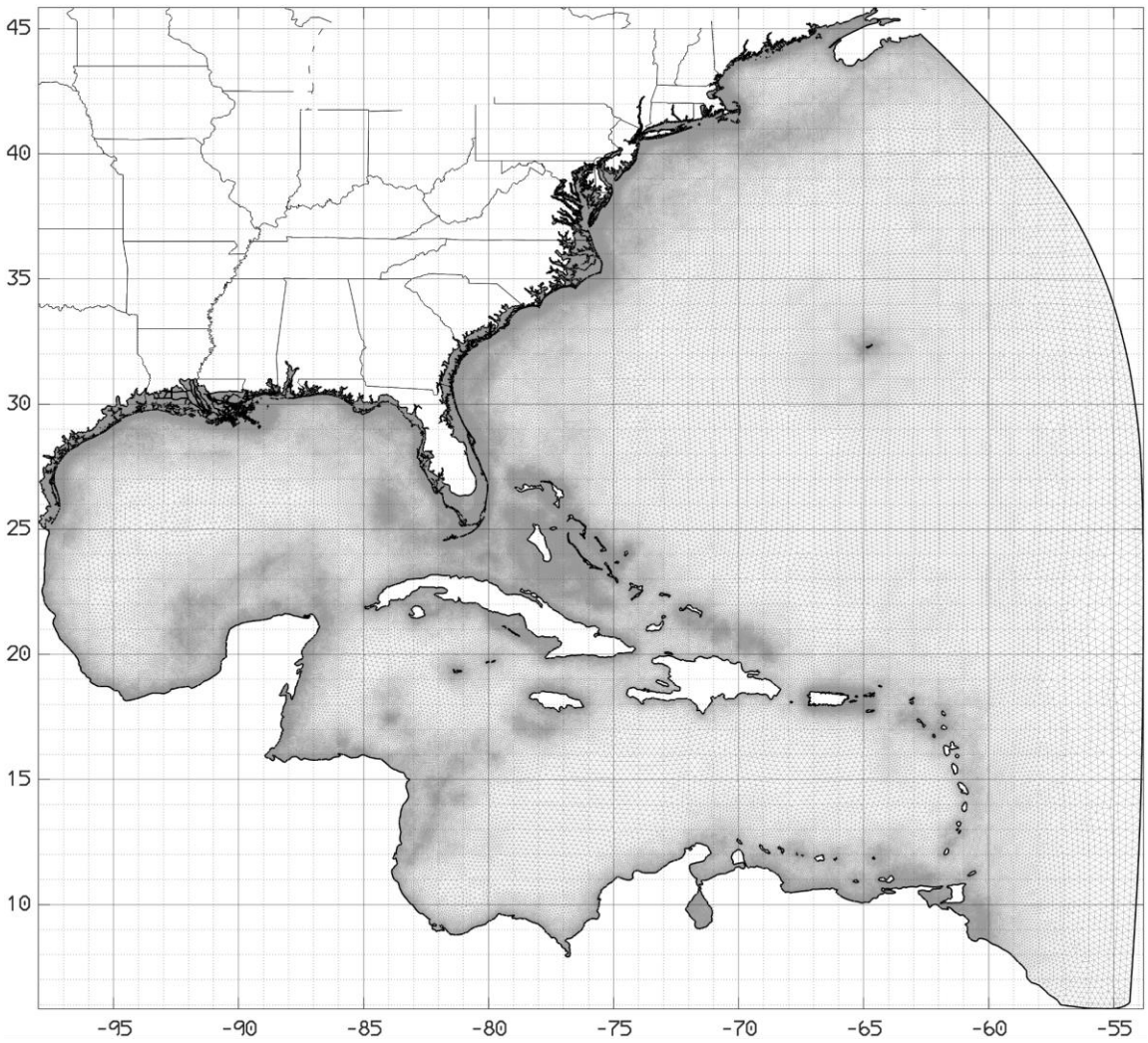

Figure 2. NOAA's Hurricane Surge On-Demand Forecast System (HSOFS) (ADCIRC) grid used for the reanalysis simulations. There are 1.8 million nodes and 3.6 million elements. Coastal resolution ranges from 200-500 m. Black contour lines represent mean sea level in the grid. Resolution in the deep ocean ranges from about 10-30 km.

Forcing Functions

Both prior and posterior simulations are forced with astronomical tides, atmospheric MSL pressure, and 10-m wind velocities. Tidal harmonics along the grid's open boundary are extracted from the version 9 of the Oregon State University TPXO (Egbert et al. 2002) tidal solution for the 1/12° North Atlantic region[12]. The harmonics used are M_2, S_2, N_2, K_2, K_1, O_1, P_1, Q_1, MM, and MF. Nodal and equilibrium values are computed for the middle of each simulation year. Tidal potential forcing terms are also included for the same harmonics.

For the meteorological forcing, the MSL pressure and 10-m U (east/west) and V (north/south) wind speeds from the ECMWF's ERA5 "hourly data on single pressure levels" collection over the reanalysis period of 1979-2022 were acquired through the Copernicus Climate Change Service's python application programming interface (API)[13]. Each hourly

[12] North Atlantic region tidal forcing data source: https://www.tpxo.net/regional
[13] https://cds.climate.copernicus.eu

global file is subsetted for the western North Atlantic Ocean and formatted for ADCIRC-specific wind and pressure inputs[14], specifically the Oceanweather Inc. (OWI)-formatted wind/pressure input type. A snapshot of the subsetted ERA5 fields for 06Z on 14 September 2018 is shown in Figure 3.

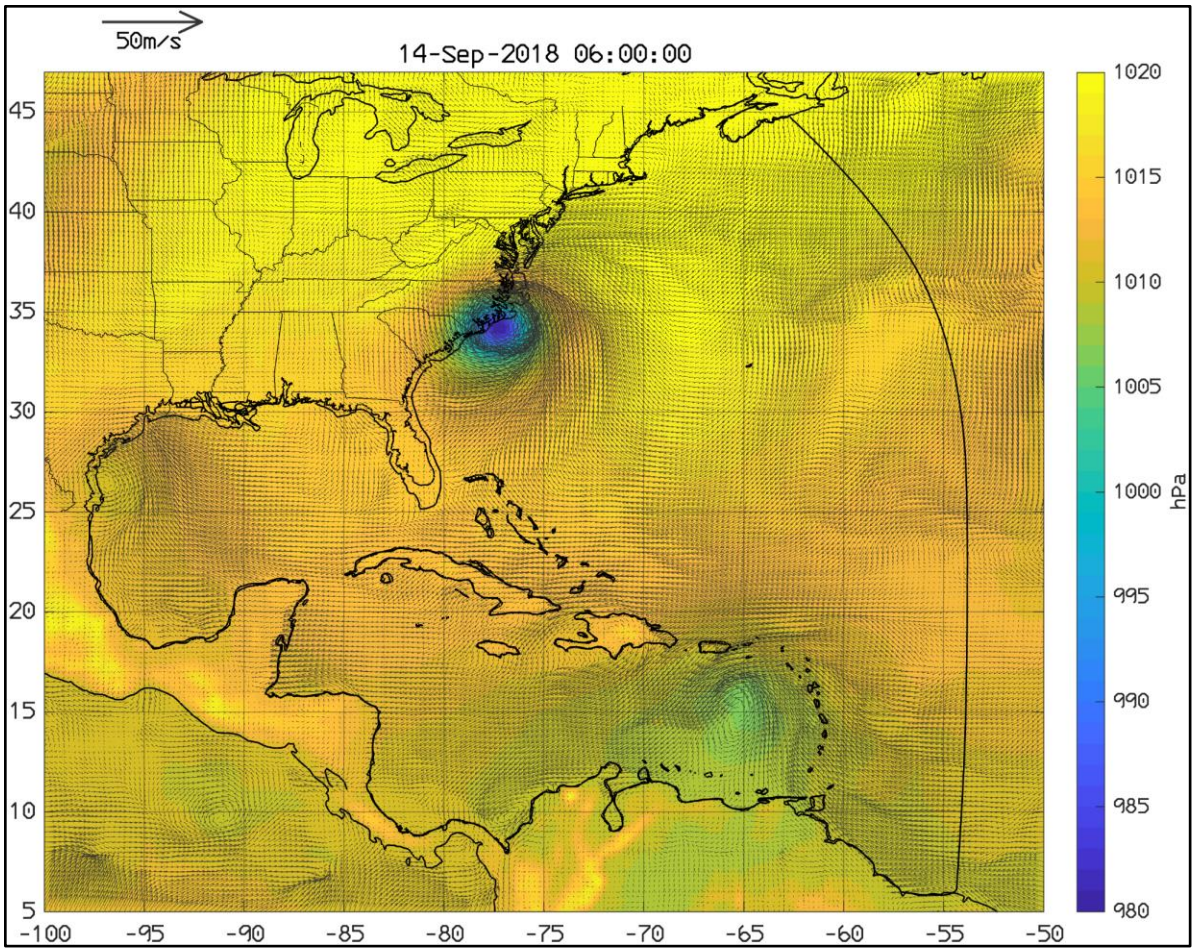

Figure 3. Snapshot of the European Centre for Medium-Range Weather Forecasts Reanalysis version 5 (ERA5) surface pressure and 10-m elevation wind speed for 14 September 2018 at 0600Z, as Hurricane Florence made landfall along the North Carolina coast. The vector scale is shown in the upper left. The highest wind speed in the plotted area is about 30 m/s in the core of the cyclone.

Datums & Units

Modeled data typically provides units as metadata. In cases where no units are provided, the International System of Units are applied: m for lengths, s for time, and m/s for speed. Horizontal coordinates are in longitude and latitude. For cases where a horizontal datum is needed, the World Geodetic System 1984 (WGS 84) is used. For cases where a vertical datum is needed, MSL relative to the 1983-2001 national tidal datum epoch (NTDE) is used, and model results should be considered to be relative to MSL. Note that 2 water level stations used in the assimilation process are currently reported in Modified 5-year tidal datum epochs. This is discussed below in the Data Sources and Processing section. Due to the mesh resolution

[14] https://adcirc.org/home/documentation/users-manual-v53/parameter-definitions#NWS

of ~500 m, the exact horizontal reference frame is not as critical. However, the vertical reference frame merits further discussion. There are 2 ways to relate modeled water levels to a real-world datum. One is through the vertical datum used to develop the mesh. The other is through the vertical adjustments to water levels that result from the assimilation. Each is discussed in the following paragraphs.

Topography and bathymetry used to build the mesh come from a variety of sources and datums, and are composited together and reprojected as necessary during the mesh construction process. For a more detailed description, see the mesh development and validation report (Riverside Technologies, Inc. and AECOM 2015). The mesh was constructed using The North American Vertical Datum of 1988 (NAVD 88), then converted to MSL once completed using a single surface provided by the NOAA Coast Survey Development Lab (CSDL) and derived from NOAA's Vertical Datum Transformation (VDatum) tool[15]. The surface was provided to mesh developers at AECOM in 2014, however the precise creation date is not known. Recent spot checks comparing the datum conversions in the surface to what is currently provided by VDatum have shown some areas differ. However, these changes are not considered to be substantially better given other uncertainties such as how the datum transformation must be extrapolated inland to cover the full mesh. As such, the original conversion surface can be used as a datum conversion for modeled water levels, though in cases where datum information has changed noticeably, caution is advised. A file with the conversions used to go between NAVD88 and MSL at each node in the mesh, hsofs_nomad_msl2navd88.grd, is archived with the OCM and can be provided to interested users.

A more relevant means of interpreting the vertical datum of modeled water levels comes from the assimilation during modeling. The ADCIRC numerical model solves for water levels resulting from forces acting upon it. If there is no forcing, all water levels are equal to 0 everywhere; this is known as an equipotential surface. In this case, 0 can be likened to MSL, though it is not intrinsically tied to any particular vertical datum. ADCIRC does not include some of the physical processes driving changes in MSL, like vertical land motion or glacial melt. Rather, data assimilation is used to give water levels an exact reference frame by comparing simulated water levels with observations from a set of coastal NWLON water level stations, discussed in greater detail below. These corrections are what lead to the reproduction of processes like sea level rise in modeled water levels. Therefore, the NOAA stations data used for the assimilation, and particularly the spatial surfaces that result, are what best characterize the vertical datum of modeled results. As such, there may be some additional uncertainty in the modeled MSL in places far from an assimilation site, as discussed in the Limitations subsection under the Water Level Results section.

4. DATA ASSIMILATION

The main goal for the CORA reanalysis of coastal water levels is to compute a superior prediction of coastal water levels based on a prior prediction, a prior error analysis, and a data assimilation approach for the ADCIRC model (Asher et al. 2019). We have focused on 2 important "bands" for the error analysis and posterior prediction, namely a monthly-mean component and subtidal weather band.

Given the importance of sea level increases over longer terms, incorporation of this component in the posterior is essential. Since the prior prediction (forced only by astronomical

[15] NOAA Vertical Datum Transformation Tool: https://vdatum.noaa.gov/

tides and the ERA5 meteorological reanalysis surface pressure and winds) does not contain a sea level rise component, long-term sea level changes are introduced through analysis of the prior error and subsequent assimilation of a low-frequency term that represents sea level stand, expressed in monthly mean sea levels (relative to a specific tidal datum epoch).

The subtidal weather band water level variability is due to several contributing factors, including meteorological forcing, astronomical tides, steric effects, and offshore influences on the coastal zone (e.g., Florida current or Gulfstream transport variations [Noble and Gelfenbaum 1992)]. In the prior prediction, only meteorology and tides are used as forcing. Thus, the data assimilation process improves the prior prediction irrespective of the physical source of the error (e.g., inaccurate meteorology, absent baroclinic forcing).

In both cases, errors in the prior prediction of coastal water levels are analyzed and used to improve the accuracy of the posterior prediction in these 2 critical parts. As such, the observations are split into 2 components: monthly means and residuals. Monthly mean time series are analyzed and processed to develop a complete (no temporal gaps) dataset over the 44-year period. The residual time series are filtered to remove tidal variability, after removing the monthly means, but gaps are not filled unless there is a nearby station that exhibits high correlation and little bias. The details of the data processing are given below.

Data Sources and Processing

The observational data source for the CORA reanalysis is the water level observations available from the CO-OPS NWLON stations data archive. Both verified hourly and monthly mean data were retrieved using the CO-OPS API[16] and accessed using the python package noaa-coops (version 0.3.2)[17]. All data was retrieved in meters relative to MSL defined by the 1983-2001 NTDE except for 2 stations (see below). Fifty-three stations were selected for assimilation based primarily on the criteria that 1) the station was within the model domain and 2) the station was in relatively open-coast conditions (Table 1).

Table 1. National Water Level Observation Network (NWLON) stations used in development of the data assimilation component.

ID	Long	Lat	State	Name
8410140	-66.982903	44.904598	ME	Eastport
8413320	-68.204278	44.392194	ME	Bar Harbor
8418150	-70.24417	43.65806	ME	Portland
8443970	-71.05028	42.3539	MA	Boston
8447930	-70.671112	41.523613	MA	Woods Hole
8449130	-70.096703	41.285	MA	Nantucket Island
8461490	-72.095556	41.371667	CT	New London
8467150	-73.183969	41.175819	CT	Bridgeport
8510560	-71.959444	41.048333	NY	Montauk

[16] https://api.tidesandcurrents.noaa.gov/api/prod/
[17] https://github.com/GClunies/noaa_coops

ID	Long	Lat	State	Name
8516945	-73.7649	40.810299	NY	Kings Point
8518750	-74.014167	40.700556	NY	The Battery
8531680	-74.009399	40.4669	NJ	Sandy Hook
8534720	-74.418053	39.356667	NJ	Atlantic City
8536110	-74.959999	38.9683	NJ	Cape May
8551910	-75.571944	39.558333	DE	Reedy Point
8557380	-75.119278	38.782833	DE	Lewes
8570283	-75.091086	38.328267	MD	Ocean City Inlet
8574680	-76.5783	39.27	MD	Baltimore
8635750	-76.465556	37.996389	VA	Lewisetta
8638901	-76.083298	37.032902	VA	Chesapeake Channel
8651370	-75.746696	36.1833	NC	Duck
8656483	-76.671111	34.7175	NC	Beaufort Duke Marine Lab
8658163	-77.786667	34.213333	NC	Wrightsville Beach
8661070	-78.916389	33.655556	SC	Springmaid Pier
8665530	-79.923889	32.775	SC	Charleston
8670870	-80.903028	32.034694	GA	Fort Pulaski
8720030	-81.465842	30.671356	FL	Fernandina Beach
8720218	-81.427889	30.398167	FL	Mayport (Bar Pilots Dock)
8721604	-80.593056	28.415833	FL	Trident Pier Port Canaveral
8722670	-80.034167	26.612778	FL	Lake Worth Pier Atlantic Ocean
8723214	-80.161667	25.731667	FL	Virginia Key Biscayne Bay
8723970	-81.1065	24.711	FL	Vaca Key Florida Bay
8724580	-81.807899	24.5557	FL	Key West
8725110	-81.8075	26.131667	FL	Naples Gulf of Mexico
8726520	-82.626944	27.761111	FL	St. Petersburg Tampa Bay
8727520	-83.0317	29.135	FL	Cedar Key
8729108	-85.664444	30.149722	FL	Panama City
8729840	-87.211197	30.4044	FL	Pensacola
8735180	-88.075	30.250278	AL	Dauphin Island

ID	Long	Lat	State	Name
8747437	-89.3258	30.326342	MS	Bay Waveland Yacht Club
8761305	-89.673	29.8683	LA	Shell Beach
8761724	-89.956667	29.263333	LA	Grand Isle
8768094	-93.342889	29.768167	LA	Calcasieu Pass
8770822	-93.841797	29.689301	TX	Texas Point Sabine Pass
8772471	-95.294197	28.935699	TX	Freeport Harbor
8774770	-97.046667	28.021667	TX	Rockport
8775870	-97.21666	27.579999	TX	Bob Hall Pier Corpus Christi
8779770	-97.215528	26.061167	TX	Port Isabel
9751401	-64.753799	17.6947	VI	Lime Tree Bay
9751639	-64.925806	18.330583	VI	Charlotte Amalie
9755371	-66.116417	18.458944	PR	San Juan La Puntilla San Juan Bay
9759394	-67.162444	18.218833	PR	Mayaguez
9759938	-67.938208	18.089289	PR	Mona Island

Monthly Mean Observations

The monthly mean time series for the 53 stations were further processed by merging data from nearby station locations to have as complete and practical a dataset with minimal gaps. This process was informed by Zervas (2009; Table 2) that defines groups of stations relatively close to each other and that are merged to temporally extend the monthly mean series beyond the currently active station times. The specific station pairs used for this data analysis is reported in Table 2.

Table 2. List of stations and corresponding nearby station(s) used to merge monthly mean time series.

Current/Active Station (as of December 2022)		Augmenting Station(s)	
Station ID	**Name**	**Station ID**	**Name**
8419870	Seavey Island, ME	8423898	Fort Point, NH
8516945	Kings Point, NY	8516990	Willets Point, NY
8570283	Ocean City Inlet, MD	8570280	Ocean City, Fishing Pier, MD
8638901	CBBT, Chesapeake Channel, VA	8638863	Chesapeake Bay Bridge Tunnel, VA
8720218	Mayport, FL	8720220	Mayport (Ferry Depot), FL
8723214	Virginia Key, FL	8723080 8723170	Haulover Pier, N. Miami Beach, FL Miami Beach, FL

Current/Active Station (as of December 2022)		Augmenting Station(s)	
Station ID	**Name**	**Station ID**	**Name**
8747437	Bay Waveland Yacht Club, MS	8747766	Waveland, MS
8761724	Grand Isle, LA	8761720	Grand Isle, LA
8770570	Sabine Pass North, TX	8770590	Sabine Pass, TX
8770822	Texas Point, Sabine Pass, TX	8770570	Sabine Pass North, TX
8771341	Galveston Bay Entrance, TX	8771510	Galveston Pleasure Pier, TX
8772471	Freeport Harbor, TX	8772447 8772440	USCG Freeport, TX Freeport, TX
8779748	South Padre Island CG Station, TX	8779750	South Padre Island, Brazos Santiago Pass, TX

An example of the merging result is shown in Figure 4 for Kings Point, NY (8516945), and Freeport Harbor, TX (8772471). In both cases, there exists some small temporal overlap of the contributing stations' data used to adjust the monthly means in the composite series. Green vertical lines indicate where gaps still exist after the merging. These gaps are filled with the relevant monthly mean from the entire series, adjusted for the long-term trend. The most recent data of a station pair always takes precedence over older data.

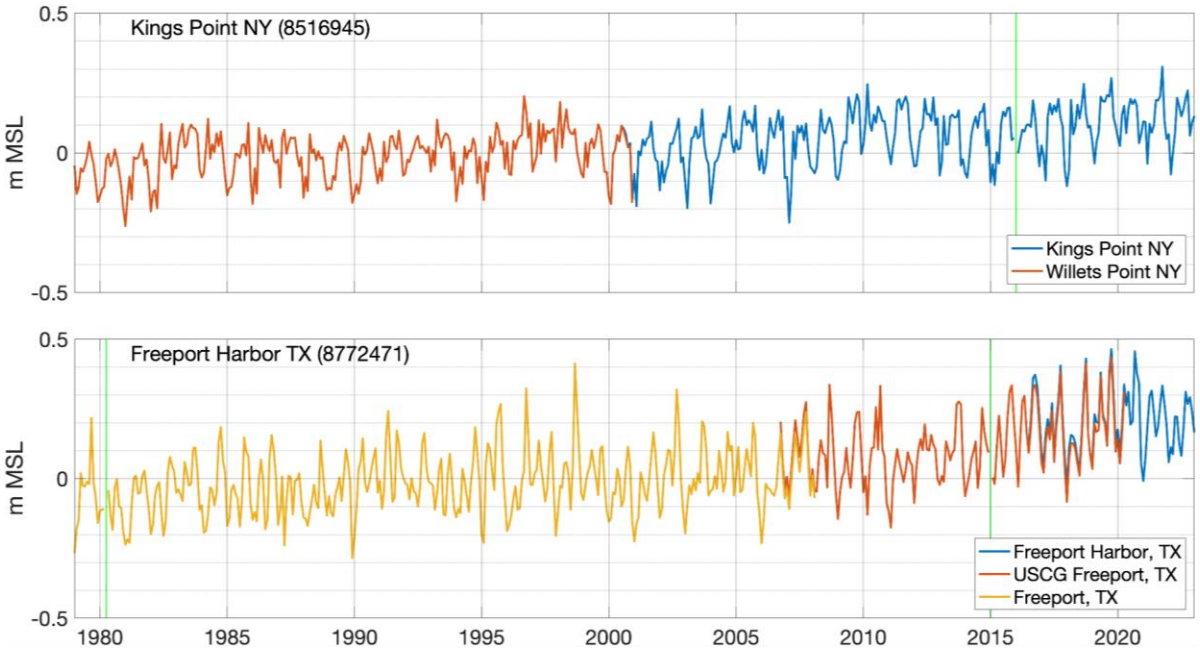

Figure 4. Example of merging data from nearby stations to extend the temporal record of each monthly mean time series. Green vertical lines indicate remaining gaps that are filled with the monthly mean from the record, adjusted for the long-term trend.

Handling of Different Tidal Datum Epochs

Of the 53 stations used for the assimilation, 2 stations are referenced to Modified 5-Year epochs, not the current 1983-2001 period. This is due primarily to rates of change of

long-term water level means. Stations Rockport, TX (8774770), and Grand Isle, LA (8761724), are referenced to the 2002/2006 and 2012/2016 NTDEs, respectively. These stations' monthly mean water levels were adjusted by 0.12 and 0.19 m, respectively, to put all stations onto the same 1983-2001 MSL reference level epoch. These offset values were determined by computing the average of the time series during the 1983-2001 epoch, thus effectively raising the water level to have approximately 0 mean during the current epoch. Figure 5 illustrates the results of this adjustment for Rockport and Grand Isle.

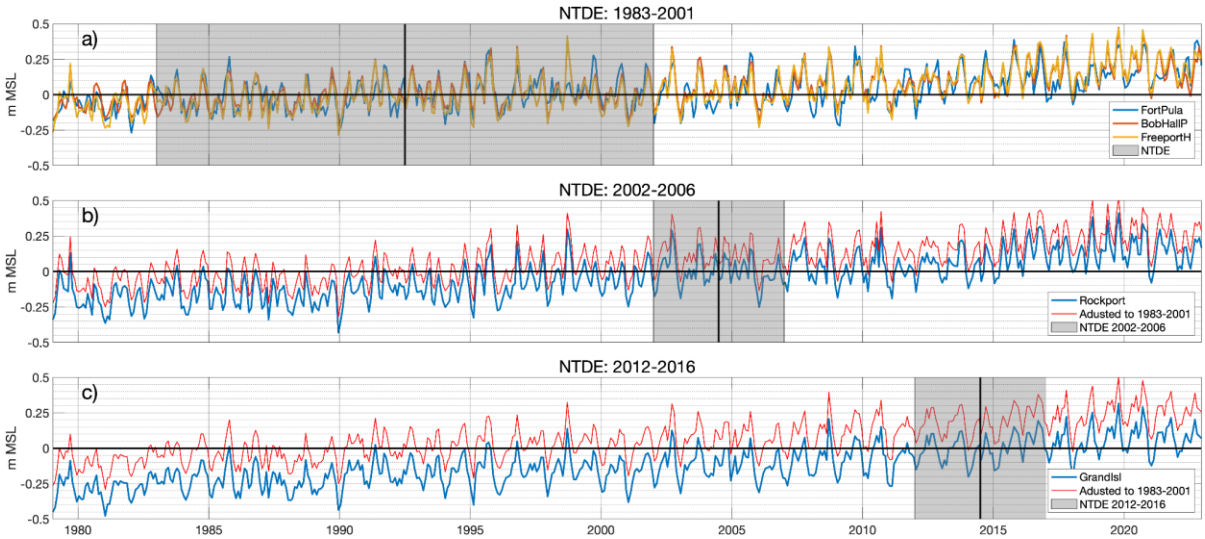

Figure 5. Monthly mean water levels for 5 stations used in the data assimilation process. The tidal datum epoch for most stations is based on the current National Tidal Datum Epoch (NTDE) period of 1983-2001. Panel a) shows 3 stations in the 1983-2001 NTDE. The average of the monthly means in this period is (by definition) close to 0. Panels b) and c) show Rockport, TX, and Grand Isle, LA, monthly mean water levels, respectively, relative to their respective Modified 5-Year tidal datum epochs and the adjusted time series (red) adjusted to the 1983-2001 NTDE.

The above 2 steps (the compositing of monthly mean time series and datum adjustments for Rockport and Grand Isle) resulted in a complete 44-year set of monthly means referenced to the same tidal datum epoch. The consequence of this complete record is that the monthly MSL component is completely defined and imposed as a part of the posterior prediction process. Figure 6 shows the adjusted and augmented monthly mean dataset. Features of note are the broad trends of long-term increasing sea level, the different characteristics of the monthly mean water levels in the Puerto Rico area (last 5 stations), and the along-coast/regional similarity of much of the variability.

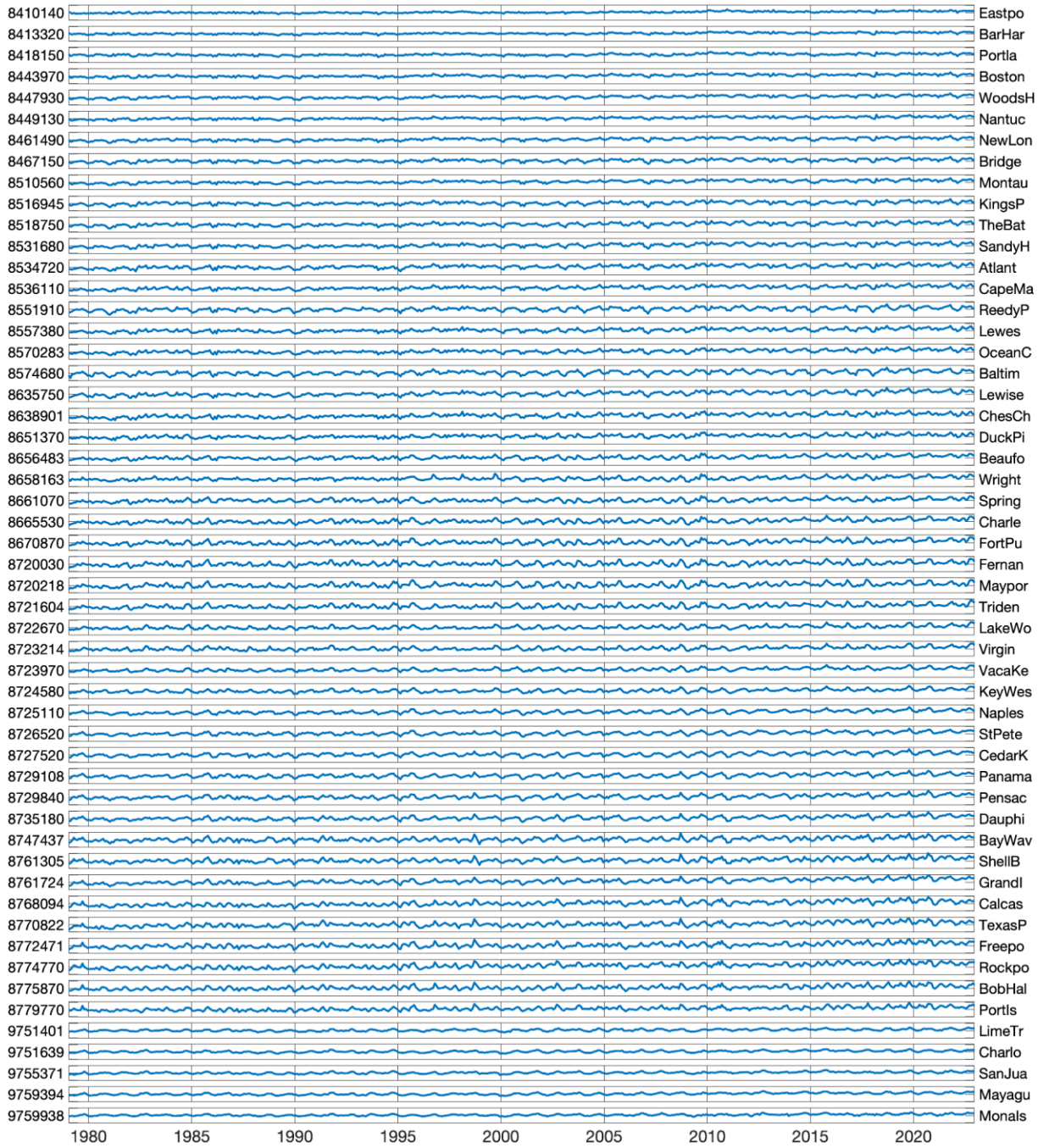

Figure 6. Monthly mean water levels for the 53 National Water Level Observation Network (NWLON) stations. Data records are completed over the 44-year period by merging nearby highly correlated time series.

Residual Time Series Processing

The next data processing step was to compute the residual water level by removing the monthly means from the hourly data. Small gaps (<24 hrs) in the hourly data were filled by linear interpolation. Larger gaps were not filled. The residual was low-pass filtered with a 4th order Butterworth filter (Roberts and Roberts 1978), with a cutoff of 36 hrs, to remove tidal variations, errors which we do not want to assimilate. Data near the edges of large gaps were eliminated. The final data preparation step was to augment the (filtered) hourly residual data

in a similar manner where nearby stations could be identified that were highly correlated. For example, the Chesapeake Bay station 8638863 (Chesapeake Bay Bridge Tunnel [CBBT], VA) was retired on 28 September 2017, and 8638901 (CBBT, Chesapeake Channel, VA) was installed in October 2016, approximately 8 km to the northeast (Figure 7) toward the center of the bay entrance. The 2 stations are in generally the same physical environment with very similar tidal and non-tidal characteristics; the correlation is very high (0.99) with a mean difference (active-retired) of -0.04 m. In this case, the 2 time series were merged by adding the mean difference to the retired station data.

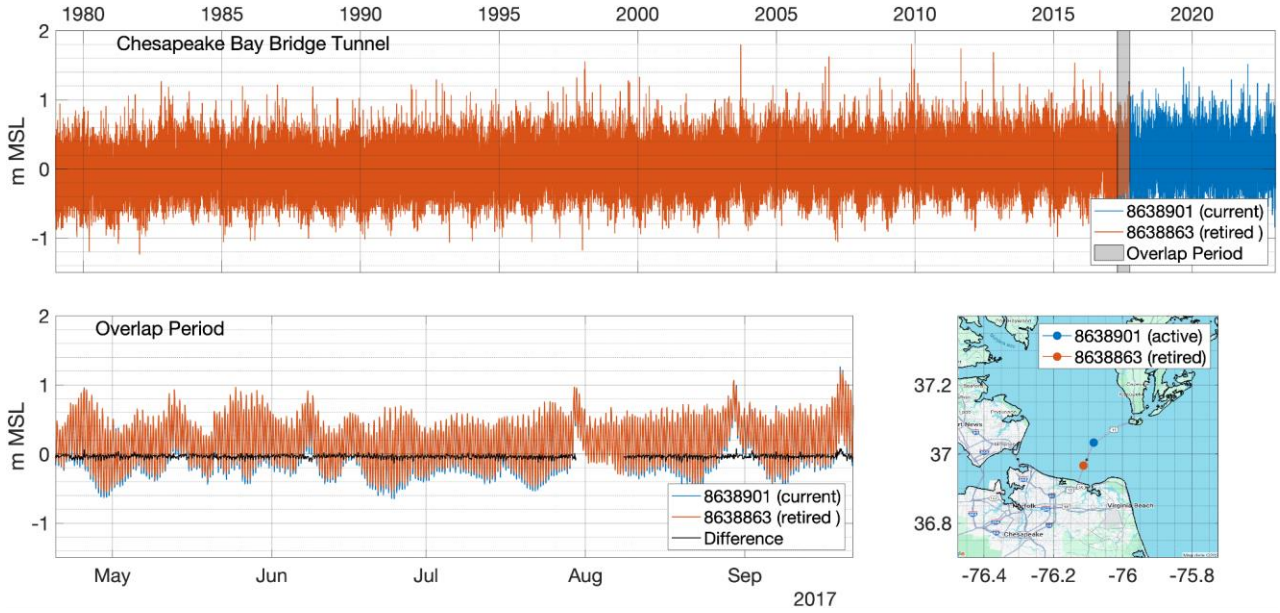

Figure 7. Water level time series at the 2 Chesapeake Bay Bridge stations (Station IDs: 8638901, 8638863). The top panel shows data over the Coastal Ocean Reanalysis (CORA) time period. The bottom panel shows data in the overlap period and the locations of the 2 stations in the lower Chesapeake Bay.

Figure 8 shows the station hourly water level availability for the study period after the merging of several stations similar to the processing of the lower Chesapeake Bay station. Nonetheless, large gaps still exist where companion stations could not be identified, notably Ocean City, MD; Wrightsville Beach, NC; Lake Worth Calcasieu Pass, LA; Mayaguez, PR; and Mona Island, PR. The consequence of these and smaller gaps are that the prior and posterior predictions, in the residual component of water levels, are the same since the error is assumed to be 0.

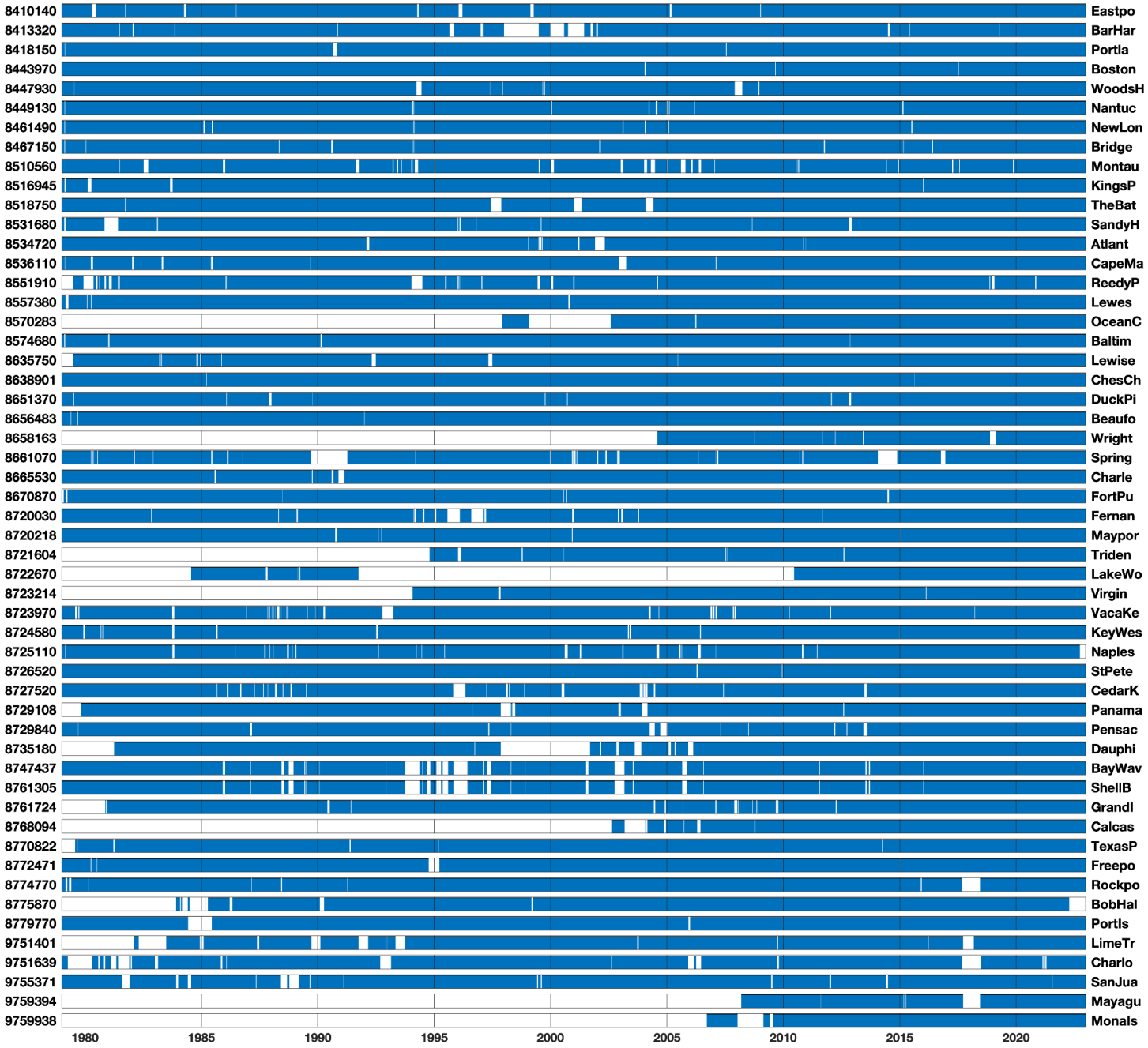

Figure 8. A plot of water level data availability for 53 stations used in the reanalysis process over the period of 1979-2022, with station IDs on the left y-axis and a station name abbreviation on the right. Dates from 1980-2020 are listed in 5-year increments with blue lines delineating completeness of each station's water level time series. Stations are listed from Maine to Texas to Puerto Rico (last 5).

Error Processing

At the end of each annual prior prediction simulation, the time series of water levels at the assimilation stations are processed in a manner identical to how the observations were treated (except that there are no gaps in the modeled predictions). Each station's prior

prediction is split into a monthly mean series and a residual series. The error is then computed as prior observations for each component, where negative errors indicate an under-prediction of the observations. Figure 9 illustrates the process for Fort Pulaski, GA (Station ID: 8670870), where each panel shows the residual and monthly mean water levels. Note that the hourly data is not shown for clarity. Observations, prior prediction, and the prior error are shown from top to bottom. Several important features are to be noted. There is a clear sea level rise signal in the observed monthly means (upward long-term trend, with larger increase from about 2010 forward). The monthly mean in the prior prediction (middle panel) has clear seasonal variability but no long-term trend. This is consistent with the forcing in the prior simulations being only tides and ERA5 meteorology. The prior error (difference between middle and top panels) thus shows the long-term sea level component as a negative error.

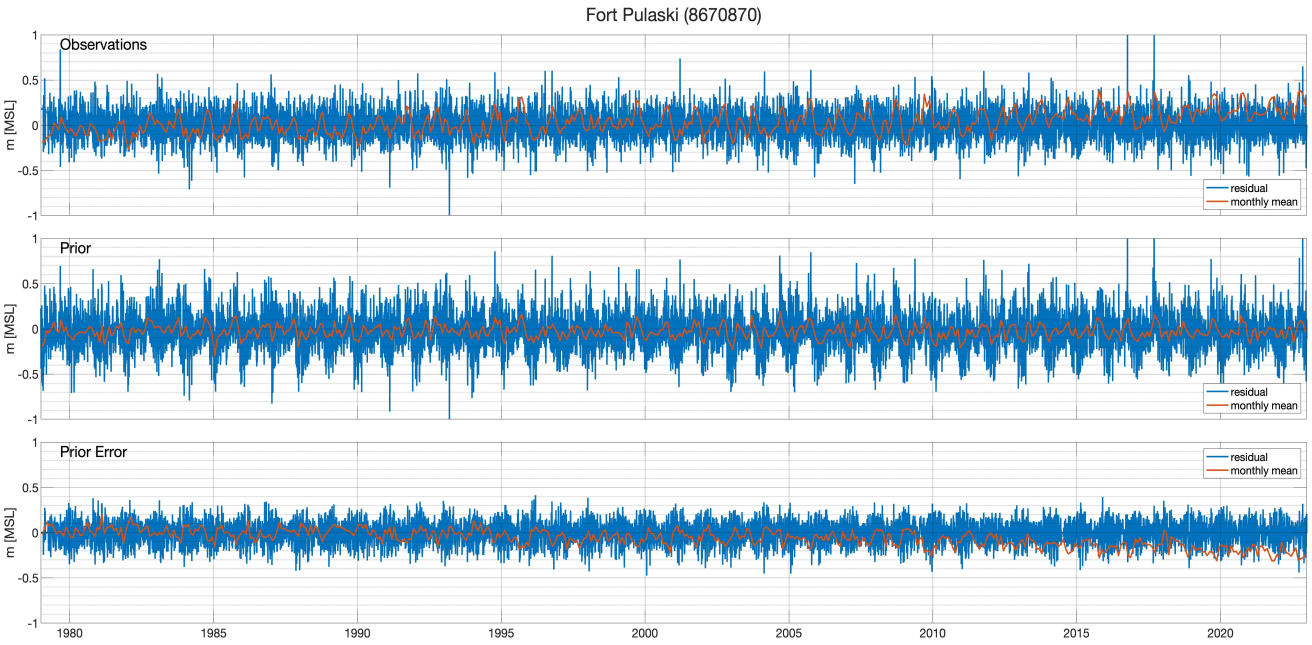

Figure 9. A time series of monthly means and residual water levels at the Fort Pulaski, GA, station (Station ID: 8670870). Observations, prior prediction, and prior errors are shown from top to bottom, each panel showing residual and monthly mean water levels. For clarity, raw hourly data are not shown.

The assimilation of errors into the posterior prediction occurs by specifying a dynamic water level correction (DWLC) field at each model node that accounts for the errors at each assimilation station every 6 hrs of model time (ADCIRC linearly interpolates between the error surfaces to the current model time). For each component of the error (monthly mean and residual), the errors for all assimilated stations are mapped onto the ADCIRC grid nodes using radial basis functions. The surface interpolation is constrained to be 0 offshore. Figure 10 shows the error surfaces on 14 September 2018 at 00Z. The sum (right) is the field used in the posterior prediction simulations. At this specific time, the monthly mean error surface is negative everywhere, indicating that the prior prediction lacks the expected long-term sea level increase. Note also that the surface is very smooth in the along-coast direction, consistent with the relatively large scales associated with monthly mean water level variability. The residual

component exhibits much higher spatial variation, with areas of prior over-prediction on the upper middle Atlantic Bight and along the eastern Texas coast.

Note that for all stations, the monthly mean prior error is defined for all times, resulting in the consistent incorporation of the long-term sea level signal. However, for stations with large gaps in the hourly data, the gaps persist in the residual error, meaning that no residual correction is made when and where these gaps exist.

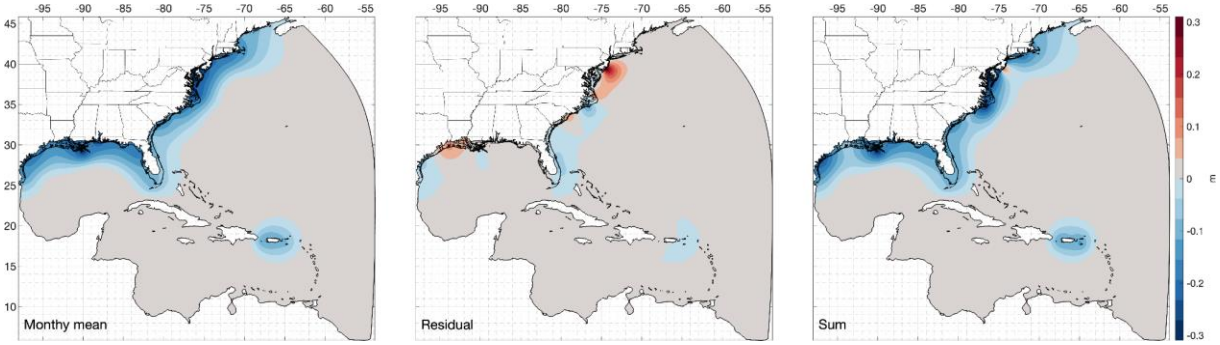

Figure 10. Error surfaces for the monthly mean (left), the residual (middle), and their sum (right) for 14 September 2018 at 00Z. The sum is the dynamic water level correction applied in the posterior prediction simulations.

Before the details of the statistical characteristics of the prior and posterior predictions are covered in the Water Level Results section (below), we show here the impact of the data assimilation process on the prior and posterior error. Figure 11 shows the prior and posterior errors at Fort Pulaski. The posterior errors are computed in the same manner as the prior errors by computing the posterior monthly mean prediction (for example) and subtracting off the observed monthly mean. The top panel shows the monthly mean prior (blue) and posterior (red) error. The prior error (same as Figure 9 above, lower panel) becomes more negative over time due to the long-term sea level component missing in the prior prediction, but the assimilation of the prior error into the posterior prediction has accounted for most of this error component. In the residual component, both the prior and posterior errors are relatively unbiased, but the variance in the posterior error has been substantially reduced, from about 11 cm to 1 cm.

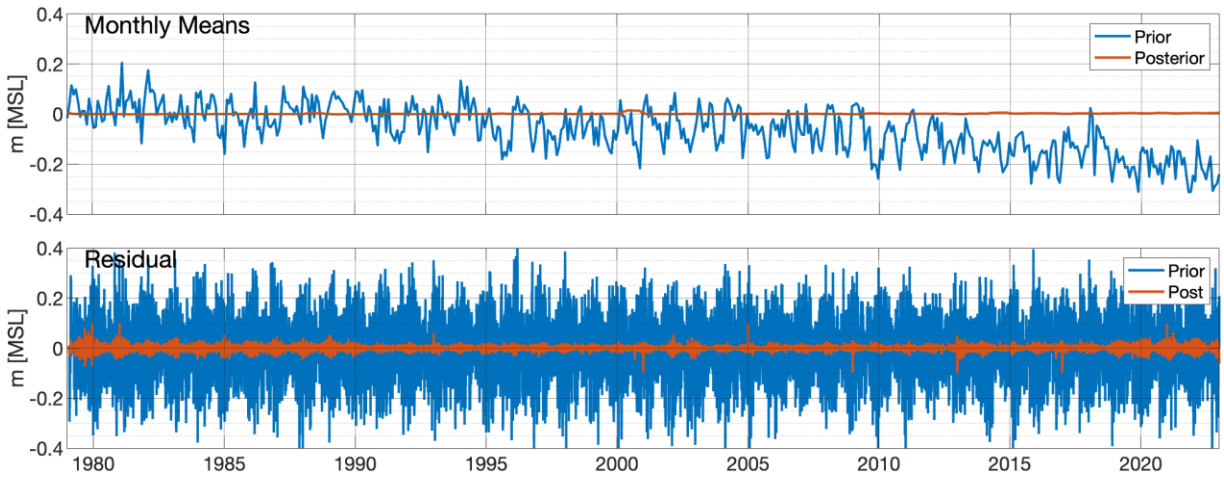

Figure 11. Time series of monthly mean and residual water level errors (model minus observations) at the Fort Pulaski, GA (8670870), station. Monthly means and residual errors are shown in the top and bottom panels, respectively, and each panel shows the prior (blue) and posterior (red) errors. For clarity, hourly raw errors are not shown.

5. WATER LEVEL RESULTS

Observation Data

A total of 112 NOAA stations (Figure 12) were used to assess model performance, of which 53 stations were used in assimilation, as described above, and 59 stations were used for validation. There are many differences between the 2 sets of stations, an element important to consider when evaluating model performance. Stations used in assimilation are generally at or near the open coast to better represent conditions across a broader geographic range. Conversely, unassimilated stations tend to be in more sheltered areas, like estuaries or rivers. Such locations tend to be tougher to get good results at than the open coast because of local effects like the accuracy of local topobathy, sheltered winds, or rainfall. Many validation stations are concentrated in the Chesapeake Bay area.

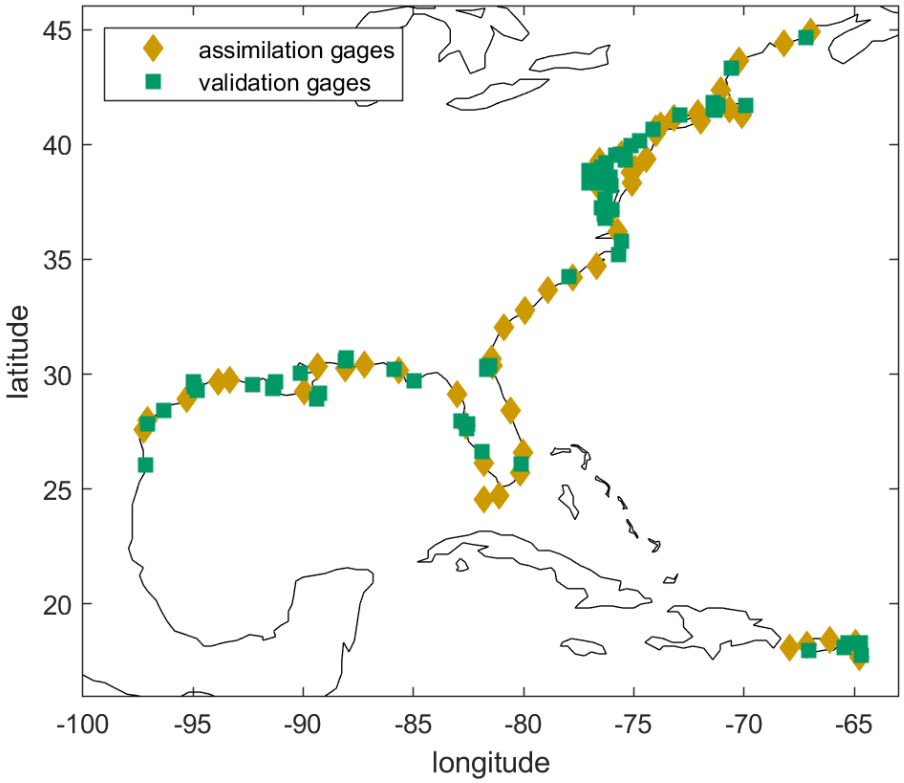

Figure 12. Location of NOAA National Water Level Observation Network (NWLON) stations used to assess model performance.

Stations also have very different temporal coverages. Figure 13 shows the fraction of hourly data from 1979-2022 that is absent for each station. There are a total of 28,868,587 observations[18] across 112 stations, equaling close to 3293 years of hourly data. Most missing data is from time periods before a station was first installed, though some stations also have substantial gaps in the middle of their records. Data is especially sparse in the western Gulf of Mexico.

[18] To calculate this, a value is counted as missing if either an observation is missing or the modeled value is missing (i.e., the model is "dry" at that point in time). However, the model is dry at stations <0.001% of the time, so almost all the missing values are due to missing observations.

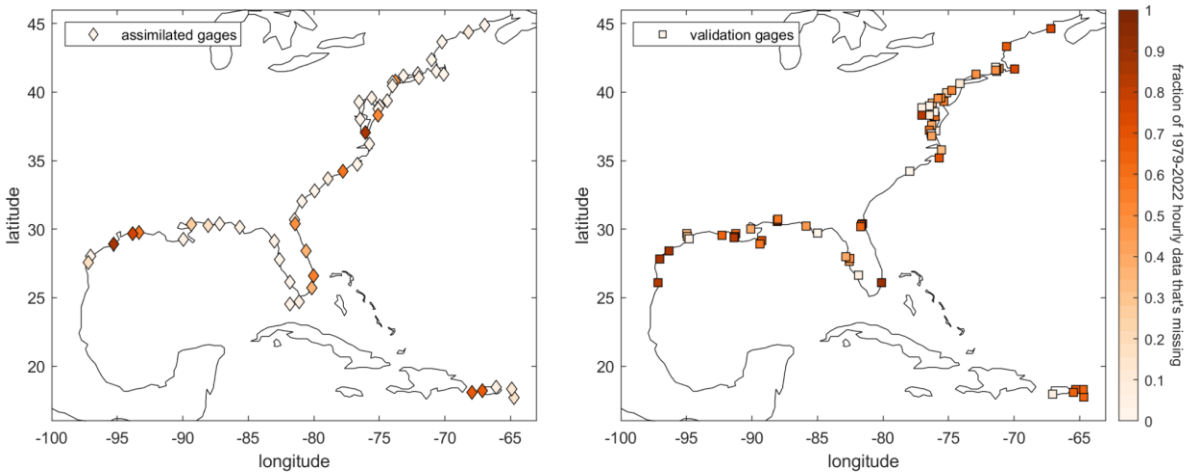

Figure 13. Fraction of missing data from 1979-2022 at each station: 0 indicates full data coverage over the 1979-2022 time period, and 1 indicates no data at all. The left map shows assimilation sites and the right map shows validation sites.

Model Performance

To assess the model, simulated water levels were extracted at or near station locations, interpolating the solution on the selected triangular element of the mesh as is consistent with the finite element formulation of the model. Error distributions are shown in Figures 14 and 15, and Table 3 shows summary statistics. Overall, errors are generally less than 20 cm for the baseline unassimilated model and are reduced by 30-60% in the assimilated model. Errors are about 2 cm larger at validation sites than assimilated sites in the unassimilated run, which may result from these being locations where water levels are harder to predict as noted in the previous section. Validation site errors are also larger in the assimilated run, by 2-3 cm, though the increased error here is also likely driven by the assimilation not being as accurate at these sites. Looking at the mean and standard deviation (STD) columns in Table 3, there is also a reduction in the variational error due to assimilation though the mean errors are substantially reduced. This pattern is also visible in Figure 14, where the breadth of the errors narrows and the low bias in the unassimilated simulation (primarily a result of sea level rise) is almost completely removed in the assimilated simulation. It should be noted that the slight increase in small positive errors in Figure 14 is a natural result of the bias removal and not caused by some tendency of the assimilation to overpredict.

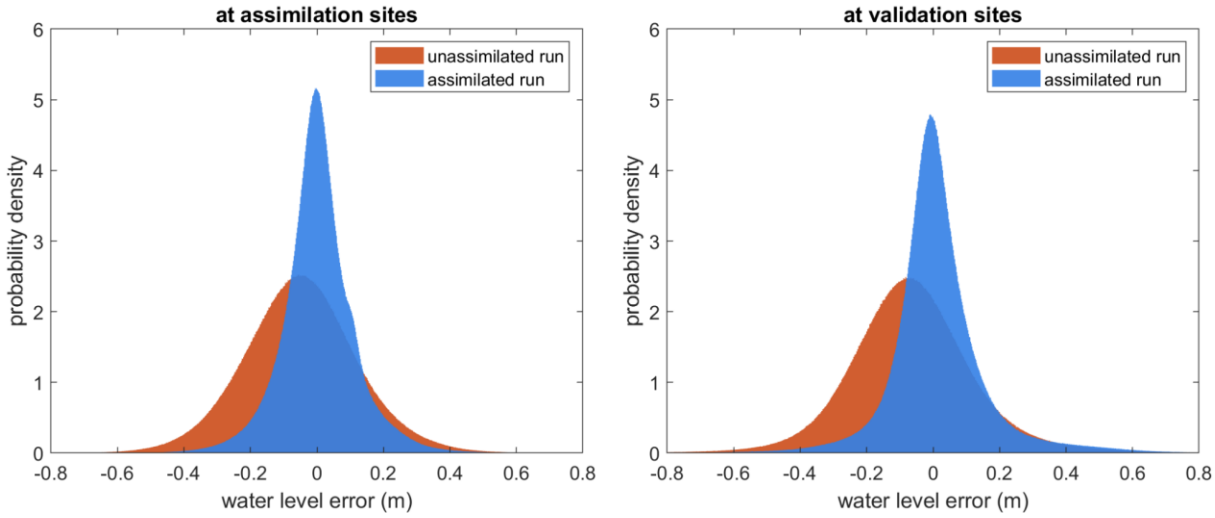

Figure 14. Distribution of errors for all hourly water levels. Shown for unassimilated and assimilated runs at both assimilation sites and validation sites.

Errors are shown geographically, with STD represented in Figure 15 and mean absolute error (MAE) in Figure 16. Geography also contributes to error trends, like the Gulf of Maine where errors are higher due (at least partly) to large tidal variations. Conversely, errors are particularly small in Puerto Rico and the U.S. Virgin Islands thanks to an especially small tidal range and nearly no shelf. Errors increase steadily up the Delaware River, likely due to a lack of riverine inflow and because mesh resolution does not support extending the river far into the domain.

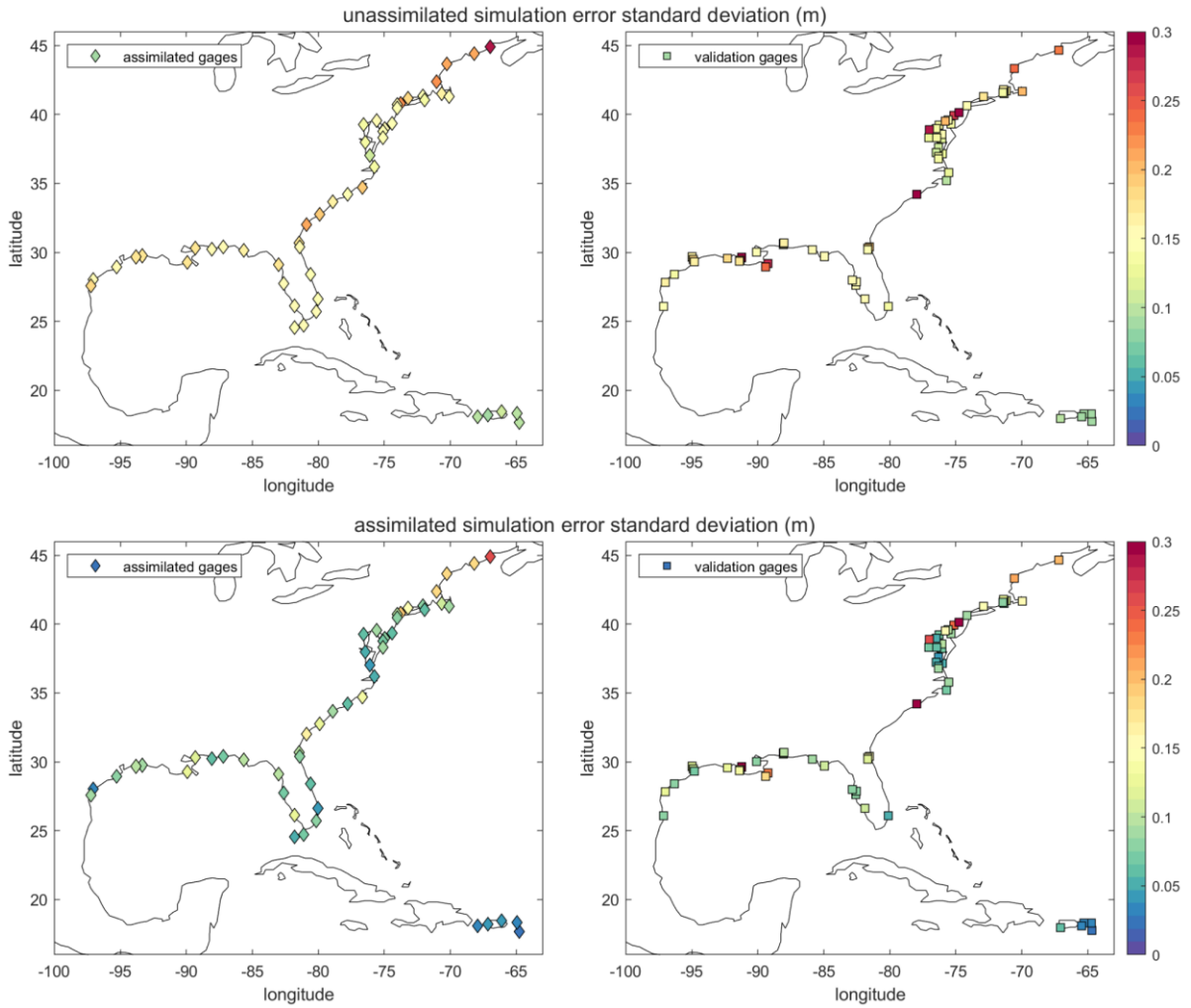

Figure 15. Standard error (STD) in meters (m) at each station. Maps on the left show assimilated sites, and maps on the right show validation sites. The top row of maps show results for unassimilated simulations, while the bottom row of maps show assimilated sites.

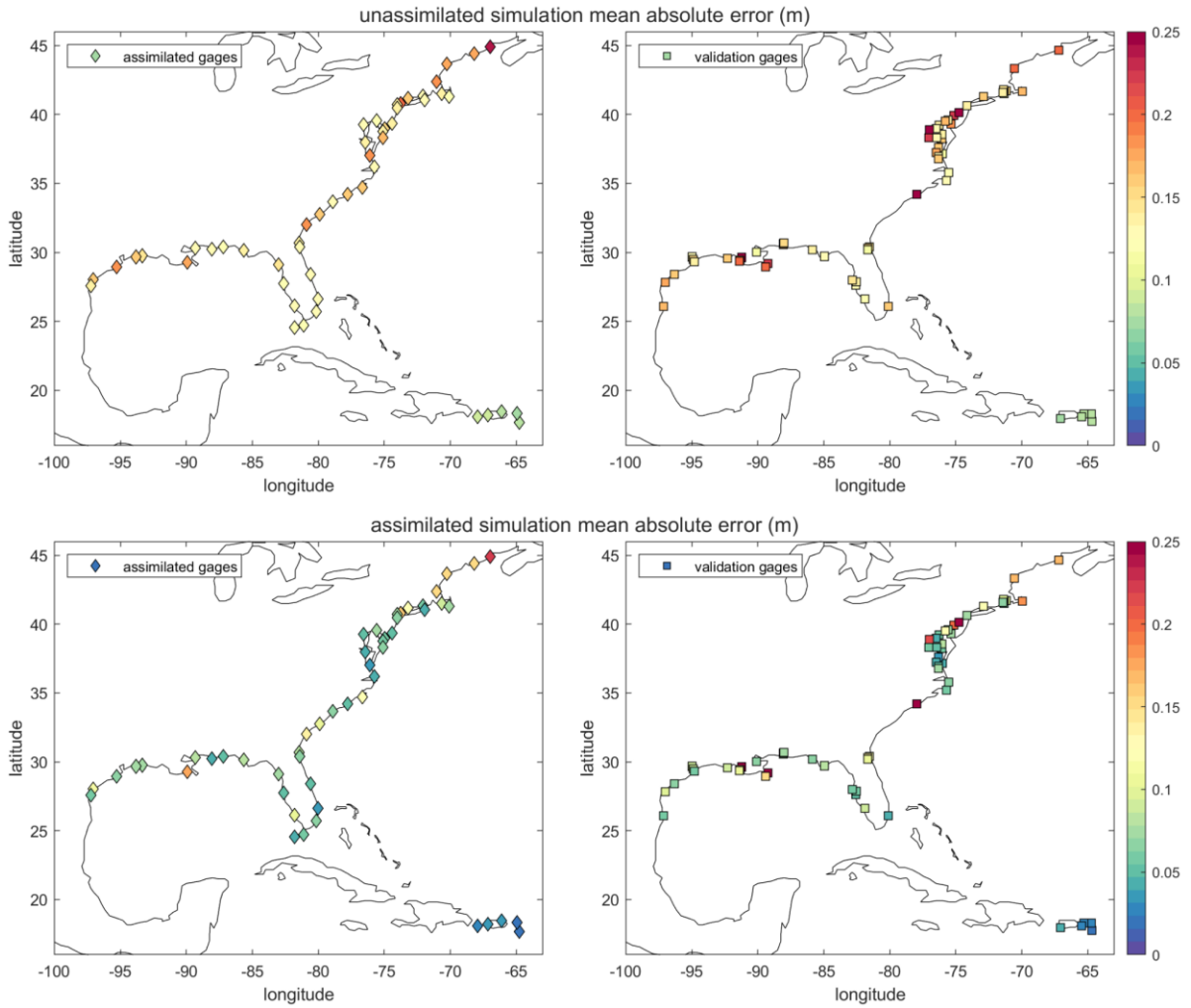

Figure 16. Mean absolute error (MAE) at each station. Maps in the left column show assimilated sites, and maps on the right show validation sites. The top row of maps shows results for the unassimilated simulation, and the bottom row of maps shows assimilated simulation.

Geographic patterns depend on how the data is viewed, however, as demonstrated in Figure 17. Here, the STD of error at each station has been normalized to that station's STD of observed water levels. This shows that normalized errors are much smaller in areas dominated by more easily predicted tides; meanwhile, normalized errors are quite large for the unassimilated simulation at sites whose variability is strongly driven by seasonal water levels and sea level rise. Viewing results this way also strongly demonstrates the benefits of assimilation: errors exceed 70% (0.7 in the figures) at almost all stations in the Gulf and on islands before assimilation, but after assimilation, they are generally less than half that.

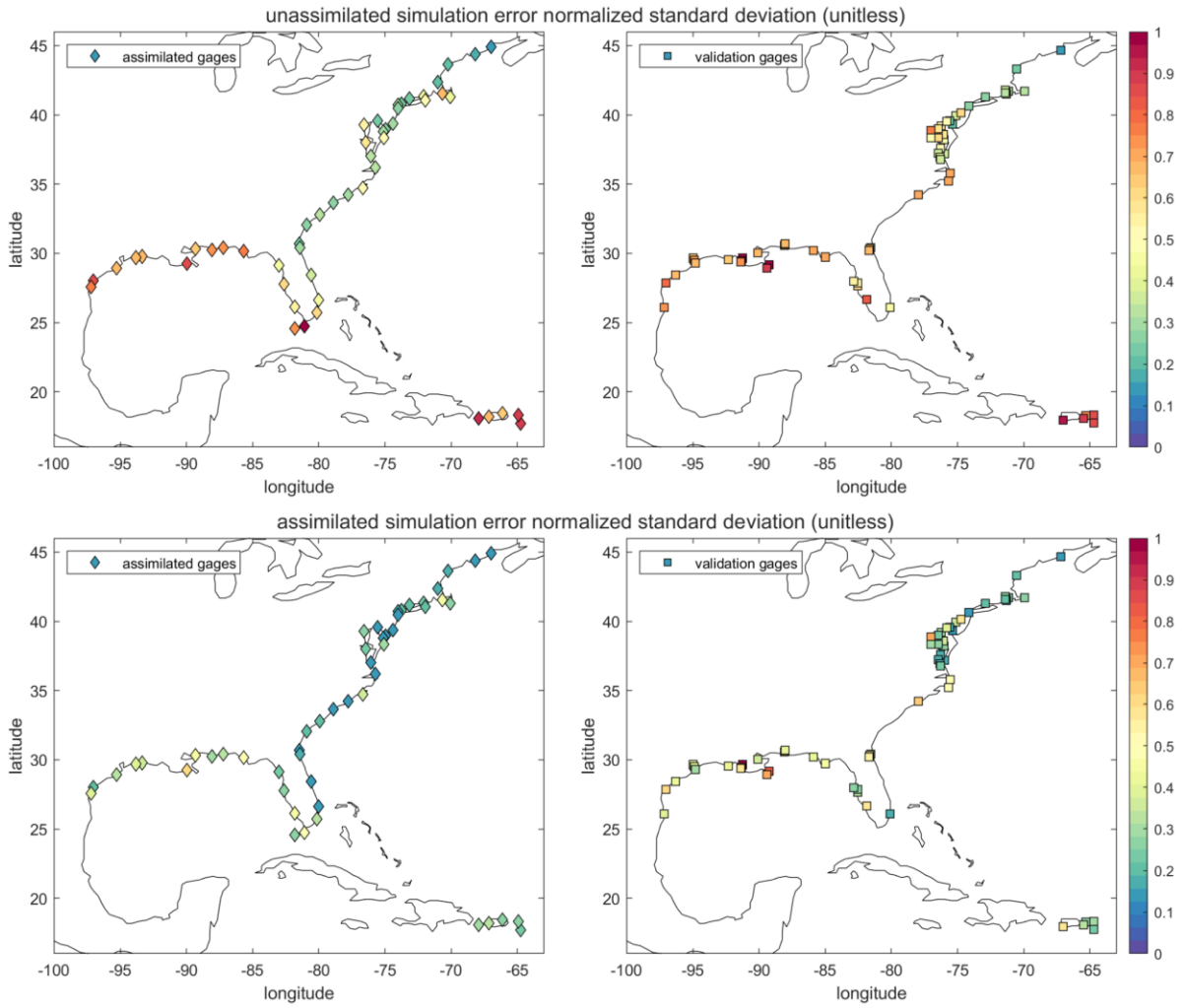

Figure 17. Standard deviation (STD) normalized to the STD of observed water levels at each station. The left column shows assimilated sites, and the right column shows validation sites. The top row shows results for the unassimilated simulation, and the bottom row shows assimilated sites.

Table 3. Summary error statistics for all stations based on hourly data from 1979-2022. RMSE = root mean square error; MAE = mean absolute error; STD = standard deviation.

errors (m)		**RMSE**	**MAE**	**mean**	**STD**
unassimilated simulation	all	0.187	0.151	-0.067	0.166
	assimilated	0.179	0.140	-0.050	0.172
	validation	0.205	0.155	-0.065	0.194
assimilated simulation	all	0.110	0.089	0.009	0.106
	assimilated	0.111	0.080	0.004	0.111
	validation	0.150	0.097	0.012	0.150

As noted in the Observation Data subsection, the differences in data availability through space and time can complicate interpreting model results. As an example of this, the top left pane of Figure 18 shows averaged MAE across all (available) stations through time as well as the number of stations available for the assimilated simulation. Data have been run through a 30-day moving mean window to make the figures less noisy (note that "high-frequency" errors are deliberately not completely removed since the MAE is used). Errors are noticeably higher for the first ~3 years and again increase beginning in the early 2010's. When we subset the data as shown in the other panes, these patterns change. Looking at the number of available stations, we can see a strong correlation with the changes in apparent errors. This implies that at least part of the change in average error is due to new stations coming online. What is consistent in these plots is that errors have a seasonal component such that they are highest in mid-to-late January and lowest in the summer months. This may be related to increased storminess in winter months and, at least at validation sites, to errors in seasonal water levels.

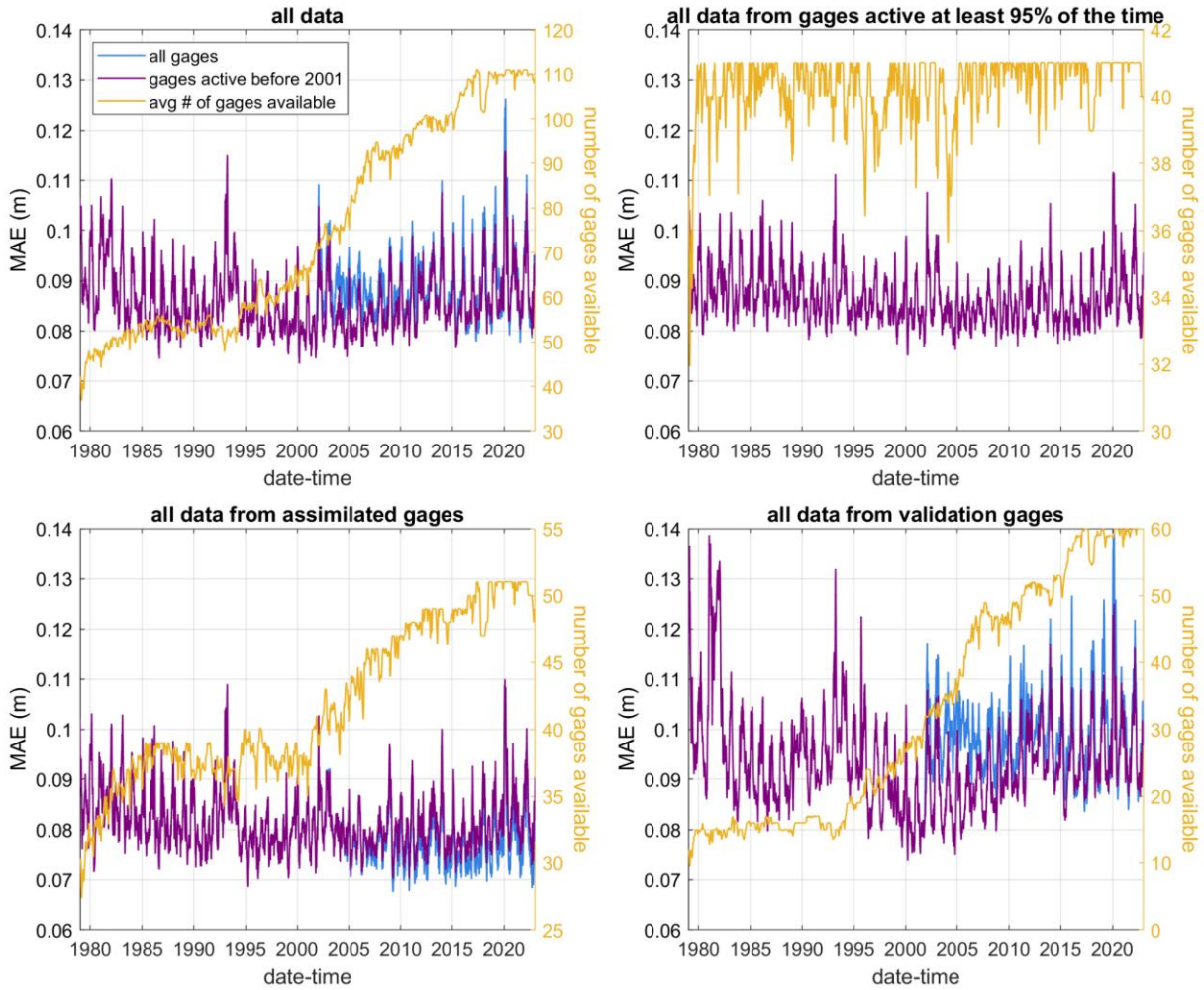

Figure 18. Averaged mean absolute errors (MAEs) in the assimilated simulation over time for all stations and subsets of them, as noted in the subtitle of each pane. In all panes, yellow lines show the number of active stations. Purple lines indicate stations active before the year 2001, and blue lines denote all stations.

Spectral Errors & Effects of Assimilation

Spectral analysis helps illustrate the different components of the water level signals. The power spectra[19] for select Atlantic locations are plotted in Figure 19, while Gulf locations are shown in Figure 20. In all cases, the correction signal closely follows the error from the unassimilated (i.e., prior) simulation, as is expected, and the error in the assimilated (i.e., posterior) run drops off significantly around the correction's cutoff frequency of 1.5 days. This shows that it adeptly removes low-frequency errors.

Spectral signals for the unassimilated simulation and the correction itself help to illustrate the effects and relative importance of correction, which varies spatially and with frequency. This can be seen in the figures by comparing the magnitudes of the correction and the unassimilated water level and determining which of these more closely follows observations. The spectra can be characterized by 3 distinct regions in the Atlantic (Figure 19). In the Gulf of Maine, which includes northern Massachusetts, the correction's effects are key for periods longer than 5-10 days. South of Cape Hatteras, NC, unassimilated water levels do poorly for periods longer than 1-3 days, although there is some skill for Georgia and Florida stations. For most of the mid-Atlantic Bight (e.g., Connecticut to upper North Carolina), the unassimilated simulation is fairly accurate for periods up to 90 or even 180 days. For the lowest frequencies, correction becomes more influential across the entire Atlantic coast. Similarities between observations and unassimilated water levels across a wide frequency band indicates that low-frequency simulations are driven by surface winds and pressure.

Spectra for the Gulf of Mexico are plotted in Figure 20 and show similar patterns to the Atlantic, emphasizing the role of meteorology for fairly long-period signals, though correction tends to dominate for periods beyond 30-60 days, perhaps due to stronger steric effects. Gulf sites also have 2 erroneous spectral humps, centered roughly on 35 and 16 hours. These correspond to naturally excited modes in the Gulf of Mexico, and testing indicates they are due to reflections off the open boundary along the eastern (open ocean) edge of the mesh. An inverted barometer boundary condition along the open boundary was implemented to help limit this spurious energy; the hump was larger in preliminary versions of the reanalysis. The assimilation tamps down some of the remaining error introduced by the 35-hour resonance. It is worth noting that an earlier version of the reanalysis that used a 24-hour cutoff frequency for the correction also had limited success in modulating the 35-hour resonance. This implies that the phases of the resonances changed in response to the water level correction, and so they may be driven in part by the correction.

In Puerto Rico and the U.S. Virgin Islands, at periods longer than ~2 days, there is about an order of magnitude more spectral energy than in observations for the unassimilated simulation. The cause of this is unknown; however, errors are removed during correction, and the assimilated solution closely matches observations.

At higher frequencies where the correction does not come into play, the unassimilated and assimilated runs are nearly identical, as are their errors. The model does well representing the primary tidal harmonics at most sites, as is to be expected based on the original validation study (Riverside and AECOM 2015). However, it does underestimate some harmonics, particularly some of the lesser constituents like the J1.

[19] To ensure a fair comparison between all signals, any time there is a gap in data in either the observations or the model, both were set to 0 prior to spectral analysis. This has the side effect of "smearing" some of the spectral energy, though this effect is barely visible in the figures and does not influence conclusions.

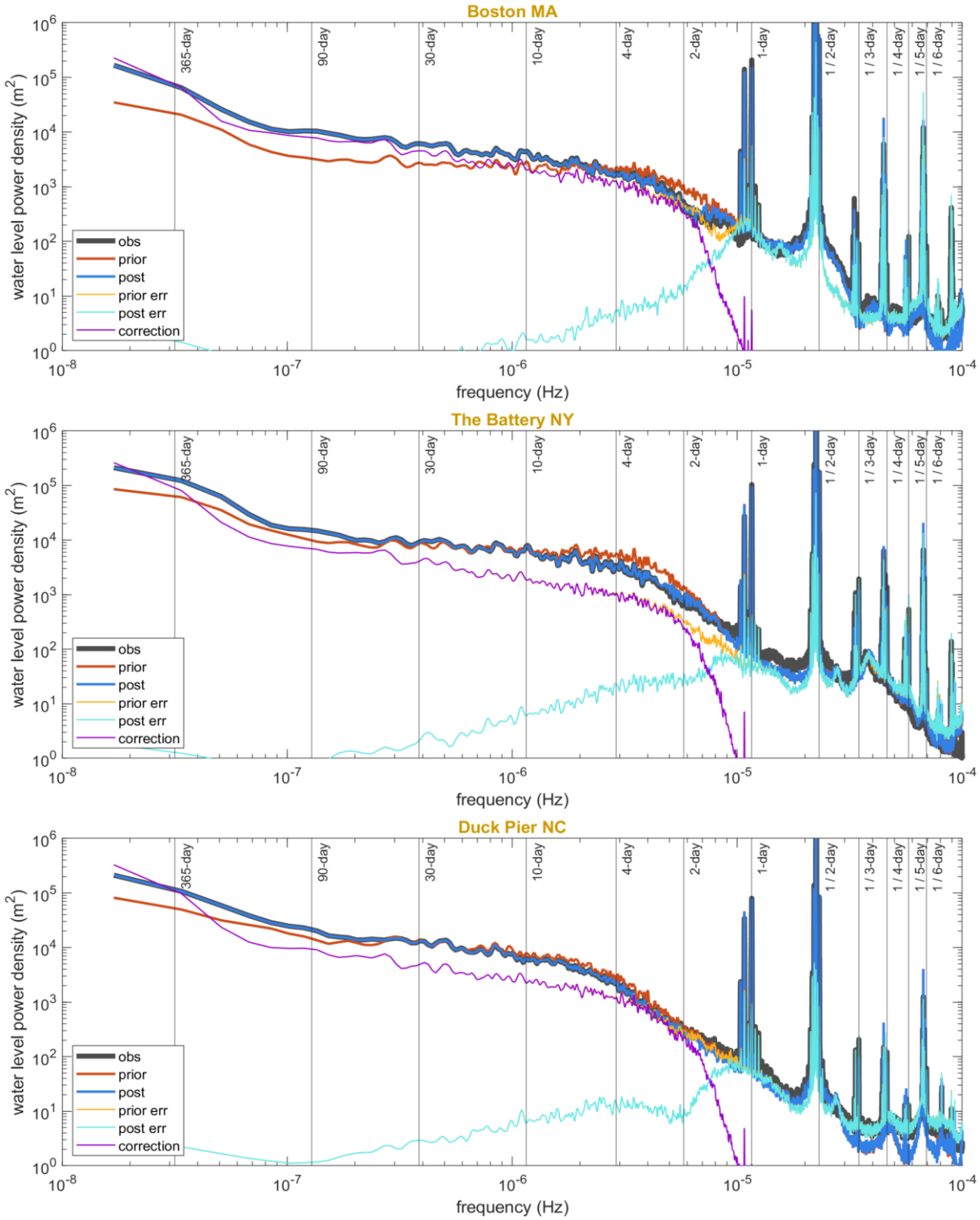

Figure 19. Water level spectra for Atlantic locations. Labeled vertical lines in the plots indicate the number of days corresponding to certain frequencies. Thick blue lines represent observations. Red lines represent prior simulations, and blue lines represent posterior simulations. Model error is noted in yellow for prior simulations and light blue for posterior simulations. Corrections are noted in purple.

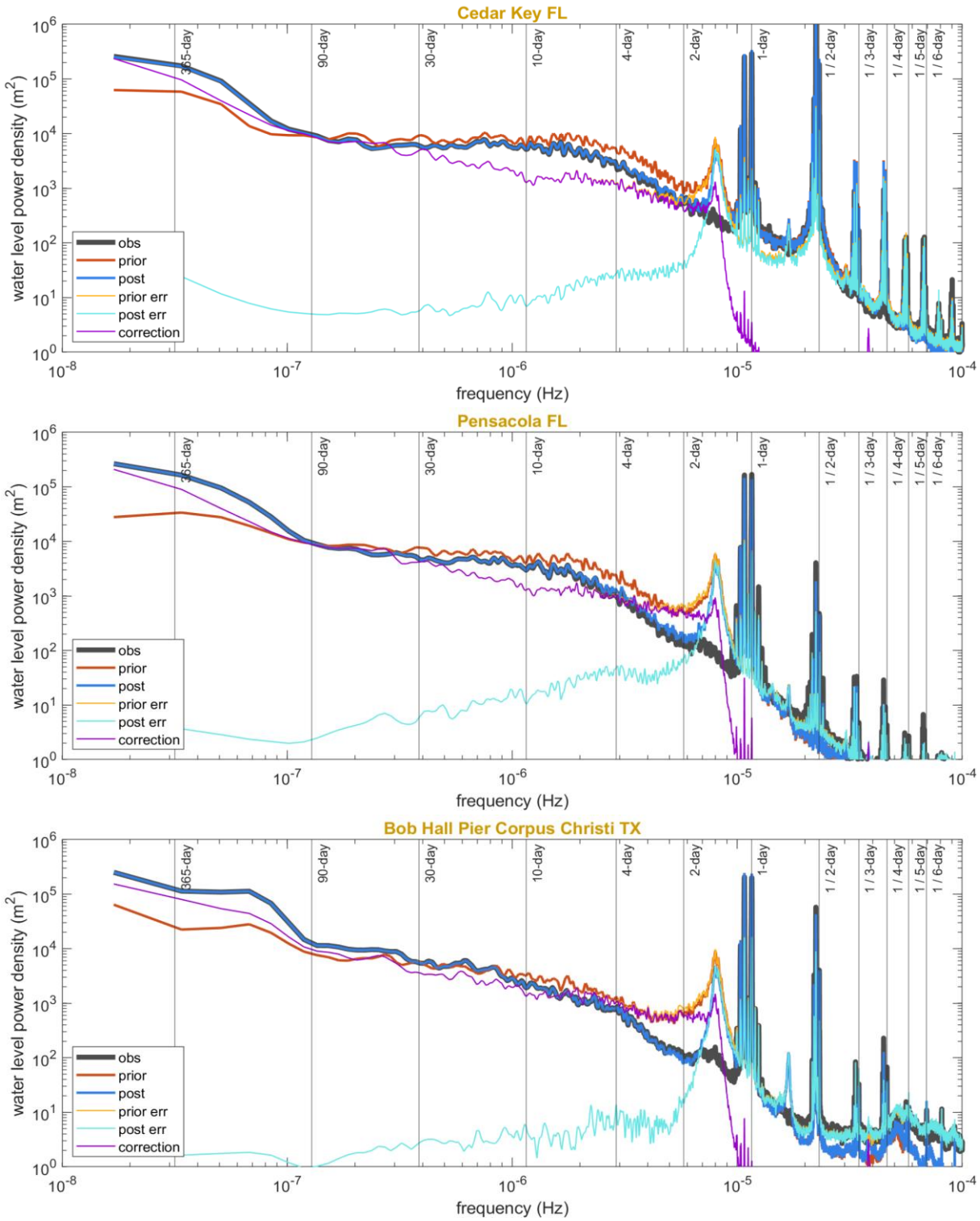

Figure 20. Water level spectra for Gulf of Mexico locations. Labeled vertical lines in the plots indicate the number of days corresponding to specific frequencies. Thick blue lines represent observations. Red lines represent prior simulations, and blue lines represent posterior simulations. Model error is noted in yellow for prior simulations and light blue for posterior simulations. Corrections are noted in purple.

Extreme Value Analysis

Much of the analysis thus far has focused on bulk statistics and other averaged properties, which, while informative, do not help understand the model's performance under more extreme cases. Since such extremes, particularly major surge events, are of particular interest, this section presents an extreme value analysis to illuminate errors under these circumstances. Observed and modeled water levels are taken as-is, without detrending or other adjustments[20]. A block maxima approach, specifically annual maxima, was taken, and the generalized extreme value (GEV) distribution was fit to these data. As in other analyses in the Water Level Results section, only data for the 1979-2022 period are used, and whenever observed or modeled data are missing, the other dataset's values (for the same time period) are also treated as missing. This ensures a fairer comparison between the model and observations and is particularly important since stations are often damaged during extreme events. Stations with less than 10 years of data were excluded to avoid poor fitting to small amounts of data. Note that these methods are somewhat simplistic, as this work is only meant to be illustrative. For instance, better methods might use more robust fitting techniques, evaluate other approaches like peaks-over-threshold, address stationarity concerns, consider seasonality and the existence of subpopulations, and address missing data, among other things.

Results at select stations are shown in Figure 21. Overall, the modeled results closely follow the observations, and the assimilation improves results. Importantly, with assimilation, improvement is seen in both the individual peak water level estimates and the fitted distributions. Aggregated results at the 5- and 10-year levels in Figures 22 and 23 show that across almost all sites, estimated extreme water levels are improved compared to the unassimilated run. However, there is a small low bias compared to observations, averaging -4.6 cm for the assimilated estimate of the 5-year water level and 6.3 cm for the 10-year.

[20] Due to sea level rise, this violates the stationarity assumption of typical extreme value analyses, but the sea level rise signal is smaller than other trends that cannot be readily removed, such as interannual trends in sea levels and storminess.

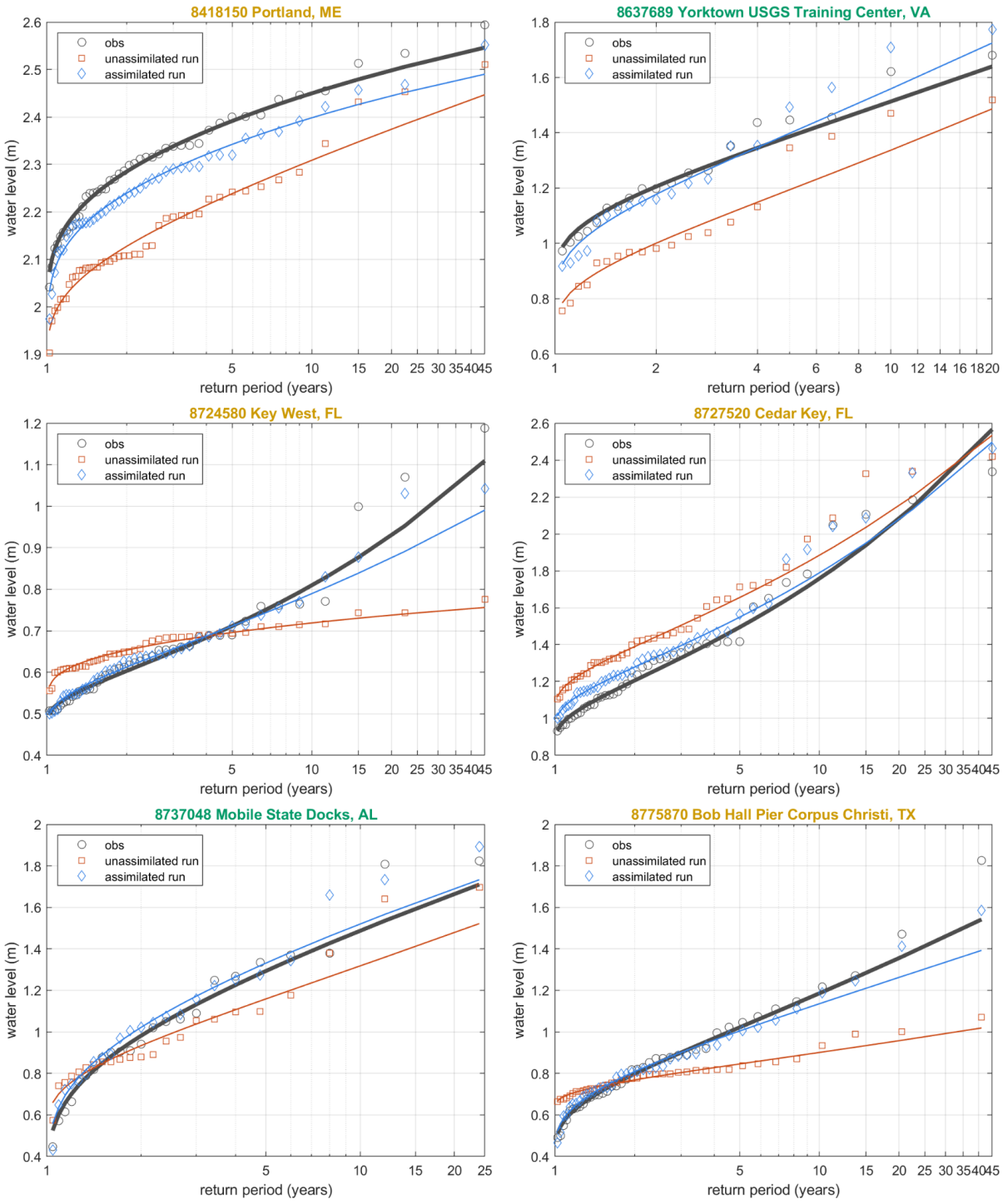

Figure 21. Extreme value plots at select stations. Green titles indicate validation stations, yellow titles indicate assimilated stations. Empirical data plotted using the Weibull plotting position formula n/(N+1).

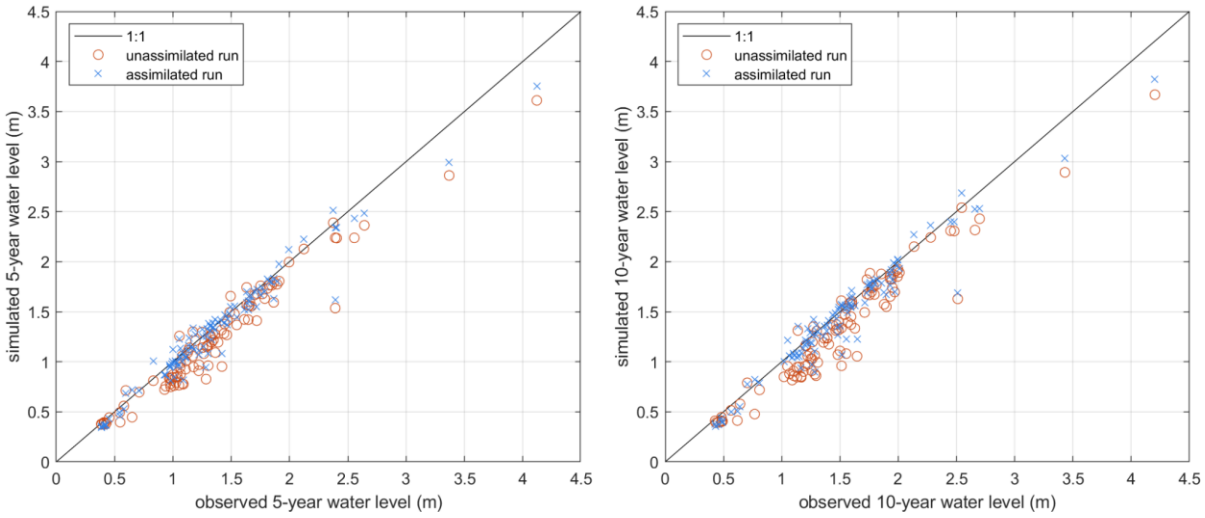

Figure 22. Comparison of observed and simulated estimates of 5- and 10-year water levels at all sites.

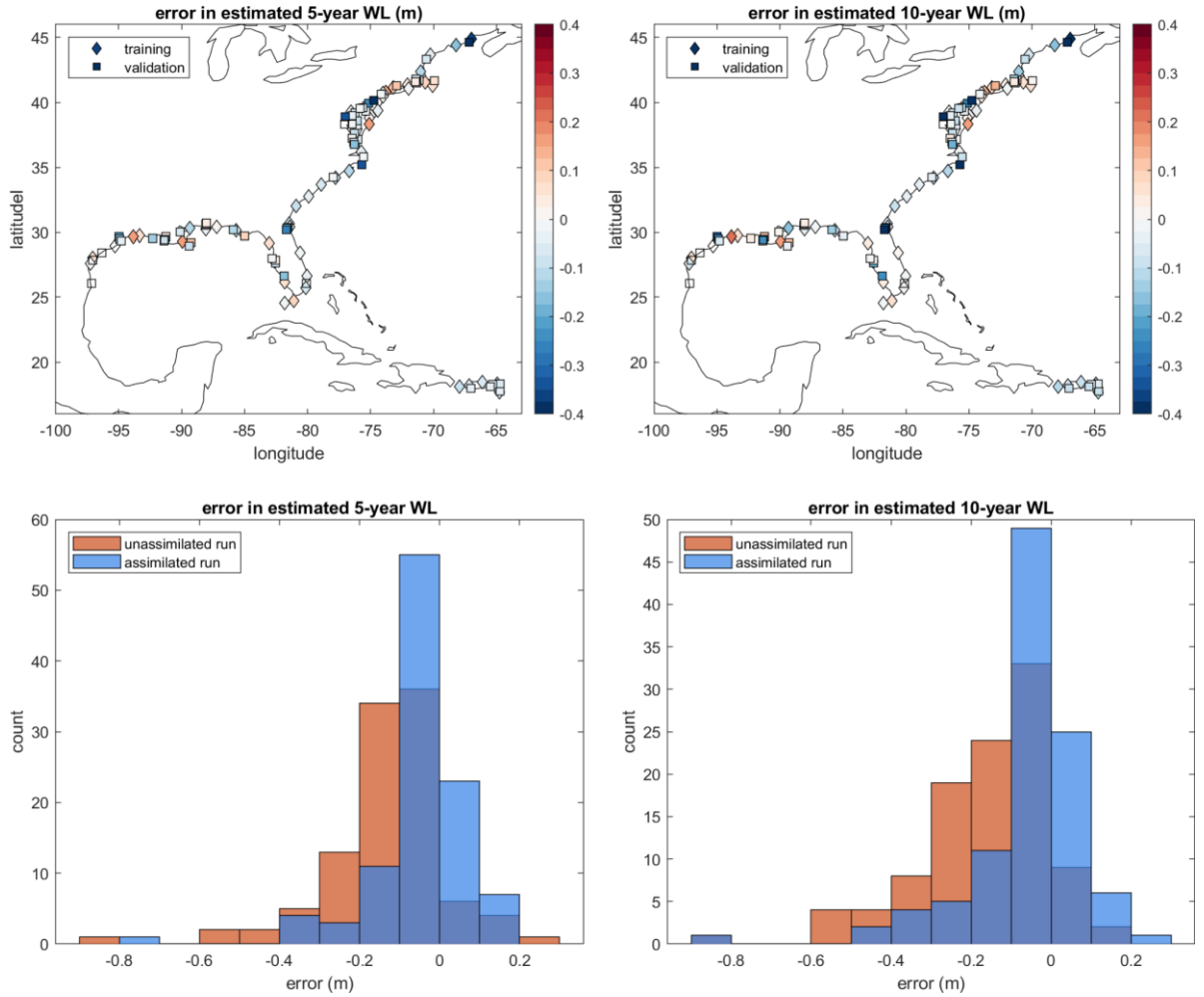

Figure 23. Errors in estimated return period water levels. The top row shows geographic distributions for the assimilated run, and the bottom row shows histograms. For the histograms, red denotes unassimilated results, and blue denotes assimilated.

Limitations

The meteorological forcing for this coastal water level reanalysis is from ECMWF's ERA5. At a spatial resolution of 0.25°, its representation of tropical cyclones tends to be less intense than observed (e.g., Hodges et al. 2017; Dulac et al. 2024), particularly for peak winds associated with stronger and/or more concentrated storms. For instance, peak surges from Hurricane Sandy are well captured by the model because, although strong, Sandy was a very broad storm, with a radius of maximum winds (RMW) about 4 times ERA5's ~30-km resolution. The highest surges in Hurricane Katrina, however, are underestimated. Katrina's winds were much stronger than Sandy's, and its RMW was considerably smaller. Hurricane Ike falls somewhere between Katrina and Sandy: It was both powerful and large, with a 30 nautical mile (nm) radius of maximum winds at landfall. Model performance is also intermediate, with a good representation of the surge at most open-coast locations, though surge is low at Galveston Pier, TX, by about half a meter.

Ike also exemplifies another key limitation of this reanalysis, which is the model's resolution. Modern coastal flood models often have coastal resolutions an order of magnitude higher than the mesh used here (400-500 m). This likely causes an underestimation of wave setup, particularly during storms, which may have been a factor in the errors at Galveston Pier. Furthermore, the mesh is too coarse to capture narrow, deep channels, such as the heavily dredged shipping lanes of many Texas inlets. As a result, the modeled surges in Galveston Bay are much smaller than observed. Note that the validation study that developed this mesh (Riverside and AECOM 2015) demonstrated that with high-resolution meteorological forcing (i.e., meteorology that closely matches the storm), peak surges from Sandy and Katrina were accurate, as were those for Ike along the open coast; it was again in the bays where Ike's surges were noticeably underpredicted. The next version of CORA, currently in development, will include more accurate representations of tropical cyclone wind and pressure fields. This should reduce errors associated with ERA5 tropical cyclone intensities but will not help errors due to mesh resolution/accuracy.

A significant part of low-frequency water levels comes from the assimilation. Based on the spectral analysis previously presented, the assimilation represents the majority of the water level signal starting at periods around 2 to 90 days. Therefore, the quality of the model's prediction of low-frequency water level characteristics (e.g., daily/monthly/annual averages, sea level trends) depends on: (A) the availability of nearby observations, (B) the quality of the assimilation methods, and (C) the spatial interpolation of the corrections between stations. For instance, along the 130 km of coast between the Sandy Hook and Atlantic City, NJ, NOAA stations, the long-term sea level rise rates in the assimilated model results are a result of the way corrections were interpolated between those stations, and so, to a degree, the sea level rise rates along that coast are also interpolated. Since rates generally vary slowly, this should be reasonable. But in inland waterways where sea levels may also be affected by rainfall/runoff, which is not part of the model, such trends are not captured in the model.

Like many coastal hydrodynamic models, ADCIRC uses an algorithm to determine when to reposition the front between the edge of the ocean and the land as water levels rise and fall; in models, this is termed "wetting/drying." The algorithm generally works well but can have trouble in low-gradient terrain, especially when neighbored by steep banks. As a result, places like southern Louisiana and other wide coastal marshes can sometimes retain water on top of these low-elevation areas. In such cases, the actual water level is better understood by looking at nearby channels that are draining properly.

In locations with large, localized changes in MSL and/or vertical land motion, the modeled results may not accurately reflect the local sea level change if there are no assimilated stations nearby. The highly localized and rapid subsidence at Pilots Station, LA (Figure 24), is an example of this. The nearest assimilation site is Grand Isle, and though modeled sea level trends at Pilottown, LA, are reasonable, those at Pilots Station, LA, seem to greatly underestimate the local subsidence-driven sea level rise.

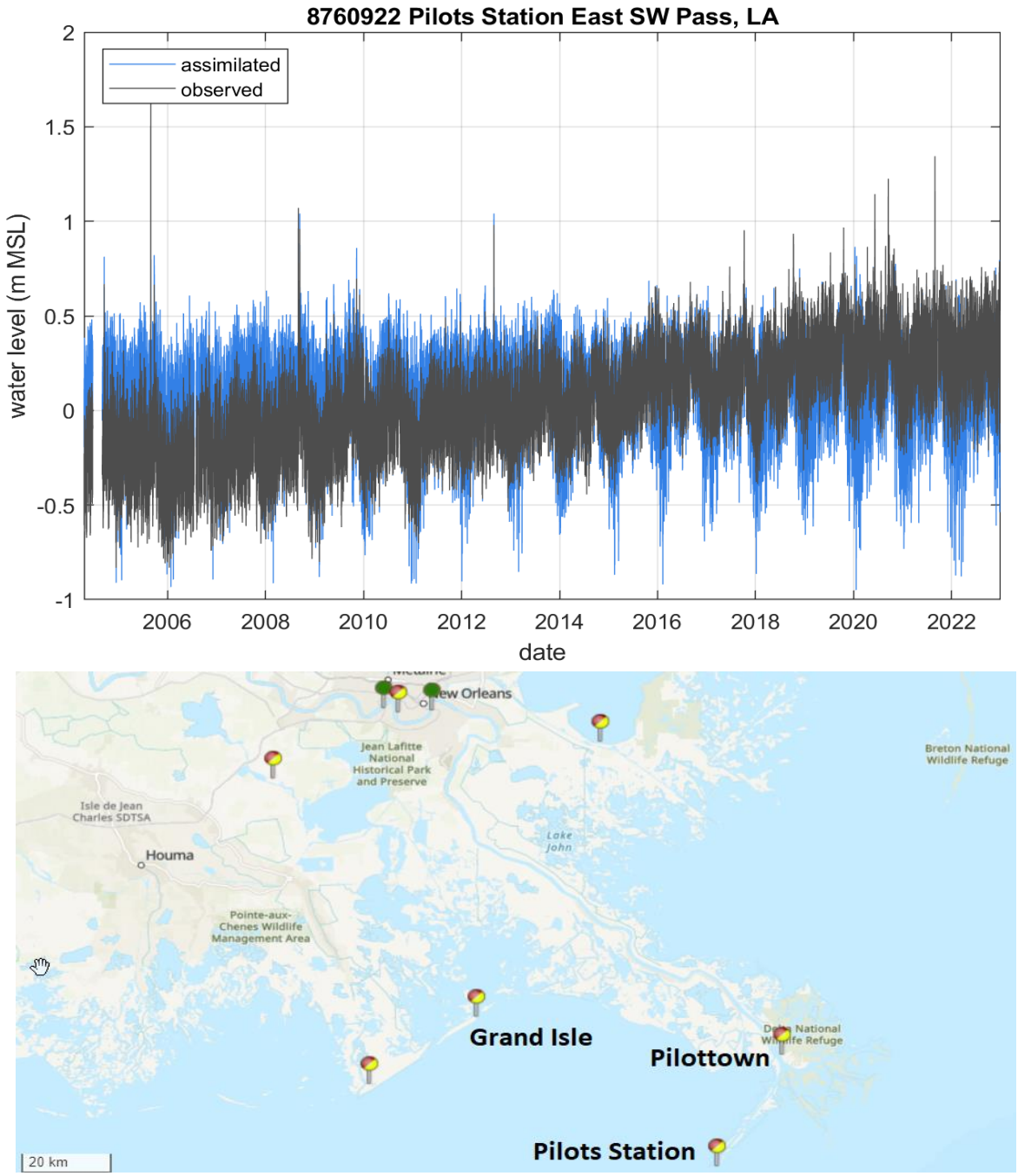

Figure 24. Time series of observed and modeled water levels at southern Louisiana stations and a map with select Gulf sites highlighted.

Water Level Results Summary

Model performance is quite good overall with errors for hourly data and extremes typically within 10 cm, though accuracy can depend heavily on the location, time period, and use case of interest. No modifications were made to the modeled solution based solely on intermediate results at validation sites, so these can be considered a reasonably independent (relative to the modelers' workflow and to the assimilated stations) estimate of model accuracy. Based on the model assessments herein, key areas for future improvement are in increased model resolution, better meteorological forcing for tropical cyclones, improved assimilation methods, and improved accuracy of tides. Model resolution is key to improving accuracy under all conditions and in all areas, but it is especially important to improve accuracy behind barrier islands and up rivers. A better representation of tropical cyclone winds and pressures is already under way and should be included in the next release of CORA. The assimilation methods have a powerful but difficult-to-quantify effect on the accuracy at un-gauged locations, and the current assimilation methods are relatively simple; although complex assimilation methods may be unnecessary, there is considerable room for improvement here (Asher et al. 2019). Though tidal accuracy is good, it is a limiting factor in the accuracy of the Gulf of Maine and plays a noteworthy role in errors in parts of the Atlantic.

6. WAVE RESULTS

The wind-wave model SWAN is formally coupled to ADCIRC (Dietrich et al. 2011a,b) and was included in the ADCIRC prediction to better account for the contribution of wave setup (Dean and Walton 2009) to the total water level at the coast. Although not the focus of this reanalysis, wave setup is an important contribution to coastal surge, particularly on relatively narrow continental shelves (Niedoroda et al. 2008, 2010), such as the southeastern Florida coast and the Cape Hatteras area of North Carolina.

Bulk wave parameters (significant wave height, direction, and period) were included in the ADCIRC output configuration. In the coupled ADCIRC/SWAN system, both models share the same spatial finite element grid, substantially simplifying the overall modeling process. The coupling timestep between ADCIRC and SWAN was set to 30 minutes. The spectral grid was defined with 36 direction bins (a constant resolution of 10°) and 30 frequency bins, with a lower frequency cutoff of 0.03 Hz. Other key parameters are specified in Table 4.

Table 4. Key settings for physics and numerics for Simulating WAves Nearshore (SWAN) model reanalysis.

Wave Physics	GEN3 KOMEN AGROW
White Capping	KOMEN 2.36E-5 3.02E-3 2.0 1.0 1.0
Wave Breaking	included
Friction	JONSWAP CFJON=0.038
Propagation scheme	BSBT (backward space, backward time)
Spectral Grid	36 directions, 30 frequencies, flow=0.03 Hz

Observations of waves from National Data Buoy Center (NDBC) buoys 44097, 44014, 44056, 41009, 41013, 42020, and 42022 were acquired using the NDBC[21] python package. A summary of the SWAN wave model performance is described below for the year 2018 at the Frying Pan Shoals, NC, buoy location. Other buoys are reported in Appendix A2.

Model Performance

The bulk wave predictions were compared to the NOAA NDBC buoy observations at several locations for the year 2018. Figure 25 shows significant wave heights at the near coastal water station 41013 (Frying Pan Shoals, NC) for observations and SWAN. The model generally underpredicts the larger wave heights, noted by the linear fit (red line) in the 3-5 m range. (We note that at this location, the large peak in mid-September is due to Hurricane Florence for which the model overpredicts the observations.) The bias is -0.087 m, indicating an overall underprediction, and a root mean square error (RMSE) of 0.31 m. Mean wave directions are similar, with observations and SWAN showing that much of the bulk wave field arrives generally from the southeastern quadrant. The SWAN prediction for peak period is not as correlated with the observations and generally underpredicts periods. Wave model skill plots for additional buoy locations are reported in Appendix A2.

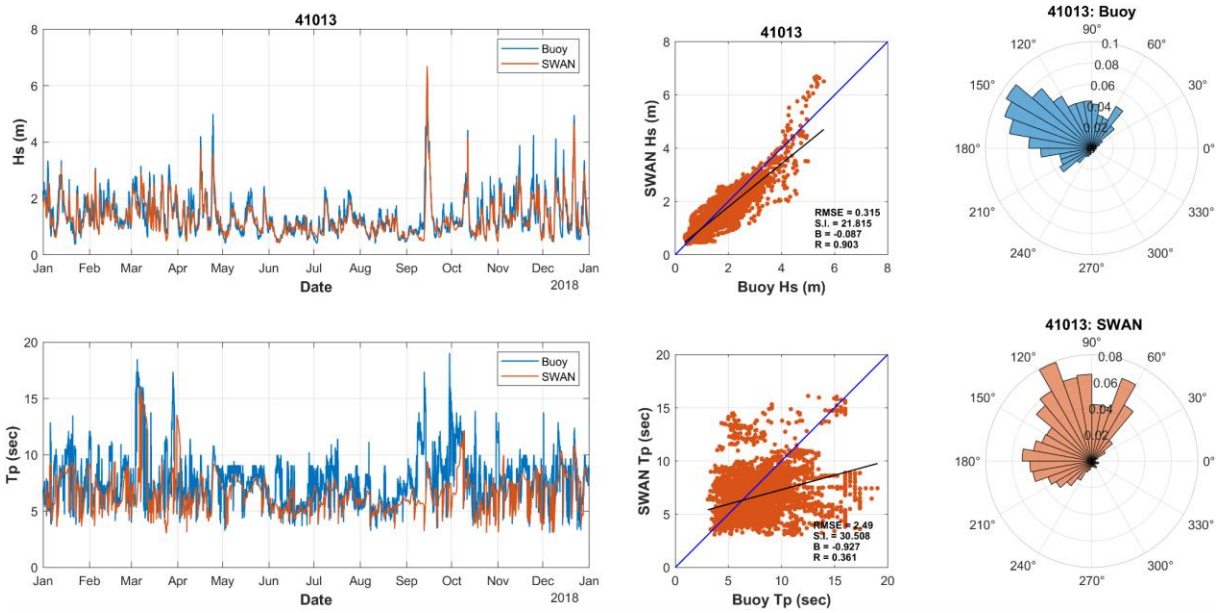

Figure 25. Bulk wave characteristics at National Data Buoy Center (NDBC) buoy 41013 (Frying Pan Shoals, NC, 33 m water depth) for observations and SWAN for the year 2018. The left column shows the time series of significant wave heights [m] and peak periods [sec]. The middle column shows a scatter plot and statistics for root-mean-square error (RMSE), scatter index (SI), bias (B), and correlation coefficient (R) for significant wave heights and peak periods. The right column shows polar density plots of the wave direction (to which waves are propagating), plotted with 0° being true east and increasing angles in the counterclockwise direction.

[21] National Data Buoy Center (NDBC) https://pypi.org/project/NDBC/

7. APPLICATIONS AND USE CASES

Background

Datasets produced from this reanalysis are foundational in their own right. CORA is sustained by the authoritative data provided by NOAA's NWLON then builds upon it by modeling the areas between observations with sound and validated modeling. In order to make both the datasets more accessible, geospatially compatible, and useful for flood assessment products and services, ADCIRC results are interpolated onto a grid to create a uniform and spatially continuous dataset.

The 500-m grid structure developed as part of this effort represents a standardized and repeatable approach for data analysis and delivery of the underlying ADCIRC model output. ADCIRC relies on an unstructured mesh consisting of nodes. In and along coastal areas, the model nodes capture pertinent details of coastal features in order to correctly model the flow of water in and around these features. This approach allows the model to capture detail where needed and to reduce detail or resolution where it's not needed, such as in the open ocean. One drawback of this approach is the lack of consistency in spatial detail and resolution along the coastline, which is often needed in geospatial applications. In addition, an unstructured mesh requires that users have the capacity and technical skill to transform the data in order to perform analysis with other geospatial datasets. By associating the model mesh with a standardized grid structure, model output and derived products can be delivered and analyzed in a consistent manner and can be more easily incorporated into geographic information systems and web applications. Further, decoupling the model mesh from the grid for associated products enables future updates to the mesh or even the entire model structure, without needing to rederive the spatial delivery of downstream products. A 500-m resolution was chosen as the grid resolution because it captures enough detail of coastal areas yet is still coarse enough to not over-represent model output, which generally has a point spacing of 250-400 m along the coast. The grid's horizontal extent was specified to cover coastal areas out to 3 km offshore (Figure 26).

Grid Development

The 500-m grid for CORA-GEC version 1.1 was developed using geospatial software and standard geospatial data types. The first step was to develop a polygon dataset of 500-m by 500-m cells that covered the model extent in coastal areas and out to 3 km offshore. This was accomplished through the "Create Fishnet" tool in Esri's ArcGIS Pro software. In order to ensure a consistent areal coverage of each grid cell throughout the CORA-GEC region, the polygon layer was created in the Albers Equal Area Projection, specifically the North American Datum of 1983 Conus Albers (EPSG: 5070)[22], which has a horizontal unit of meters. This projection is commonly used in national datasets because of its preservation of area while minimizing distortion in shape in the mid-latitudes of the continental United States[23]. Using a projection also enables future seamless additions to the grid extent if needed. This is accomplished by specifying the origin coordinate of new features in 500-m increments.

Due to the nature of how the Fishnet tool creates data, the output covered the full extent of the CORA-GEC study area (Maine to Texas), including inland areas. To refine the output to just coastal areas and out to 3 km, the ADCIRC model mesh was used to identify nodes that

[22] https://epsg.io/5070-1252

[23] Snyder, J. P. (1987). Map Projections: A Working Manual. U.S. Geological Survey Professional Paper 1395. Washington, DC: United States Government Printing Office. https://doi.org/10.3133/pp1395

had an elevation of 10 m or less. Once identified, these nodes were used to create a spatial footprint that, in combination with a 3-km offshore buffer, was used to select 500-m cells. The large spatial extent of the selected cells allows for changes to the inland extent of future modeled water levels without having to create a new 500-m grid. The result was consistent coastal coverage at a 500-m resolution (see Figure 27).

Once the final set of 500-m cells was created, and in preparation for associating the model output with the grid through interpolation, the polygon data were converted to centroids, where a single point is created for each cell at the cell's center. In addition, a unique ID was associated with each centroid to ensure that interpolated values can be tracked and associated with the correct cell. This step was needed in order to use a triangulated linear interpolation method to associate the model's nodal output with each cell. This approach assigns a value to a location, in this case a cell centroid, based on where that location or point falls on a distance-weighted triangular plane created from the model nodes. The triangular plane takes into account a centroid's distance from each node and weights the resulting value based on those distances, where closer nodes have more influence over the final value. Further, all 3 nodes making up the triangular element that a centroid falls in need to be wetted in order for the 500-m grid cell to be wetted for that time step. This spatially conservative approach to assigning values to the grid cells is appropriate due to the ADCIRC algorithms that define the models' wetting front, as noted in the Limitations section above.

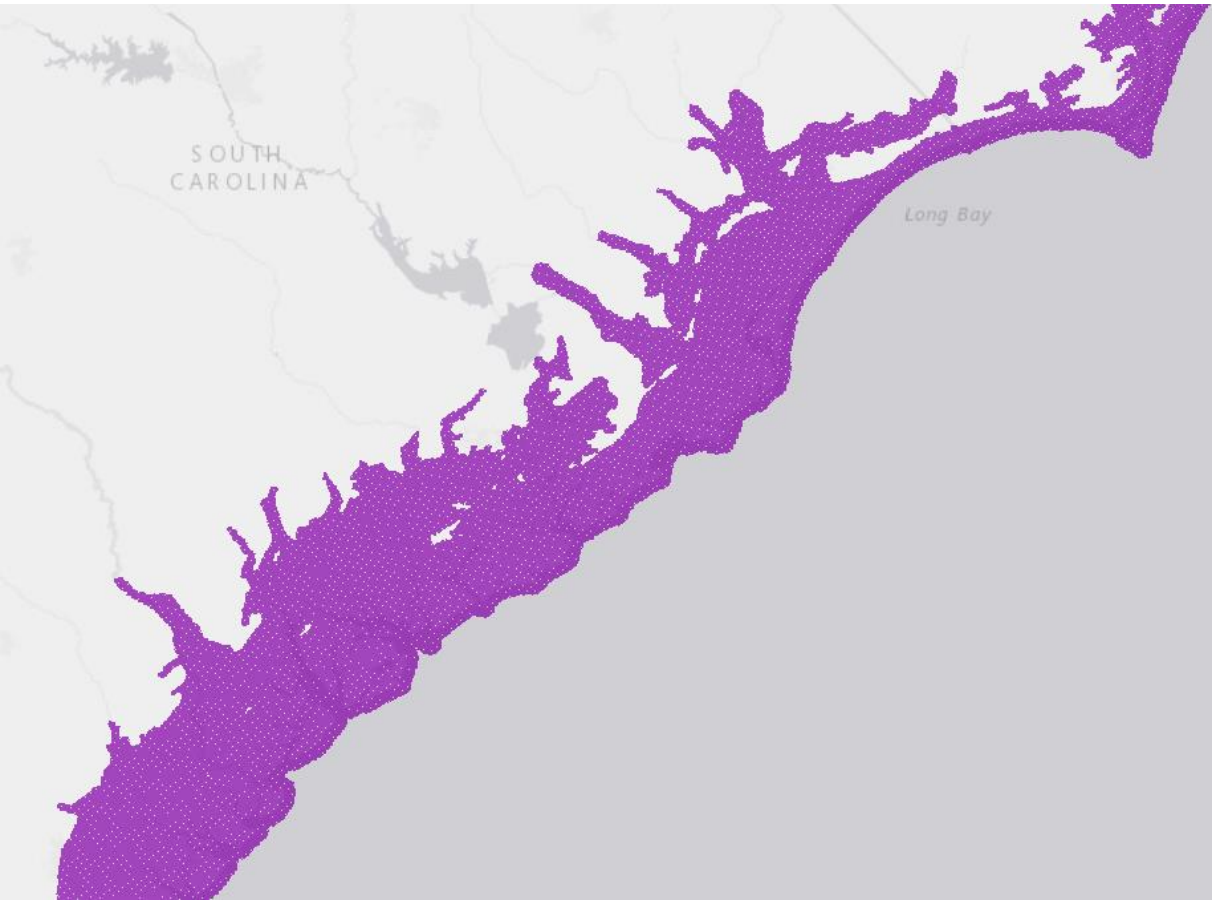

Figure 26. Five hundred-meter grid coverage for the coast of South Carolina.

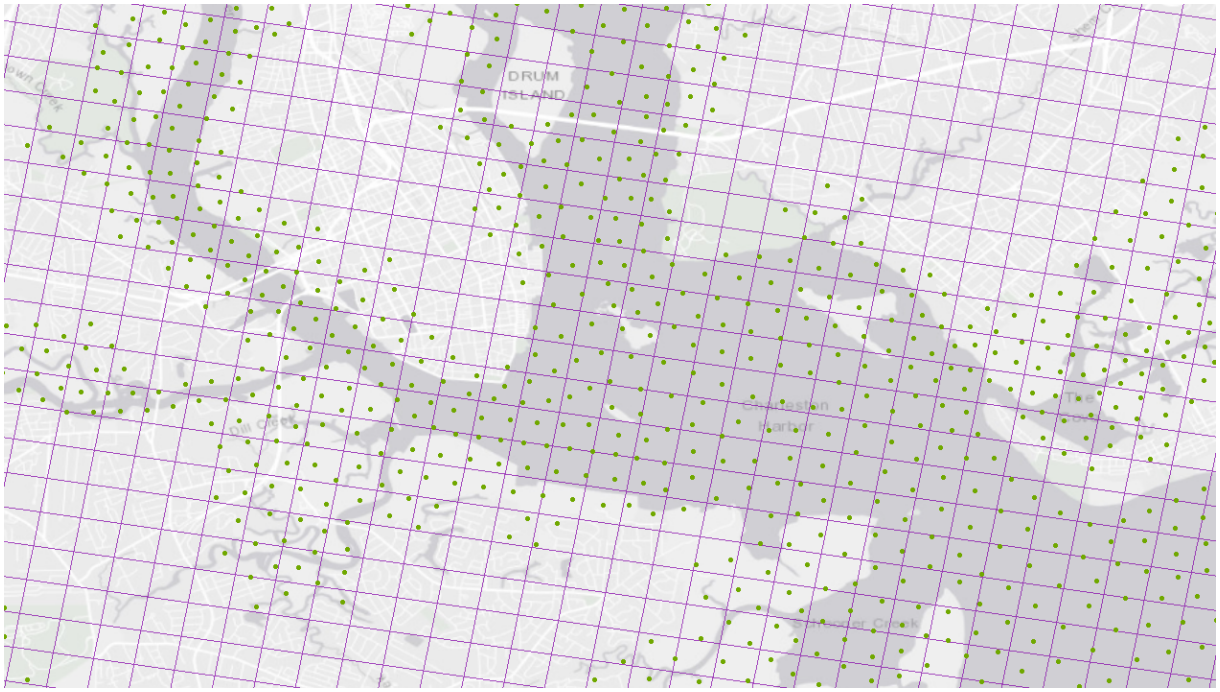

Figure 27. Five hundred-meter grid coverage and resolution (purple outlines) in Charleston, SC, compared to ADvanced CIRCulation (ADCIRC) model nodes (green points) that are wetted at least once a year.

Potential Use Cases

There are countless potential applications for data the CORA project has created, with the overall goal of helping coastal regions from state and federal initiatives to coastal community planning and preparedness and assistance for emergency responders.

1. Flooding extent and severity during extreme events

The use case noted in Rose et al. (2024) shows just how impactful coupling models with observations can be by illustrating elements associated with inundation along the Miami Beach, FL, coast in the wake of Hurricane Irma in 2017. The Rose et al. (2024) case study analyzes a roughly 20 x 30 km stretch of coastal Miami-Dade County centered around the Virginia Key, FL, tide station that provided hourly data when Category 4 Hurricane Irma made landfall. This example uses the posterior ADCIRC mesh to denote areas of land, ocean, and intertidal variability to illustrate where flooding took place during extreme water levels (Figure 28). When coupled with a layer showing areas most susceptible to flooding, including those with critical facilities and high percentages of development, CORA data paints a picture of where flooding has happened and where it's likely to occur during extreme weather events, evidence that can help emergency managers and city officials prioritize resilience.

2. Monthly HTF Outlook

We plan to utilize the 500-m gridded output to expand the NOAA monthly HTF outlook away from tide stations and to provide daily HTF likelihood every 500 m along most of the U.S. coastline. Comprehensive, high-resolution data will also greatly benefit sub-seasonal to seasonal forecasting in flood-prone areas.

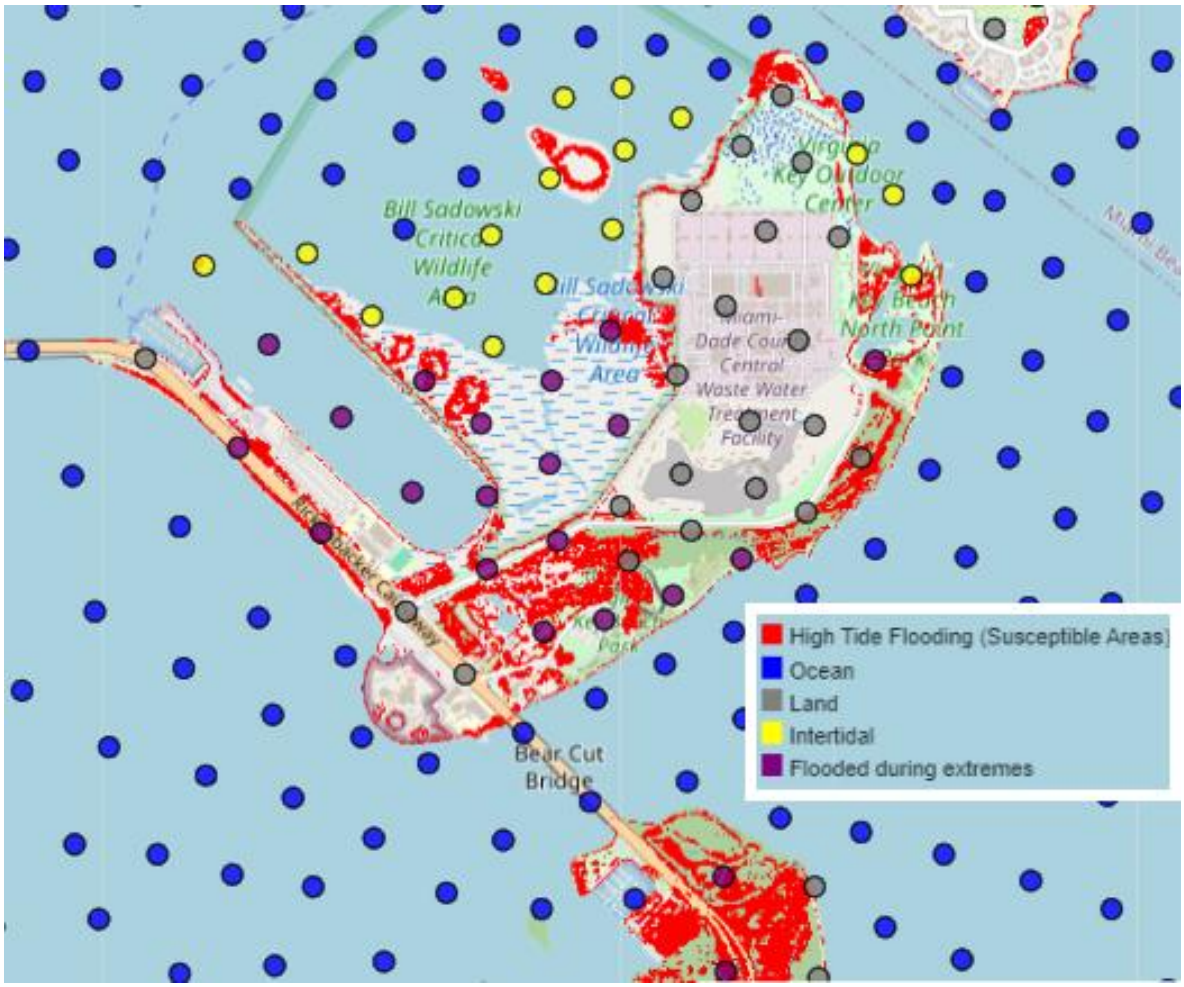

Figure 28. An example of water levels offshore Miami in the wake of Hurricane Irma in 2017. The ADvanced CIRCulation (ADCIRC) mesh is used to denote land (gray), areas susceptible to high tide (red), intertidal space (yellow), the ocean (blue), and flooding during extremes (purple).

3. Annual HTF Outlook

The Annual HTF Outlook analyzes the temporal trend of annual high tide flood frequency and assesses the influence of the El Nino Southern Oscillation (ENSO) phase (Sweet et al. 2018). Currently, the Annual HTF Outlook is only provided at NWLON water level stations. However, coastal flooding is hyperlocal, and the counts of flood days could be significantly different at locations a short distance away as influenced by local topography and distance to waterways. ENSO also brings changing wind, precipitation, and ocean patterns that can drive higher water levels at some locations. Assimilating water level observations will inherently include contributions from river discharge and other factors. CORA results will help assess the frequency of flooding to better understand implications of sea level rise.

4. Tidal datums and predictions

NOAA CO-OPS produces tidal datums (e.g., MSL, mean higher-high water, tidal range) and tidal predictions only at water level stations in open water bodies. These are issued primarily for navigation and safety purposes but have also been used in many other applications, such as assessing changes in sea level and non-tidal residual over time or inundation depth/duration curves for coastal natural resource areas. CORA results would allow

for a better understanding of how various datum surfaces and the harmonic constituents that comprise tidal predictions can change spatially within estuaries, bays, or river systems.

5. Impact on transportation networks

Transportation networks are one of the primary infrastructure assets impacted by flooding, including coastal flooding caused by surge and high tides. State and county transportation and emergency management agencies pay close attention to the status of low-lying roads and critical byways during storm events, however monitoring such a large area is extremely difficult and usually relies on placement of sensors in those locations. Many assets cannot be continuously monitored, and the impact of flooding on them is only found at a later time, if at all. CORA results will help agencies determine the frequency, severity, and duration that unmonitored roads have been likely impacted in the past, information which can be used to assess damages or extrapolating trends into the future for development planning and management purposes.

6. Climate resilience planning

Projections of temperature, precipitation, and sea level rise, and their associated impacts, are critical to climate resilience and mitigation planning across the country, from local community town plans to national-level initiatives. Risk and vulnerability assessments based on these projections help emergency managers, natural resource managers, and decision makers better prepare for changing environmental conditions. Output from CORA will help assess the larger spatial scale of coastal inundation threats to infrastructure, natural resources, public safety, and the economy, as well as the appropriate potential adaptation and mitigation responses.

ACKNOWLEDGEMENTS

We would like to acknowledge and thank the dedicated groups of individuals across NOAA and academia that contribute to the ongoing success of this reanalysis. To the CORA Technical Team for providing insight and innovative ideas for every aspect of the reanalysis, from modeling to naming conventions, your contributions have been pivotal and steadfast. Thank you members of the bimonthly working group meetings that turned a multi-faceted collaboration into fruition. Huge thanks to Tetrach Tech's RPS Group for applying modern cloud-based data optimization techniques and processes to the modeling approach, grid interpolation, data brokerage, and notebook development. You have been key partners at every step of the way. Thank you, IOOS, for believing in this effort to couple observations and modeling to provide data between tide stations. Your funding initiatives through the South East Coastal Ocean Observing System (SECOORA) paved the way for the entire GEC reanalysis and still remain vital to the success of the next stages of this project through your Cloud Sandbox. Special thanks to Kylee Lewis for analyzing and exploring datasets early on in the project. Thank you, CO-OPS Data Processing Team, for routinely analyzing and verifying water level observations foundational to the assimilated datasets that increase the accuracy of this reanalysis. Special thanks to everyone who helped review this technical report; your seasoned expertise and contributions helped elevate the results of this first chapter in NOAA's Coastal Ocean Reanalysis, including Technical Editing by Hailey Foglio. Additional thanks to partners in the U.S. Army Corps of Engineers, the U.S. Geological Survey, and the Stevens Institute of Technology for your insights and recommendations during external peer review. Through meaningful engagement and transparent analysis, we continue to advance capabilities in applied science and technology.

REFERENCES

Asher TG, Luettich Jr RA, Fleming JG, Blanton BO. 2019. Low frequency water level correction in storm surge models using data assimilation. Ocean Model. 144:101483.

Blanton B, Seim H, Luettich R, Lynch D, Werner F, Smith K, Voulgaris G, Bingham F, Way F. 2004. Barotropic tides in the South Atlantic Bight. J Geophys Res. 109:C12024.

Blanton B, McGee J, Fleming J, Kaiser C, Kaiser H, Lander H, Luettich R, Dresback K, Kolar R. 2012. Urgent computing of storm surge for North Carolina's coast. Procedia Comput Sci. 9(0):1677-1686. Proceedings of the 2012 International Conference on Computational Science. 2012 June 4-6; Omaha (NE).

Booij N, Ris R, Holtuijsen L. 1999. A third-generation wave model for coastal regions, part I: model description and validation. J Geophys Res. 104(C4):7649-7666.

Cyriac R, Dietrich J, Fleming J, Blanton B, Kaiser C, Dawson C, Luettich R. 2018. Variability in coastal flooding predictions due to forecast errors during Hurricane Arthur. Coast Eng. 137:59-78.

Dean RG and Walton TL. (2009). Wave setup. In: Kim YC, editor. Handbook of coastal and ocean engineering. Singapore: World Scientific. p. 1-23.

Dietrich J, Zijlema M, Westerink J, Holthuijsen L, Dawson C, Jensen RA Jr, Smith J, Stelling G, Stone G. 2011a. Modeling hurricane waves and storm surge using integrally-coupled, scalable computations. Coast Eng. 58(1):45-65.

Dietrich JC, Westerink JJ, Kennedy AB, Smith JM, Jensen RE, Zijlema M, Holthuijsen LH, Dawson C, Luettich RA Jr, Powell MD, et al. 2011b. Hurricane Gustav (2008) waves and storm surge: Hindcast, synoptic analysis, and validation in southern Louisiana. Mon Weather Rev. 139(8):2488-2522.

Dresback K, Fleming J, Blanton B, Kaiser C, Gourley J, Tromble E, Leuttich RA Jr, Kolar R, Hong Y, Cooten S, et al. 2013. Skill assessment of a real-time forecast system utilizing a coupled hydrologic and coastal hydrodynamic model during Hurricane Irene (2011). Cont Shelf Res. 71:78-94.

Dulac W, Cattiaux J, Chauvin F, Bourdin S, Fromang S. Forthcoming. Assessing the representation of tropical cyclones in ERA5 with the CNRM tracker. Clim Dyn.

Dusek G, Sweet W, Widlansky M, Thompson P, Marra J. 2022. A novel statistical approach to predict seasonal high tide flooding. Front Mar Sci. 9.

Egbert GD and Erofeeva SY. 2002. Efficient inverse modeling of barotropic ocean tides. J Atmos Ocean Technol. 19(2):183-204.

Fleming J, Fulcher C, Luettich R, Estrade B, Allen G, Winer H. 2008. A real time storm surge forecasting system using ADCIRC. In: Spaulding M, editor. Estuarine and Coastal Modeling. Reston (VA): ASCE.

Hill DF, Griffiths SD, Peltier WR, Horton BP, Törnqvist TE. 2011. High-resolution numerical modeling of tides in the western Atlantic, Gulf of Mexico, and Caribbean Sea during the Holocene. J Geophys Res Oceans. 116:C10014.

Hodges K, Cobb A, Vidale PL. 2017. How well are tropical cyclones represented in reanalysis datasets? J Clim. 30(14):5243-5264.

Hope ME, Westerink JJ, Kennedy AB, Kerr P, Dietrich JC, Dawson C, Bender CJ, Smith J, Jensen RE, Zijlema M, et al. 2013. Hindcast and validation of Hurricane Ike (2008) waves, forerunner, and storm surge. J Geophys Res Oceans. 118(9):4424-4460.

[IPCC] Intergovernmental Panel on Climate Change. 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva (Switzerland): 151 p.

Jean-Michel L, Eric G, Romain B-B, Gilles G, Angelique M, Marie D, Clement B, Mathieu H, Olivier LG, Charly R, et al. 2021. The Copernicus Global 1/12 deg Oceanic and Sea Ice GLORYS12 Reanalysis. Front Earth Sci. 9.

Kerr P, Donahue A, Westerink JJ, Luettich R Jr, Zheng L, Weisberg RH, Huang Y, Wang HV, Teng Y, Forrest DR, et al. 2013. US IOOS coastal and ocean modeling testbed: Inter-model evaluation of tides, waves, and hurricane surge in the Gulf of Mexico. J Geophys Res Oceans. 118(10):5129-5172.

Luettich R, Westerink J, Scheffner N. 1992. ADCIRC: an advanced three-dimensional circulation model for shelves, coasts and estuaries, report 1: theory and methodology of ADCIRC-2DDI and ADCIRC-3DL. Dredging Research Program Technical Report DRP-92-6, USACE/ERDC, Waterways Experiment Station. Vicksburg (MS): United States Army Corps of Engineers.

Niedoroda A, Das H, Slinn D, Dean R, Weaver R, Reed C, Smith J. 2009. The role of wave set-up during extreme storms. In: Coastal Engineering 2008. Singapore: World Scientific. p. 950-961.

Niedoroda A, Resio D, Toro G, Divoky D, Das H, Reed C. 2010. Analysis of the coastal Mississippi storm surge hazard. Ocean Eng. 37(1):82-90.

Riverside Technologies, Inc. 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 US Gulf-Atlantic Coast. Report Prepared for: NOAA/NOS Coast Survey Development Laboratory, Office of Coast Survey.

Roberts J and TD Roberts. 1978. Use of the Butterworth low-pass filter for oceanographic data. J Geophys Res. 83(C11):5510-5514.

Rose L, Widlansky MJ, Feng X, Thompson P, Asher TG, Dusek G, Blanton B, Luettich RA Jr, Callahan J, Brooks W, et al. 2024. Assessment of water levels from 43 years of NOAA's Coastal Ocean Reanalysis (CORA) for the Gulf of Mexico and East Coasts. Front Mar Sci. 11:1381228.

Smagorinsky J. 1963. General circulation experiments with the primitive equations, I: The basic experiments. Mon Weather Rev. 91(3):99-164.

Sun K and Pan J. 2023. Model of storm surge maximum water level increase in a coastal area using ensemble machine learning and explicable algorithm. Earth Space Sci. 10(12):e2023EA003243.

Sweet WV, Dusek G, Obeysekera J, Marra JJ. 2018. Patterns and projections of high tide flooding along the U.S. coastline using a common impact threshold. US Dept Commer NOAA CO-OPS Technical Report CO-OPS 086. 52 p.

Sweet WV, Hamlington BD, Kopp RE, Weaver CP, Barnard PL, Bekaert D, Brooks W, Craghan M, Dusek G, Frederikse T, et al. 2022. Global and regional sea level rise scenarios for the United States: updated mean projections and extreme water level probabilities along U.S. coastlines. US Dept Commer NOAA Technical Report NOS 01. 111 p.

Tanaka S, Bunya S, Westerink JJ, Dawson C, Luettich RA. 2011. Scalability of an unstructured grid continuous Galerkin based hurricane storm surge model. J Sci Comp. 46:329-358.

Westerink J, Luettich R, Feyen J, Atkinson J, Dawson C, Roberts H, Powell M, Dunion J, Kubatko E, Pourtaheri H. 2008. A basin-to channel-scale unstructured grid hurricane storm surge model applied to Southern Louisiana. Mon Weather Rev. 136:833-864.

Zijlema M. 2010. Computation of wind-wave spectra in coastal waters with swan on unstructured grids. Coast Eng. 57(3):267-277.

Zervas C. 2009. Sea Level Variations of the United States, 1854-2006. US Dept Commer NOAA Technical Report NOS CO-OPS 053.

APPENDIX

Wave Model Results

This section shows SWAN wave model comparisons for additional buoy observations for the year 2018. Each plot shows time series and scatter plots of significant wave height and peak wave period, as well as directional histograms for wave direction (counterclockwise and relative to true east).

44097 - Block Island, RI (154)

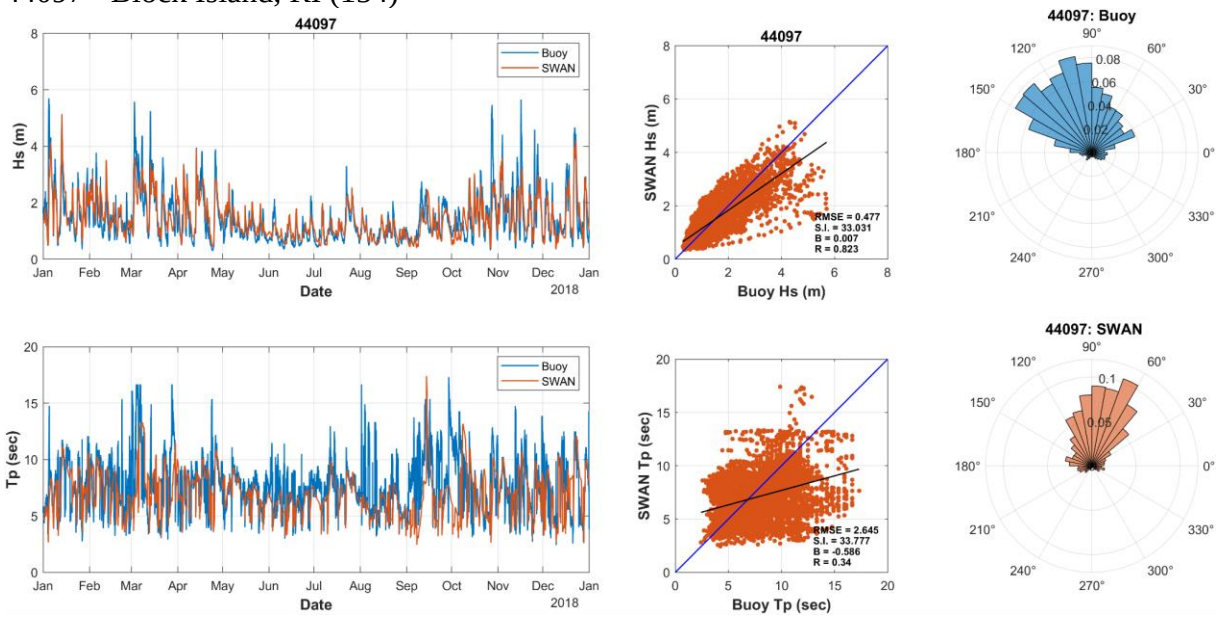

Figure 29. Comparison of buoy and Simulating WAves Nearshore (SWAN) model bulk wave parameters at National Data Buoy Center (NDBC) station 44097 (Block Island, RI (154), 49 m water depth).

44014 (LLNR 550, Virginia Beach)

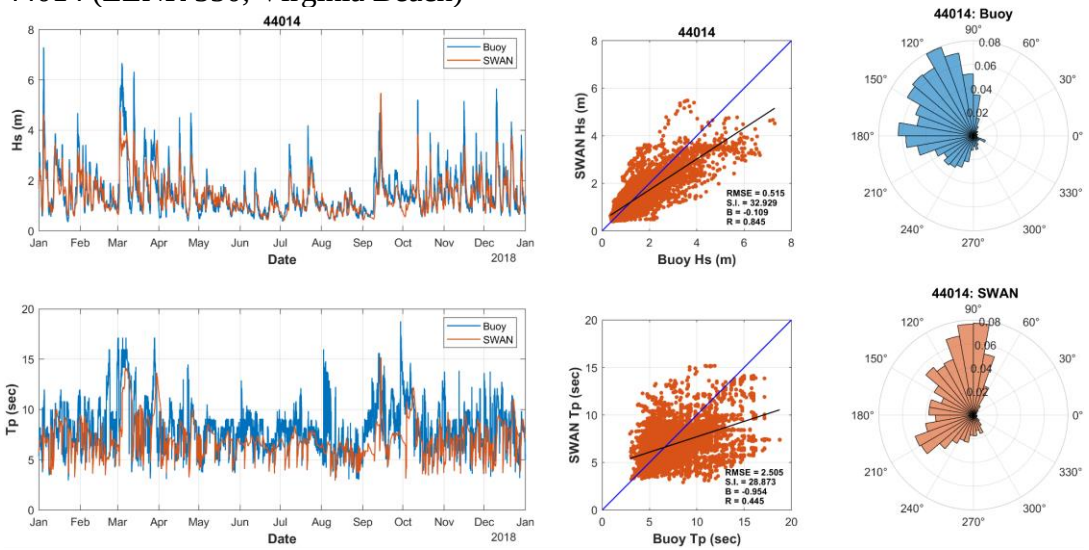

Figure 30. Comparison of buoy and Simulating WAves Nearshore (SWAN) model bulk wave parameters at National Data Buoy Center (NDBC) station 44014 (LLNR 550, Virginia Beach, VA, 49 m water depth).

44056 (Duck FRF, NC)

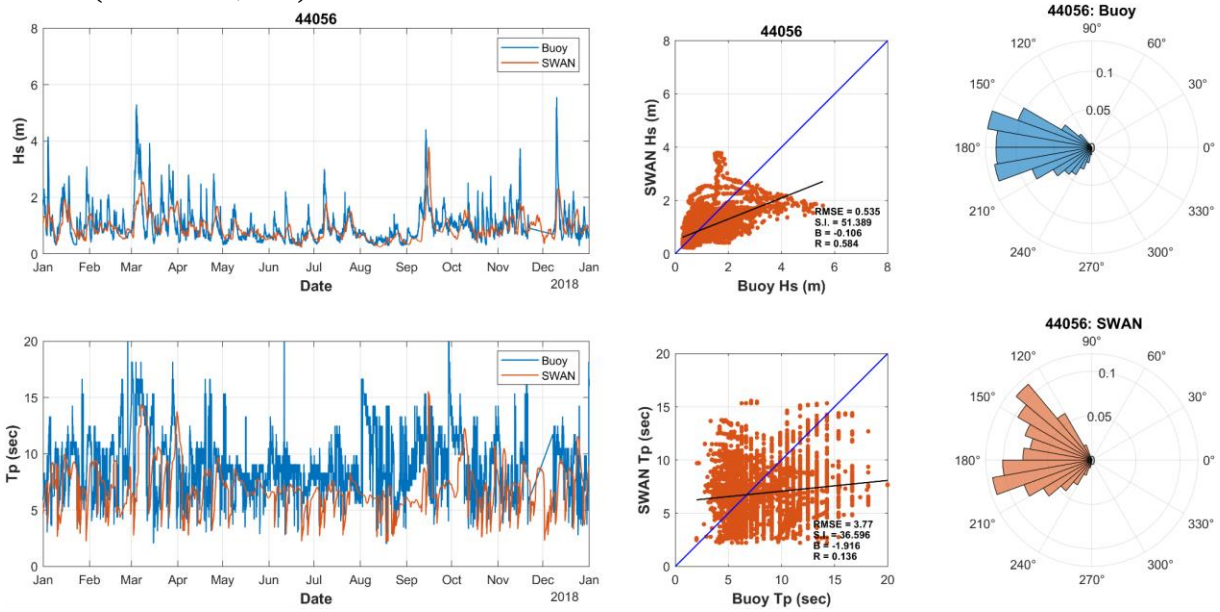

Figure 31. Comparison of buoy and Simulating WAves Nearshore (SWAN) model bulk wave parameters at National Data Buoy Center (NDBC) station 44056 (Duck FRF, NC, 17.8 m water depth).

41009 (LLNR 840, Cape Canaveral)

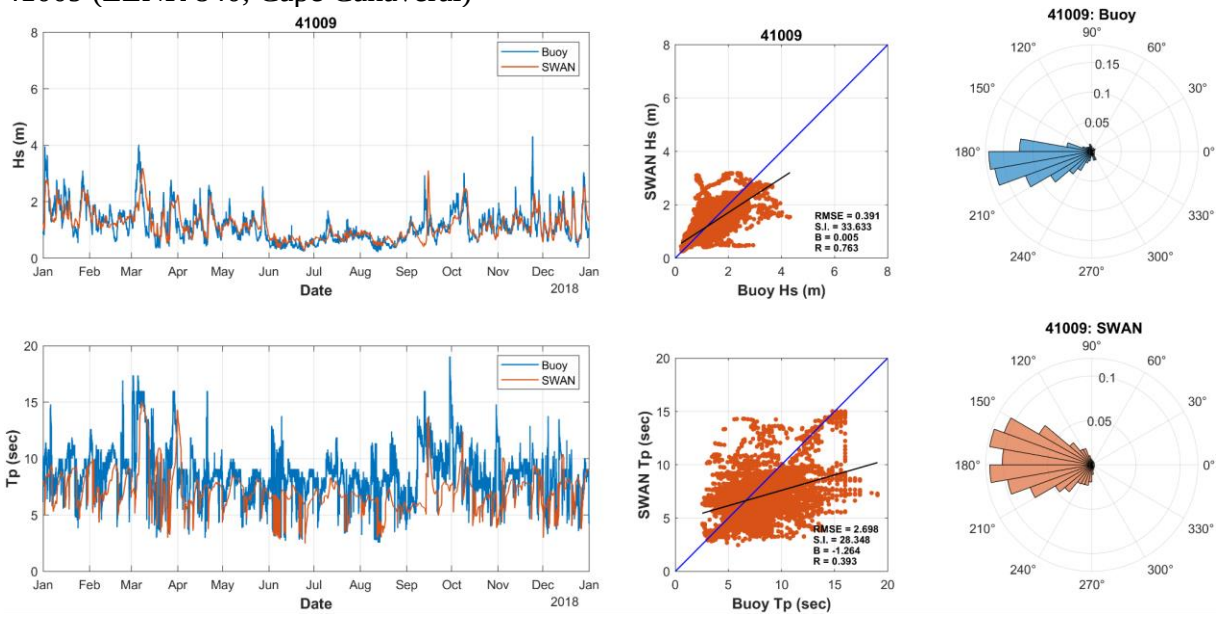

Figure 32. Comparison of buoy and Simulating WAves Nearshore (SWAN) model bulk wave parameters at National Data Buoy Center (NDBC) station 41009 (LLNR 840, Cape Canaveral, FL, 42 m water depth).

42040 (LLNR 245, Luke Offshore)

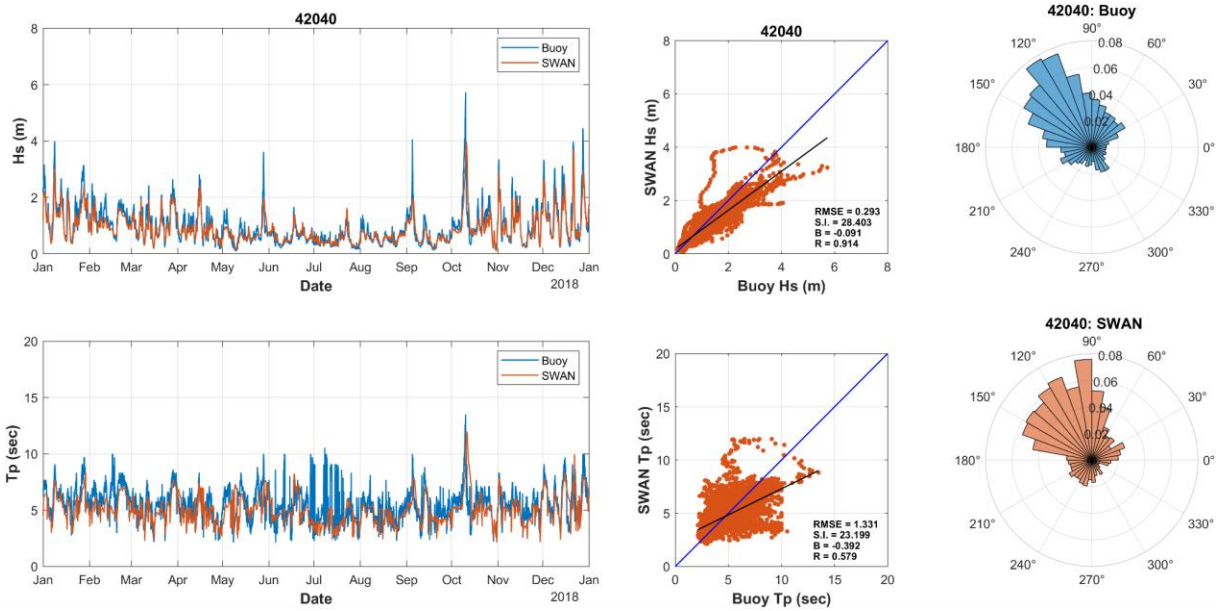

Figure 33. Comparison of buoy and Simulating WAves Nearshore (SWAN) model bulk wave parameters at National Data Buoy Center (NDBC) station 42040 (LLNR 245, Luke Offshore Platform, AL, 192 m water depth).

42002 (LLNR 1470, TX)

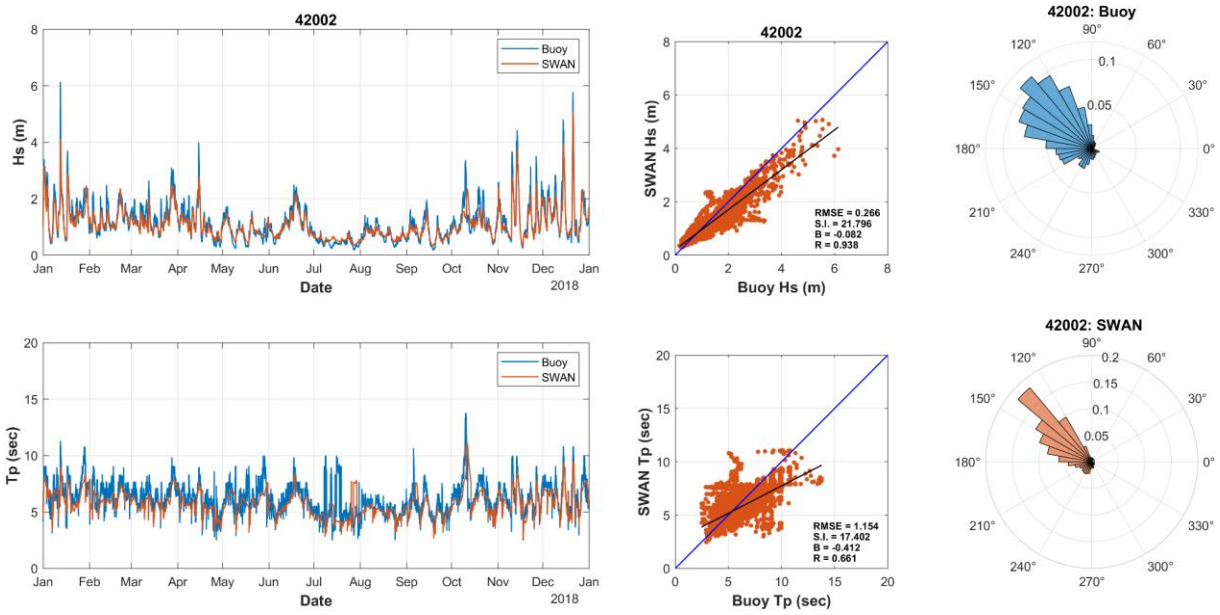

Figure 34. Comparison of buoy and Simulating WAves Nearshore (SWAN) model bulk wave parameters at National Data Buoy Center (NDBC) station 42002 (LLNR 1470, TX, 3088 m water depth).