Offshore wind energy forecasting sensitivity to sea surface temperature input in the Mid-Atlantic

Stephanie Redfern¹, Mike Optis¹, Geng Xia¹, and Caroline Draxl¹

¹National Renewable Energy Laboratory, Golden, Colorado, USA

Correspondence: Stephanie Redfern (stephanie.redfern@nrel.gov)

Abstract. As offshore wind farm development expands, accurate wind resource forecasting over the ocean is needed. One important yet relatively unexplored aspect of offshore wind resource assessment is the role of sea surface temperature (SST). Models are generally forced with reanalysis data sets, which employ daily SST products. Compared with observations, significant variations in SSTs that occur on finer time scales are often not captured. Consequently, shorter-lived events such as sea breezes and low-level jets (among others), which are influenced by SSTs, may not be correctly represented in model results. The use of hourly SST products may improve the forecasting of these events. In this study, we examine the sensitivity of model output from the Weather Research and Forecasting Model (WRF) 4.2.1 to two different SST products—a daily, spatially coarser resolution data set (the Operational Sea Surface Temperature and Ice Analysis, or OSTIA), and an hourly, spatially finer resolution product (SSTs from the Geostationary Operational Environmental Satellite 16, or GOES-16). We find that in the Mid-Atlantic, although OSTIA SSTs validate better against in situ observations taken via a buoy array in the area, the two products result in comparable hub-height wind characterization performance on monthly time scales. Additionally, during flagged events that show statistically significant wind speed deviations between the two simulations, the GOES-16-forced simulation outperforms that forced by OSTIA.

Copyright statement. This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Wind Energy Technologies Office and by the National Offshore Wind Research and Development Consortium under Agreement No. CRD-19-16351. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

1 Introduction

The United States Atlantic coast is a development site for upcoming offshore wind projects. There are 15 leasing areas located throughout the Atlantic Outer Continental Shelf, where a number of offshore wind farms are planned to be developed (Bureau of Ocean Energy Management, 2018). Therefore, characterizing offshore boundary layer winds in the region has risen in
importance. Accurate forecasting will provide developers with a better understanding of local wind patterns, which can inform wind farm planning and layout decisions (Banta et al., 2018). Additionally, improved weather prediction will allow for real-time adjustments of turbine operation to increase their operating efficiency and protect them against unnecessary wear and tear (Gutierrez et al., 2016, 2017; Debnath et al., 2021).

The Mid-Atlantic Bight (MAB) is an offshore cold pool region spanning the eastern United States coast from North Carolina up through Cape Cod, Massachusetts, and it overlies the offshore wind leasing areas. The cold pool forms during the summer months, when the ocean becomes strongly stratified and the thermocline traps colder water near the ocean floor. During the transition to winter, as sea surface temperatures (SSTs) drop, the stratification weakens and the cold pool breaks down. Thus, the cold pool generally persists from the spring through the fall. Southerly winds that drive surface currents offshore will result in coastal upwelling of this colder water. And, at times, strong winds associated with storm development can mix the cold pool upward, cooling the surface and influencing near-surface temperatures and winds (Colle and Novak, 2010; Chen et al., 2018; Murphy et al., 2021).

Accurate representation of the MAB in forecasting models is important because SSTs are closely tied to offshore winds. Horizontal temperature gradients between land and the ocean, as well as vertical temperature gradients over the ocean—which can form, for example, when SSTs are anomalously cold as with the MAB—help define offshore airflow. In particular, variations in temperature can lead to or impact short-lived offshore events occurring on hourly time scales, such as sea breezes and low-level jets (LLJs). Sea breezes are driven by the land-sea temperature difference, which, if strong enough (around 5°C or greater), can generate a circulation between the water and the land (Stull et al., 2015). With a relatively colder ocean, as during summer months, this leads to a near-ground breeze blowing landward, with a weak recirculation toward the ocean aloft (Miller et al., 2003; Lombardo et al., 2018). Similarly, the near-surface horizontal and air-sea temperature differences dictate the strength of stratification over the ocean. Studies have found a robust link between atmospheric stability and LLJ development, so accurately representing SSTs is key to modeling near-surface stability and, accordingly, LLJs (Gerber et al., 1989; Kaellstrand, 1998; Kikuchi et al., 2020; Debnath et al., 2021). Both of these phenomena can affect individual wind turbine and whole farm operation, so forecasting them correctly can improve power output and turbine reliability (Nunalee and Basu, 2014; Pichugina et al., 2017; Murphy et al., 2020; Xia et al., 2021).

Typical climate and weather model initialization and forcing inputs are reanalysis products, such as ERA5 and MERRA2, which are global data sets that assimilate model output with observations to create a comprehensive picture of climate at each time step considered (Gelaro et al., 2017; Hersbach et al., 2020). These data sets primarily include global SST products that are produced at lower temporal and spatial resolutions than what can be available via regional, geostationary satellites. These coarser-resolution data sets, therefore, do not capture observed hourly and, in many cases, diurnal fluctuations in SSTs, which may influence their ability to properly force sea breezes and LLJs. Some preliminary comparisons between weather simulations, forced with different SST products, indicate that this particular input can have a significant impact on modeled offshore wind speeds (Byun et al., 2007; Chen et al., 2011; Dragaud et al., 2019; Kikuchi et al., 2020).

Few studies have examined the impact of finer-temporal resolution SST products specifically on wind forecasting, and to the authors’ knowledge, none so far have focused on the Mid-Atlantic. There have been studies looking at numerical weather
prediction model sensitivity to SST, but they have considered other regions or different, often coarser spatial and temporal resolution, products (Chen et al., 2011; Park et al., 2011; Shimada et al., 2015; Dragaud et al., 2019; Kikuchi et al., 2020). In this article, we explore the effects of forcing the Weather Research and Forecasting Model (WRF), a numerical weather prediction model used for research and operational weather forecasting, with different SST data sets characterized by different spatial and temporal resolutions, in the Mid-Atlantic region during the summer months. Specifically, we address differences in model performance on monthly time scales and then contrast characterization effectiveness during shorter wind events. Section 2 lays out the data, model setup, and methods used in this study. Section 3 explains the findings of our simulations, and Section 4 explores their implications. Finally, Section 5 summarizes the intent of the study as well as its findings.

2 Methods

We run two model simulations with identical setups, aside from the input SST data, off the Mid-Atlantic coast for June and July of 2020. The output data are compared with in situ measurements taken at buoys (SSTs) and floating lidars (winds) in the region. We evaluate performance primarily via a set of validation metrics calculated on monthly time scales. We then flag specific events during which the model generally captures regional winds, but output from the two simulations deviate significantly (defined in this study as one or more standard deviations from their mean differences) from one another. Again, validation analysis is performed for these periods.

2.1 In Situ Data

This study makes use of both SST and wind profile observational data for model validation. SSTs are provided by the National Buoy Data Center (NBDC) at several locations along the Mid-Atlantic Coast, as listed in Table 1 and shown in Fig. 1. Buoy data located at the Atlantic Shores Offshore Wind location are also used (Fig. 1). Wind data have been taken from the Atlantic Shores floating lidar and the two New York State Energy Research & Development Authority (NYSERDA) floating lidars, whose locations are listed in Table 1. These lidars provide wind speed and wind direction at 10-minute intervals from either 10 m (Atlantic Shores) or 40 m (NYSERDA) up through 250 m above sea level. There are periods of missing data for all buoys and lidars.

2.2 Model Setup

WRF Version 4.2.1 is the numerical weather prediction model employed in this study (Powers et al., 2017). WRF is a fully compressible, non-hydrostatic model that is used for both research and operational applications. Our model setup, including key physics and dynamics options, are outlined in Table 2.

The study area spans the majority of the MAB, with the nested domain running from the mid-Virginia coast up through Cape Cod to the north (Fig. 1).
**Figure 1.** The selected study area (nested domain only)—with lidars depicted by stars and buoys by hollow circles. The Atlantic Shores location is the site of both a buoy and a lidar. Leasing areas are outlined in brown and white.

**Table 1. Buoy and Lidar Locations**

<table>
<thead>
<tr>
<th>Buoy or Lidar</th>
<th>Name</th>
<th>Latitude</th>
<th>Longitude</th>
<th>SST Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lidar &amp; Buoy</td>
<td>NYSERDA E05</td>
<td>40.1614</td>
<td>-72.7396</td>
<td>0.8 m</td>
</tr>
<tr>
<td>Lidar &amp; Buoy</td>
<td>NYSERDA E06</td>
<td>39.6273</td>
<td>-73.4123</td>
<td>0.8 m</td>
</tr>
<tr>
<td>Lidar &amp; Buoy</td>
<td>Atlantic Shores</td>
<td>39.2717</td>
<td>-73.8892</td>
<td>1 m</td>
</tr>
<tr>
<td>Buoy</td>
<td>44017</td>
<td>40.693</td>
<td>-72.049</td>
<td>1.5 m</td>
</tr>
<tr>
<td>Buoy</td>
<td>44025</td>
<td>40.251</td>
<td>-73.164</td>
<td>1.5 m</td>
</tr>
<tr>
<td>Buoy</td>
<td>44065</td>
<td>40.369</td>
<td>-73.703</td>
<td>1.5 m</td>
</tr>
<tr>
<td>Buoy</td>
<td>44075</td>
<td>40.363</td>
<td>-70.883</td>
<td>1 m</td>
</tr>
<tr>
<td>Buoy</td>
<td>44076</td>
<td>40.137</td>
<td>-70.775</td>
<td>1 m</td>
</tr>
<tr>
<td>Buoy</td>
<td>44077</td>
<td>39.940</td>
<td>-70.883</td>
<td>1 m</td>
</tr>
<tr>
<td>Buoy</td>
<td>44091</td>
<td>39.768</td>
<td>-73.770</td>
<td>0.46 m</td>
</tr>
<tr>
<td>Buoy</td>
<td>44097</td>
<td>40.967</td>
<td>-71.126</td>
<td>0.46 m</td>
</tr>
</tbody>
</table>
Table 2. WRF Physics Options Used in This Study

<table>
<thead>
<tr>
<th>WRF Parameter</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Domains</td>
<td>2</td>
</tr>
<tr>
<td>Domain Resolution</td>
<td>6 km (Parent), 2 km (Nest)</td>
</tr>
<tr>
<td>Output Time Resolution</td>
<td>5 minutes</td>
</tr>
<tr>
<td>Vertical Levels</td>
<td>61</td>
</tr>
<tr>
<td>Reanalysis Data</td>
<td>ERA5</td>
</tr>
<tr>
<td>Microphysics</td>
<td>Ferrier</td>
</tr>
<tr>
<td>Radiation Scheme</td>
<td>RRTMG (longwave &amp; shortwave)</td>
</tr>
<tr>
<td>Planetary Boundary Layer</td>
<td>Nakanishi and Niino (MYNN), 2006</td>
</tr>
<tr>
<td>Surface Layer Parameterization</td>
<td>MYNN</td>
</tr>
<tr>
<td>Land Surface Scheme</td>
<td>Unified Noah land-surface model</td>
</tr>
<tr>
<td>Cumulus Parameterization</td>
<td>Kain-Fritsch</td>
</tr>
<tr>
<td>Upper-Level Damping</td>
<td>Rayleigh at 5km depth</td>
</tr>
</tbody>
</table>

Table 3. SST Data Sets Considered in This Study

<table>
<thead>
<tr>
<th>Parameter</th>
<th>OSTIA</th>
<th>GOES-16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satellite Coverage</td>
<td>Global</td>
<td>Regional</td>
</tr>
<tr>
<td>Temporal Resolution</td>
<td>Daily</td>
<td>Hourly</td>
</tr>
<tr>
<td>Spatial Resolution</td>
<td>0.054°</td>
<td>0.02°</td>
</tr>
<tr>
<td>Processing Level</td>
<td>Gridded &amp; assimilated</td>
<td>Gridded only</td>
</tr>
<tr>
<td></td>
<td>with in situ observations</td>
<td></td>
</tr>
<tr>
<td>Gap-Filling</td>
<td>Released product is filled</td>
<td>DINEOF needed</td>
</tr>
</tbody>
</table>

2.3 Sea Surface Temperature Data

We compare how well three different SST datasets validate against buoy observations and subsequently select the two best-performing data sets to force our simulations (Table 3). Aside from these different SST product inputs, the rest of the model parameters in the simulations remain identical.

Our coarser-resolution selection, the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system, is a daily global product that combines in situ observations taken from buoys and ships, model output, and multiple remotely sensed SST data sets. Because of this, it is a complete product in that it has no missing data. OSTIA has a spatial resolution of 0.05°
Additionally, as a daily product, OSTIA provides a foundation SST, which is measured deep enough in the water to discount diurnal temperature fluctuations (Stark et al., 2007; Donlon et al., 2012; Fiedler et al., 2019).

Our finer-resolution product is taken via GOES-16, which is a geostationary, regional satellite with a spatial resolution of 0.02° x 0.02°. This product does not assimilate its measurements with in situ data. While GOES-16 does not offer global coverage and, therefore, cannot be used for certain world regions, it does cover the Mid-Atlantic Bureau of Ocean Energy Management offshore wind lease areas, which is our region of interest. Because GOES-16 has an hourly resolution, it can capture diurnal changes in SST and, therefore, measures surface temperature (Schmit et al., 2005, 2008).

We have selected the ERA5 global reanalysis data set to force our simulations. OSTIA is the SST data set native to this product (Hersbach et al., 2020). As such, when included as part of ERA5, OSTIA’s resolution has been adjusted to match ERA5’s 31 km spatial resolution and hourly temporal resolution. For our simulations, however, we overwrite these SSTs with OSTIA data at its original resolution of 0.05° and with GOES-16 data at its resolution of 0.02°.

Due to the lesser level of processing in the GOES-16 data set, it contains numerous data gaps that must be filled. This missing data arises due to a post-processing algorithm applied prior to release of the data set, which flags pixels that fall below a specified temperature threshold. This filter is in place to remove cloud cover. While this method is effective with regard to its set intent, it can also erroneously discard valid pixels that capture the cold water upwelling typical to the MAB region during the warmer months. And, although this cold-pixel filter is also a common practice in global SST data sets (OSTIA), the high level of post-processing applied in those products interpolates over and fills the missing grid cells prior to release.

To gap-fill the GOES-16 data, we employ the Data INterpolating Empirical Orthogonal Function algorithm, or DINEOF, which is an open-source application that applies empirical orthogonal function (EOF) analysis to reconstruct incomplete data sets. The program was originally designed to specifically gap-fill remotely observed SSTs that contain missing data due to cloud-flagging and removal algorithms, and has demonstrated strong results in past studies (Alvera-Azcárate et al., 2005; Ping et al., 2016).

We additionally apply to the GOES-16 SST data set the sensor-specific error statistics (SSES) bias field that is included with the product. This component accounts for retrieval bias via a statistical algorithm designed to correct for errors in the SST field. Compared with the SST values alone, the bias-corrected GOES-16 data offsets an inherent warm bias, bringing the values down to be more consistent with those of OSTIA.

### 2.4 Event Selection

We are particularly interested in examining model performance for forecasting shorter wind events, as LLJs and sea breezes occur on hourly time scales. Therefore, we have created a set of parameters that, when met, detect briefer events during which one simulation may be outperforming the other. Once the events have been collected, we more closely examine them to characterize the wind profiles and SSTs at those times.

For an event to be flagged, it must meet the following criteria:

1. Correlation for both models is above 0.5 at hub height, for two of the three lidar locations.
Table 4. Error Metrics Used in This Study

<table>
<thead>
<tr>
<th>Error Metric</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>$\bar{p} - \bar{o}$</td>
</tr>
<tr>
<td>Unbiased Root Mean Square Error (cRMSE)</td>
<td>$\left[\frac{1}{N} \sum_{n=1}^{N} [(p_n - \bar{p})(o_n - \bar{o})]^2\right]^{\frac{1}{2}}$</td>
</tr>
<tr>
<td>Square of Correlation Coefficient (R²)</td>
<td>$\left[\frac{1}{N} \sum_{n=1}^{N} (p_n - \bar{p})(o_n - \bar{o})\right]^2$</td>
</tr>
<tr>
<td>Wasserstein Metric / Earth-Mover’s Distance (EMD)</td>
<td>$\sum_{i=1}^{m} \sum_{j=1}^{n} M_{ij} d_{ij}$</td>
</tr>
</tbody>
</table>

2. Differences in wind speeds between the two models must be greater than one standard deviation from the monthly mean difference.

3. Gaps during which the wind speed difference drops below one standard deviation must not persist for more than 2 hours during a single event.

4. Events must last for at least one hour.

This set of event characteristics first acts to filter out periods during which WRF is generally underperforming—possibly due to model shortcomings outside of SST forcing—so that the performance difference in the selected events may be attributed to SSTs with more certainty. Then, events are located during which the two simulations forecast statistically significantly different hub-height wind speeds, which persist for a period of time long enough to substantially affect power generation.

### 2.5 Validation Metrics

To evaluate which simulation performs best during our study period, we calculate sets of validation metrics, as outlined by Optis et al. (2020). Specifically, we look at SSTs and 100-m (hub-height) wind speeds, on both monthly and event time scales. The metrics we calculate are named and defined in Table 4, and they provide a quantification of the error present in each case.

The bias provides information on the average simulation performance during the evaluation period—specifically, if the model is consistently over- or under-predicting the output variable in consideration. Unbiased root mean square error (RMSE) provides a more nuanced look at the spread of the error in the results. Correlation quantifies how well the simulations’ variables change in coordination with those of the observations. And, finally, the Wasserstein Metric, also known as the Earth-Mover’s Distance (EMD), measures the difference between the observed and simulated variable distributions.

### 3 Results

We compare the performance of GOES-16 with OSTIA by validating each data set’s SSTs and modeled hub-height winds with in situ observations collected via buoys and lidars in the Mid-Atlantic. First, we focus on the SSTs of each remotely sensed
data set and how they compare with buoy data. Following SST validation, we evaluate the model’s wind characterization performance using each input data set. We assume a hub height of 100 m and compare output winds at this altitude against measurements taken via floating lidars off the Mid-Atlantic coast. Specifically, we analyze monthly performance and then select several shorter periods during June and July of 2020 during which we compare wind characterization accuracy.

Overall, we find that on a monthly time frame, OSTIA SSTs validate better than those of GOES-16. Average hub-height winds at this resolution validate similarly for both products. At an event-scale temporal resolution—that is, on the order of hours—GOES-16 outperforms OSTIA for the selected events.

### 3.1 Sea Surface Temperature Performance

Before narrowing the scope of this study to a comparison between OSTIA and GOES-16 SSTs, we have included in the SST analysis a third data set: the Multi-scale Ultra-high Resolution SST analysis (MUR). The MUR is a daily product that has undergone foundation-level preprocessing (the same level as OSTIA). MUR has a spatial resolution of 0.25° x 0.25°, which is significantly coarser than either of the other two SST data sets (Chin et al., 2017). We evaluate SST performance by interpolating the satellite-based products to 10-minute intervals (the in situ data output resolution) and making a brief qualitative assessment followed by the calculation of each of the validation metrics selected for this study.

Taking a qualitative look at each product’s SST—specifically at Atlantic Shores—as well as the in situ measurements at the buoy, we see that while GOES-16 tracks the direnual cycle seen by observations, the other two products do not (Fig. 2). Despite this feature, however, there are a number of times during each month when GOES-16 does not capture observed dips in temperature. Moreover, during many of these periods, the daily data sets, though missing the nuance of GOES-16, better predict the colder SSTs. Despite this, at Atlantic Shores, GOES-16 does validate better than both OSTIA and MUR for both June and July at this location (Table 5).

To broaden the scope of SST validation, we take an even more nuanced look at SST product performance by calculating validation metrics at all the buoys listed in Table 1, instead of just Atlantic Shores. Because MUR markedly underperforms compared with the other two data sources, it is not included in this (and further) analysis.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Bias (°C)</th>
<th>cRMSE (°C)</th>
<th>R²</th>
<th>EMD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>JUNE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MUR</td>
<td>0.0</td>
<td>0.74</td>
<td>0.9</td>
<td>0.3</td>
</tr>
<tr>
<td>OSTIA</td>
<td>-0.3</td>
<td>0.65</td>
<td>0.93</td>
<td>0.38</td>
</tr>
<tr>
<td>GOES-16</td>
<td>-0.31</td>
<td>0.57</td>
<td>0.95</td>
<td>0.35</td>
</tr>
<tr>
<td><strong>JULY</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MUR</td>
<td>0.09</td>
<td>1.03</td>
<td>0.57</td>
<td>0.37</td>
</tr>
<tr>
<td>OSTIA</td>
<td>-0.17</td>
<td>0.7</td>
<td>0.8</td>
<td>0.37</td>
</tr>
<tr>
<td>GOES-16</td>
<td>-0.01</td>
<td>0.77</td>
<td>0.76</td>
<td>0.35</td>
</tr>
</tbody>
</table>
Over the course of June and July combined, looking across the entire buoy array, OSTIA overall outperforms GOES-16, as shown in Fig. 3. Both products have a relatively strong cold bias compared with observations at the three lidar locations. GOES-16 has a negative bias at five additional buoys, and OSTIA has a negative bias at one other buoy. Both SST products display warm biases at buoys 44017 (off the northeast coast of Long Island) and 44076 (the farthest offshore location considered, southeast of Cape Cod). Although GOES-16 follows the diurnal cycle rather than representing only the daily average SSTs, it still does not correlate with observations as well as OSTIA.
Specifically within the leasing areas (NYSERDA E05 & E06, Atlantic Shores, and buoys 44025, 44017, and 44065), OSTIA still delivers as the strong product, although GOES-16 at these sites alone delivers more promising average metrics than the average across all buoys.

### 3.2 Monthly Wind Speeds

The probability distribution functions (PDFs) of hub-height wind speeds at each lidar show that WRF, on a monthly time scale and with a 10-minute output resolution, generally captures the shape of the observed wind speed distribution at each lidar (Fig. 4), which indicates that the model itself is performing as it should. The wind speed distributions for each simulation maintain an even closer similarity in shape to one another, which helps highlight the biases directly related to the particular SST data set being used. A box plot of wind speeds across the entire domain, for each simulation and for both months, shows that although GOES-16 and OSTIA present near-identical average hub-height wind speeds, GOES-16 has a greater spread than OSTIA. Additionally, in both simulations, whole-domain winds in July tend to be significantly faster than June winds.
A comparison between each of the two simulations of monthly hub-height wind speed bias and correlation shows that although they perform overall similarly, GOES-16 tends to forecast 100-m wind speeds at each lidar with slightly more accuracy than OSTIA (e.g., Fig. 5). Performance at other heights varies, but, in general, GOES-16 continues to outperform OSTIA moving higher above sea level, which will be relevant for wind resource forecasting for larger offshore wind turbines. At each location, both products tend to over-predict winds at hub height and below. The correlation of each product’s forecasted winds at hub height with observations is above 0.65 for all times and locations.
A map of the domain showing the June and July average wind speed differences between the GOES-16 and OSTIA-forced simulations, overlaid by wind barbs depicting the average GOES-16 wind speeds, is shown in Fig. 6. In general, wind speeds deviate from each other only by small amounts on a monthly time scale. The results show maximum average wind speed differences between the two simulations of up to 0.25 m s\(^{-1}\) for each month.
3.3 Event-Scale Wind Speeds

Using the event selection algorithm detailed in Section 2.4, we locate three periods—one in June and two in July—during which 100-m wind speeds in the OSTIA and GOES-16 simulations differ from one another by statistically significant amounts and still both validate relatively well against observations ($r^2 > 0.5$ at 2 or more lidars), per the criteria outlined in Section 2.4. Lidars at which there is little observational data for these time periods are not considered. Therefore, only two locations for the June event and the second July event are considered. For all three cases, at each lidar being considered, GOES-16 delivers overall better predictions of 100-m wind speeds than OSTIA. Validation metrics vary at different heights, so vertical profiles of bias, $r^2$, RMSE, and EMD for each event can be found in Appendix A.
3.3.1 June 21-22, 2020 Event

We have identified an event beginning on 06-21-2020 at 13:40:00 and ending on 06-22-2020 15:20:00. Model output from the two simulations captured the observed trend of a wind speed increase—but model output wind speeds deviated significantly from one another leading into a first ramping event, during a second ramp event, and once wind speeds began to stabilize (Fig. 7). Observational data are only available at the two NYSERDA lidars during these periods. The Atlantic Shores lidar exhibited significant data gaps, so validation at this location is not conducted.

![Graph](https://example.com/graph.png)

**Figure 7.** Hub-height wind speed and wind direction at the NYSERDA E05 (a, c) and NYSERDA E06 (b, d) lidars during the June 21-22 event. Atlantic Shores is not shown due to a lack of observational data.

Validation metrics for the event at both lidars indicated stronger performance by GOES-16 simulation than OSTIA, particularly at vertical levels associated with typical offshore turbine hub heights. Validation metrics at hub height (100 m) for the two stimulations at NYSERDA E05 and NYSERDA E06 are shown in Fig. 8. Averaged correlation between the two sites is 0.75 for GOES-16 and 0.72 for OSTIA. The bias at each site is comparable between the two simulations. Model performance across the vertical should be taken into consideration if the rotor-equivalent wind speed is used to calculate power generation during extreme shear events as well as when looking at turbines of varying hub heights. Of note is that both RMSE and EMD are lower for GOES-16 at all heights (Appendix A).
Figure 8. Hub-height wind speed validation metrics at the NYSERDA E05 and NYSERDA E06 lidars during the June 21-22 event: bias (a), correlation (b), RMSE (c), and EMD (d).

The mean average difference in wind speeds throughout the domain is nominally zero. However, we can see via a planar depiction of wind speeds across the entire domain that, although the average difference is roughly 0.01 m s\(^{-1}\), the differences vary by location across the region. The maximum difference in average wind speeds during this event is actually 2.27 m s\(^{-1}\). This is noteworthy, as wind speed differences of this magnitude have significant implications regarding power generation.

During this event, offshore winds near the coast are southerly, with a tendency to follow the coastline as they rotate around a high-pressure system southeast of New Jersey. The differences in SST do not effect significant 100-m wind speed differences between the two simulations at the three lidars; however, in other areas of the domain, including a region planned for development just south of Rhode Island (Fig. 1), these differences are quite a bit larger (Fig. 9).
3.3.2 July 10-11, 2020 Event

We have next flagged an event that occurred between 07-10-2020 06:25:00 and 07-11-2020 09:00:00. Near-complete wind data sets exist at all three lidars for the duration of this event. The GOES-16 and OSTIA simulations output statistically significantly different 100-m wind speeds during this period. Both generally tracked the major observed wind speed patterns at each observation site, although the timing and magnitude of some longer-scale changes were missed by the models. For example, although the models captured the wind speed increase beginning at 07-10-2020 09:00:00 at Atlantic Shores, both erroneously forecasted a subsequent down- and then up-ramp event (Fig. 10(a)). And, at NYSERDA E05, the timing of a wind speed increase lagged observations by 6 hours, and a down-ramp event was missed altogether (Fig. 10(b)). These faults contribute to the relatively lower correlation of model output from both simulations with observations for this event.

Validation metrics for this event show a correlation of 0.5 or greater between model output and observations, for both simulations, at 100 m at each lidar site. Average $r^2$ is slightly stronger with GOES-16, which exceeded that of OSTIA by 0.02.
Figure 10. Hub-height wind speed and wind direction at the Atlantic Shores lidar (a, b) the NYSERDA E05 lidar (c, d), and the NYSERDA E06 lidar (e, f) during the July 10-11 event.
OSTIA only outperforms GOES-16 in terms of bias, although both present values very close to 0 m s\(^{-1}\). Hub-height (100 m) validation metrics for each simulation at each lidar are shown in Fig. 11.

![Wind speed validation metrics](image)

**Figure 11.** Wind speed validation metrics at an 100-m hub height at the Atlantic Shores, NYSERDA E05, and NYSERDA E06 lidars during the July 10-11 event: bias (a), RMSE (b), correlation (c), and EMD (d).

Synoptically, Tropical Storm Fay was observed to be moving in a southerly direction through the region during this time. The high wind speeds, which peaked at 07-10-2020 18:00:00, may be attributed to the storm. Average wind directions also reflect the storm path (Fig. 12). Differences in average wind speed between the two simulations peak at only 1.3 m s\(^{-1}\)—which is around 1 m s\(^{-1}\) less than the maximum difference during the June event. However, as with the June event, the average differences in average wind speeds across the domain are nominally zero.
3.4 July 18, 2020 Event

The last flagged event occurred between 07-18-2020 02:05:00 and 07-18-2020 14:10:00. Sufficient data for this event are present at Atlantic Shores and NYSERDA E06, but not NYSERDA E05. Over this time period, wind speeds drop at both locations from over 14 m s\(^{-1}\) to 3 m s\(^{-1}\) or lower. As with the previous two cases, the models capture the general wind speed trend of slowing throughout the duration of the event, although they both present errors in the magnitude of wind speeds and the timing of wind profile changes. Both simulations exhibit greater spread in wind speeds at each lidar than observations, and both have slower average velocities (Fig. 13).

Validation analysis for this event generally indicates that GOES-16 is the stronger product during this time, as it outperforms OSTIA in every metric at hub height except for correlation at NYSERDA E06. This superior performance holds for heights up to 100 m, at which point both of the products deliver similar results (bias, Atlantic Shores), or OSTIA begins to outperform GOES-16 (RMSE and EMD, NYSERDA E06). Of note is that PDFs of wind speeds during this time indicate bimodality in the
observed wind speed distribution. Although OSTIA somewhat captures this feature, GOES-16 misses it almost entirely—at both lidars. Validation metrics at Atlantic Shores and NYSERDA E06 are shown in Fig. 14.

Synoptically, this period was characterized by a cold front and rain having just moved northwesterly through the area, with a stationary front forming by the event’s end. This front spanned eastward into the Atlantic and aligned perpendicular to the New Jersey coast. Hub-height wind speed differences are varied throughout the entire domain, with a maximum difference in average wind speeds of 1.7 m s\(^{-1}\). Larger wind speed differences, in particular, reside near the coast, and positive/negative differences vary spatially. Again, the average difference in average event wind speeds throughout the domain is nominally zero. (Fig. 15).

4 Discussion

In Section 3, we saw varied performance between the two SST products, depending on the variable in consideration (SST or winds) and the time scale (monthly or event-scale). SST in OSTIA validated better than GOES-16 across the buoy array. However, hub-height wind speeds at each lidar location, for each month, point to a similar performance by both data sets. More nuanced events that occurred over hourly-to-daily time scales, which were selected based on differences between simulation
output and overall WRF performance, also indicate a stronger performance using GOES-16 SSTs. Events such as these are of importance to wind energy forecasts because we found that they often correlate with wind ramps, during which times wind speeds fall within the nonrated power section of turbine power curves. In this region, power output is very sensitive to fluctuations in wind speed. Therefore, obtaining the most accurate wind forecast within this regime is very important.

Planar maps of SST and hub-height winds indicate a localized relationship between wind speed, wind direction, and SSTs (Fig. 16). As mentioned in Section 3.3, despite the fact that wind speed differences in the domain tend to average out, there can still be significant deviations between the two in different regions of the domain, and this must be considered.
Although OSTIA SSTs at the lidars validate better than GOES-16, hub-height winds do not. This implies a more widespread impact on winds in the leasing area by SSTs across the full study area. More widespread in situ observations and validation of the two data sets against observations would help address how large of an effect this is. This would additionally allow for a better understanding of how much the capture of diurnal SST cycling (as in the GOES-16 product, Fig. 2) influences the accuracy of wind forecasting in the region.

Also of note are the differences in weather between June and July in the Mid-Atlantic region. June average wind speeds are faster than those in July for both simulations. June also had overall cooler SSTs throughout the region, with a weaker temperature gradient in the cold bubble offshore of Cape Cod than in July. The difference in temperature in this region between GOES-16 and OSTIA is more pronounced in July, with GOES-16 presenting warmer monthly average SSTs. The corresponding average wind speed and surface pressure differences between the products are also more distinct in July. While the two simulations validated similarly for each month, in the more turbulent July environment, OSTIA performs better. However, due to the greater number of missing pixels that required gap-filling in July as compared with June (36.86% more over the course of
each month, resulting from the increased cloud cover—and the possible stronger coastal cold pool), this indication of product superiority may be questionable.

GOES-16 SSTs without bias correction indicate a very clear warm bias throughout the entire domain, as compared with OSTIA (Fig. 16a,c). Without the added SSES field, the product clearly underperforms compared with OSTIA, so it should be included whenever using GOES-16 as a WRF input. The addition of more refined post-processing to account for incorrectly cloud-filtered pixels, as shown in Murphy et al. (2021), has been demonstrated to improve the GOES-16 SSTs even further. A comparison of simulations using OSTIA against those using the sophisticated GOES-16 product should be conducted in the future to evaluate the level of improvement delivered using the latter’s more computationally expensive data set.

5 Conclusions

In this study, we evaluated the performance of WRF to forecast winds off the Mid-Atlantic coast using two different input SST data sets. The first, OSTIA, has a 0.05° resolution, output at a daily interval. This product has been post-processed to fill any gaps in the data and assimilate it with in situ observations. The second product, GOES-16, has a finer resolution of 0.02° and is generated on an hourly interval. GOES-16 is gridded data but does not have the same level of post-processing as OSTIA. The data set contains gaps where abnormally cold pixels have been flagged as clouds and removed, and there is no assimilation with in situ observations. To account for the missing pixels, we ran an EOF process to statistically analyze trends in each monthly
data set and fill its gaps. Our findings in comparing model performance using these two data sets indicate that, while OSTIA SSTs validate better against buoy measurements, GOES-16 tends to forecast hub-height (100 m) winds with greater accuracy.

To compare the performance of the two products, WRF was run for June and July of 2020 in the Mid-Atlantic domain depicted in Fig. 1. The simulation setups are identical with the exception of the input SST field. We consider how SST values at specified buoys validate against observational data taken at these locations for each month. Specifically, we evaluate correlation, bias, RMSE, and EMD. Following the SST analysis, we examine hub-height wind speeds and validate model output against observational data taken at three separate lidars in the domain. We run the validation on both monthly and hourly-to-daily time scales. The shorter periods were identified using a flagging algorithm aimed to select periods during which there is significant difference in wind speed output from the two simulations and the general WRF performance is satisfactory.

The results show a better agreement between OSTIA and in situ SSTs than between GOES-16 and in situ SSTs. Despite this, modeled winds from the two simulations validate similarly on a monthly time scale, with GOES-16 performing slightly better in June and OSTIA performing slightly better in July. Three events were flagged using the aforementioned algorithm, and on those time scales, GOES-16 outperforms OSTIA in forecasting hub-height wind speeds.

This study shows that SST inputs to WRF do affect forecasted winds in future leasing areas in the Mid-Atlantic. There is indication that GOES-16, being of finer spatial and temporal resolution, is the better input product for forecasting offshore winds on the Mid-Atlantic coast. However, this step of the research is in its early stages and more cases need to be considered—on both monthly and event scales—before a more definitive evaluation can be made. In future work, the authors plan to continue identifying promising events, specifically including wind ramps, as they have shown to produce larger differences between GOES-16 and OSTIA-forced model output. Additionally, an advanced post-processing mechanism, as in Murphy et al. (2021), may improve GOES-16 performance even further, and a study including this method should be conducted.
Appendix A: Average validation metrics through 200 m for flagged cases

Contents of this Appendix include plots of horizontally averaged bias, $r^2$, RMSE, and EMD from sea level through a vertical height of 200 m for each flagged case, at each relevant lidar. Fig. A1 shows metrics at the two NYSERDA lidars for the June 21-22 event, Fig. A2 shows metrics at all three lidars for the July 10-11 event, and Fig. A3 shows metrics at Atlantic Shores and NYSERDA E06 for the July 18 event. The GOES-16 simulations are indicated by the blue lines, while the OSTIA simulations are shown in red. The labels AS, E05, and E06 refer to Atlantic Shores, NYSERDA E05, and NYSERDA E06, respectively.

Figure A1. Validation metrics at NYSERDA E05 and NYSERDA E06 during the June 21-22 event: bias (a), RMSE (b), correlation (c), and EMD (d). Hub height is marked by a horizontal line at 100 m.
Figure A2. Validation metrics at Atlantic Shores, NYSERDA E05, and NYSERDA E06 during the July 10-11 event: bias (a), RMSE (b), correlation (c), and EMD (d). Hub height is marked by a horizontal line at 100 m.
Figure A3. Validation metrics at Atlantic Shores and NYSERDA E06 during the July 18 event: bias (a), RMSE (b), correlation (c), and EMD (d). Hub height is marked by a horizontal line at 100 m.
Data availability.

Buoy data outside of the Atlantic Shores and NYSERDA floating lidars were collected via the National Oceanic and Atmospheric Administration’s NBDC. Satellite data were collected via the Physical Oceanography Distributed Active Archive Center.

Author contributions. SR downloaded and processed the GOES-16 data, ran the model for this SST input, and led the data analysis with contributions from MO, GX, and CD. MO designed the model setup, ran the model with OSTIA data, and contributed to the data analysis.

Competing interests. The authors state that there is no conflict of interest.

Acknowledgements. We thank Patrick Hawbecker (National Center for Atmospheric Research) for providing code to overwrite ERA5 SSTs with those of each satellite product. We also thank Sarah C. Murphy (Department of Marine and Coastal Sciences, Rutgers University), Travis Miles (Department of Marine and Coastal Sciences, Rutgers University), and Joseph F. Brodie (Department of Atmospheric Research, Rutgers University) for providing help and guidance with the DINEOF software.

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Wind Energy Technologies Office and by the National Offshore Wind Research and Development Consortium under Agreement No. CRD-19-16351. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes. A portion of the research was performed using computational resources sponsored by the Department of Energy’s Office of Energy Efficiency and Renewable Energy and located at the National Renewable Energy Laboratory.
References


