Machine Learning–Based Hurricane Wind Reconstruction

QIDONG YANG,[a,b] CHIA-YING LEE,[c] MICHAEL K. TIPPETT,[a] DANIEL R. CHAVAS,[d] AND THOMAS R. KNUTSON[e]

[a] *Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York*
[b] *Courant Institute of Mathematical Sciences, New York University, New York, New York*
[c] *Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York*
[d] *Department of Earth, Atmospheric and Planetary Sciences, Purdue University, West Lafayette, Indiana*
[e] *NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey*

(Manuscript received 18 May 2021, in final form 4 February 2022)

ABSTRACT: Here we present a machine learning–based wind reconstruction model. The model reconstructs hurricane surface winds with XGBoost, which is a decision-tree-based ensemble predictive algorithm. The model treats the symmetric and asymmetric wind fields separately. The symmetric wind field is approximated by a parametric wind profile model and two Bessel function series. The asymmetric field, accounting for asymmetries induced by the storm and its ambient environment, is represented using a small number of Laplacian eigenfunctions. The coefficients associated with Bessel functions and eigenfunctions are predicted by XGBoost based on storm and environmental features taken from NHC best-track and ERA-Interim data, respectively. We use HWIND for the observed wind fields. Three parametric wind profile models are tested in the symmetric wind model. The wind reconstruction model's performance is insensitive to the choice of the profile model because the Bessel function series correct biases of the parametric profiles. The mean square error of the reconstructed surface winds is smaller than the climatological variance, indicating skillful reconstruction. Storm center location, eyewall size, and translation speed play important roles in controlling the magnitude of the leading asymmetries, while the phase of the asymmetries is mainly affected by storm translation direction. Vertical wind shear impacts the asymmetry phase to a lesser degree. Intended applications of this model include assessing hurricane risk using synthetic storm event sets generated by statistical–dynamical downscaling hurricane models.

KEYWORDS: Wind; Machine learning; Regression

1. Introduction

Landfalling tropical cyclones (TCs) can threaten lives and bring severe economic losses to coastal societies (Geiger et al. 2016; Peduzzi et al. 2012). In addition to causing wind damage directly, TC surface wind contributes to coastal flooding through storm surge (Jordan and Clayson 2008; Needham and Keim 2014). In TC risk assessment, wind hazard is often defined as the probability of storm winds exceeding a given threshold at a particular location. The exceedance probabilities can be estimated using parametric wind models applied to thousands of synthetic storms that are derived using either statistical (e.g., Hall and Jewson 2007) or statistical–dynamical approaches (e.g., Emanuel et al. 2006; Lee et al. 2018; Jing and Lin 2020). A typical parametric wind model consists of an azimuthally symmetric wind field and asymmetric terms representing a small left-to-right asymmetry induced by storm motion (e.g., Lin and Chavas 2012). Observed surface wind fields, however, can be highly variable with differing asymmetries, especially for weaker storms, storms that encounter strong vertical wind shear, or storms that undergo extratropical transition (Loridan et al. 2015; Klotz and Jiang 2017). The discrepancy between parametric surface wind fields and observed surface winds motivates this study, and here we aim to address the problem of how to add missing asymmetries to existing parametric wind models in a manner that is consistent with observations.

Popular methods for generating the azimuthally symmetric field include empirical models such as Holland et al. (2010) and Willoughby et al. (2006), and theory-based ones, such as Chavas et al. (2015). These three models, referred as Holland10, Willoughby06, and Chavas15 hereafter, all provide a radial profile of the symmetric TC wind that sharply increases from storm center to the radius of maximum wind speed (R_{max}), and then gradually decreases outward to merge with the background wind. Wood et al. (2013) generalized this type of symmetric profile model by including multiple wind peaks. The representation of the asymmetric TC winds beyond storm motion-induced asymmetries has also improved over the years. Olfateh et al. (2017) combined Holland (1980) with the first azimuthal mode of wind asymmetry to represent impacts from the boundary layer friction, the blocking of high pressure systems, and the storm motion. Loridan et al. (2015) added a horseshoe-like asymmetry to Willoughby06 to describe wind structures associated with extratropical transitions. Uhlhorn et al. (2014) formulated asymmetric wind fields using wavenumber decomposition and approximated the full wind fields as the sum of the wavenumber-0 (symmetric) and wavenumber-1 components with dependence on vertical wind shear and storm translation. Also focusing on asymmetries induced by shear and storm translation, Chang et al. (2020) added wind asymmetries represented by harmonic functions to Holland10.

The above studies mostly aimed to capture asymmetric structures induced by one or two specific factors: wind shear, storm translation, or the extratropical transition. However,

Corresponding author: Qidong Yang, qy2216@columbia.edu

DOI: 10.1175/WAF-D-21-0077.1

TC wind asymmetries are induced by multiple factors, their interactions, and their nonlinear impacts on the vortex structure. Among all the factors, storm translation has long been recognized since Shapiro (1983) as having the clearest impact on TC wind asymmetries. Uhlhorn et al. (2014) showed that the environmental vertical wind shear influences surface wind asymmetries likely through its impacts on TC convection and boundary layer structure. Using composited scatterometer data from northwest Pacific typhoons, Ueno and Bessho (2011) showed a preferential left-of-shear and right-of-motion wind maximum. Wavenumber decomposition analysis from Uhlhorn et al. (2014) with stepped-frequency microwave radiometer data further showed that the storm motion-induced surface wind asymmetries rotate from downwind to right-of-wind as the storm moves faster, but the magnitude of the asymmetries is insensitive to the forward speed. Shear-induced asymmetries, on the other hand, rotate from downshear to left-of-shear with increasing shear magnitude, which is consistent with the findings of Klotz and Jiang (2017), who used global rain-corrected scatterometer wind data. Additionally, the blocking action from an anticyclone near a TC can cause a mesoscale condition analogous to the wall effect for a vortex (Olfateh et al. 2017), which impacts TC wind structure to certain degree. Land–sea roughness contrast (Wong and Chan 2007) and air–sea interaction (Lee and Chen 2012, 2014), as well as the interactions between TCs and midlatitude circulations (Komaromi and Doyle 2018), can influence TC surface wind structure as well.

In recent years, machine learning (ML) has made incremental progress in geophysical sciences. Several recent works have applied ML methods in TC studies. Racah et al. (2017) and Kim et al. (2019) utilized convolutional neural networks to detect TCs with weather model variables such as integrated water vapor. Our recent work, Yang et al. (2020), used a long-short-term memory model to predict TC rapid intensification with a set of storm and environment conditions as inputs and gives performance comparable to operational forecasts. Loridan et al. (2017) combined supervised (quantile regression forest) and unsupervised (principal component analysis) ML methods to model TC wind fields, suggesting that complex wind field asymmetries can possibly be captured. The successes of these ML applications stem in part from the increasing amount of observed and simulated data and the modeling capacity provided by ML algorithms, which allow for approximation of any nonlinear relationship. These works thus inspire us to apply advanced ML methods to the problem of improving modeled wind fields by adding physical dependencies that are missing in existing parametric modeling approaches. Specifically, we propose an XGBoost-based reconstruction of the symmetric and asymmetric components of the TC surface wind field based on environment and storm features as predictors. First proposed by Chen and Guestrin (2016), XGBoost is a decision-tree-based ensemble predictive algorithm. The model uses an optimization method called gradient boosting (Friedman 2001) that adjusts decision trees (Quinlan 1986) to minimize the difference between XGBoost's prediction and the ground truth. XGBoost has been widely utilized in classification and regression tasks on tabular datasets. Here we use it because it achieves outstanding performance on structured datasets and can be trained very fast with parallel processing.

The remainder of this paper is organized as follows. The observed storm, environment, and the HWIND data used to train and test the XGBoost models are described in section 2. Feasibility of the HWIND data for wind field reconstruction model development is evaluated in section 3. The model development details are specified in section 4. The performance of the reconstruction model is discussed in section 5, along with a wind field reconstruction case study. Section 6 provides further analysis on the relationship between wind field asymmetries and storm and environment variables, which play a crucial role in the reconstruction model. We summarize and discuss our findings in section 7.

2. Data and methods

a. Tropical cyclones and their environment

Here we develop the wind reconstruction model using the same set of storm and environment variables as used in Yang et al. (2020). The storm variables are derived from the National Hurricane Center (NHC; Landsea and Franklin 2013) best-track data, and the environment variables are taken from the monthly European Centre for Medium-Range Weather Forecasts interim reanalysis (ERA-Interim; Dee et al. 2011) database. The monthly environment variables are then linearly interpolated to storm time stamps. We use monthly environmental variables because our intended application is TC wind risk assessment in a changing climate using a statistical–dynamical downscaling approach that combines large-scale climate information from dynamical models or reanalysis with statistical models to produce storm-scale features (e.g., Lee et al. 2020). In this application, we need to generate wind fields for thousands of synthetic storms over a long period (e.g., 1951–2100), and with climate forcing from many global models. Using monthly environmental variables reduces the data storage burden. This also removes the need to filter modeled or observed storms from environmental variables. Of course, this approach will not capture changes that are due to sub-monthly environmental variability.

A total of 30 variables (8 storm variables and 22 environment variables) are used in this work. Storm variables describe the storm status at the time of a wind field snapshot and include quantities such as storm maximum wind speed, storm center location, and storm translation (Table 1). Environment variables, computed from a $0.25° \times 0.25°$ global grid, depict the large-scale conditions in the neighborhood of the storm and include vertical wind shear, relative humidity, outflow temperature, high-level divergence, and so on as listed in Table 2. We represent storm translation and vertical wind shear by their zonal and meridional components (i.e., SH_z and SH_m for shear and MT_z and MT_m for storm motion) to avoid the inherent ambiguity of angular variables, which can make their use in ML models cumbersome. Some environment variables are

TABLE 1. Storm variables.

Notation	Definition
lat	Storm center latitude
$V_{\max}$	Maximum wind speed of azimuthally averaged wind field
$R_{\max}$	Radius where $V_{\max}$ is achieved
$S_{\max}$	Storm maximum wind speed
dS/dt	12-h change of storm max wind speed
MT	Storm translation speed magnitude
MT_z	Storm translation speed in zonal direction in Earth coordinates
MT_m	Storm translation speed in meridional direction in Earth coordinates

averaged over a small disk centered on the storm and others are averaged over a large annulus. Disk-averaged variables are averaged over the area within 500 km from storm center while annulus-averaged ones are averaged over a ring-shaped area with inner radius of 200 km and outer radius of 800 km from storm center. The former averaging method is intended to capture environment conditions around storm center while variables averaged over a large annulus aim to describe surrounding conditions in which the storms are embedded.

TC intensity and storm structure are not unrelated—strong storms are more symmetric than weaker ones (Klotz and Jiang 2016, 2017; Sun et al. 2019). While those aforementioned variables are originally derived for TC intensity prediction (Lee et al. 2015, 2016; Yang et al. 2020), they are suitable for storm structure prediction as well. For example, vertical wind shear is typically a negative factor for intensification by tilting the storm vortex and transporting dry air into the inner core. A known impact of the above vertical wind shear process on TC structure is that the strongest convection focuses in the downshear-left quadrant. Uhlhorn et al. (2014) showed that this convective asymmetry may result in surface wind asymmetries when the shear is sufficient. The work by Knaff et al. (2017) also illustrates a successful example of predicting storm structure using intensity-related variables. Besides, as suggested in Loridan et al. (2017), more complex TC environmental features which characterize the atmospheric flow, such as the relative humidity and upper-level divergence, should also be considered for better surface wind field prediction.

Two additional variables, the maximum azimuthally averaged wind speed ($V_{\max}$) and its corresponding radius ($R_{\max}$) that are often used in the parametric wind profiles are included in the reconstruction model development, too. In this study, we derive $V_{\max}$ and $R_{\max}$ from azimuthally averaged HWIND profiles and regard them as known. In practice, $V_{\max}$ can be estimated by a linear fit of storm maximum wind speed ($S_{\max}$) provided by best track data, and $R_{\max}$ can be obtained through an empirical equation from Knaff et al. (2015), theory-based model from Chavas and Lin (2016), or a combination of theory

TABLE 2. Environment variables. Variables presented here are derived at storm center, averaged over a disk, or averaged over an annulus. Disk-averaged variables are averaged over the area within 500 km from storm center while annulus-averaged ones are averaged over a ring-shaped area with inner radius of 200 km and outer radius of 800 km from storm center.

Notation	Definition
PI_p	Potential intensity in minimum sea level pressure at the storm center
$\overline{\mathrm{PI}_{p_d}}$	Potential intensity in minimum sea level pressure, disk averaged
PI_s	Potential intensity in maximum wind speed at storm center
$\overline{\mathrm{PI}_{s_d}}$	Potential intensity in maximum wind speed at storm center, disk averaged
T_{100}	Ocean temperature averaged over top 100 m, at storm center
$\overline{T_{100_d}}$	T_{100} but averaged over 100 km storm-centered, disk averaged
T_{200}	Temperature at 200 hPa at storm center
$\overline{T_{200_d}}$	T_{200}, annulus-averaged
SH	Vertical wind shear magnitude
SH_z	Vertical wind shear (zonal) between 850 and 200 hPa at storm center in Earth coordinates
$\overline{\mathrm{SH}_{z_a}}$	SH_z, annulus averaged
SH_m	Vertical wind shear (meridional) between 850 and 200 hPa at storm center in Earth coordinates
$\overline{\mathrm{SH}_{m_a}}$	SH_m, annulus averaged
D_{200}	Divergence field at 200 hPa
$\overline{D_{200_a}}$	D_{200}, annulus averaged
$d\mathrm{PI}_p/dt$	Change in potential intensity of minimum sea level pressure over 12 h
RH_l	Low level relative humidity (850–700 hPa) at storm center
$\overline{\mathrm{RH}_{l_a}}$	RH_l, annulus averaged
RH_h	High level relative humidity (500–300 hPa)
$\overline{\mathrm{RH}_{h_a}}$	RH_h, annulus averaged
SLM	Storm location mark: land or ocean
$P - S$	$\overline{\mathrm{PI}_{s_d}} - S_{\max}$

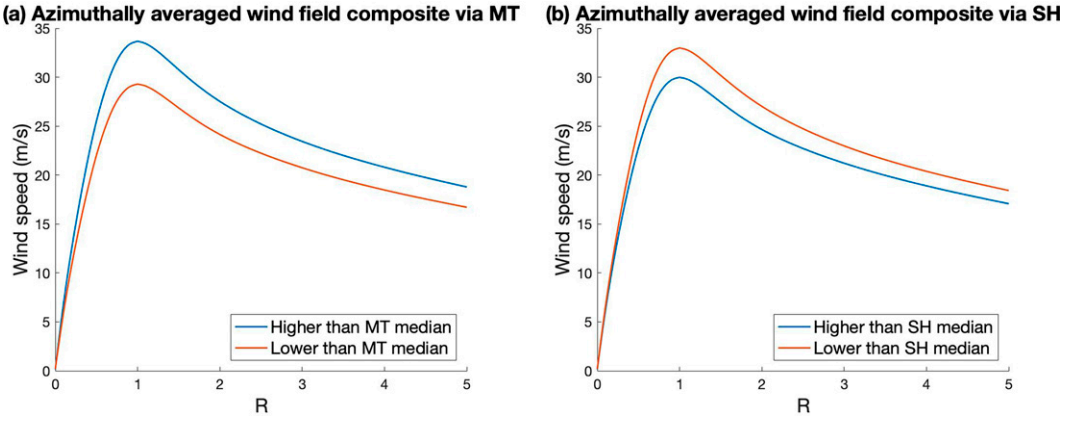

FIG. 1. (a),(b) Radial wind speed profile composites of the azimuthally averaged HWIND wind fields. The x axis shows radii normalized by radius of maximum wind speed (R_{max}). In (a) and (b), blue (orange) curves are composites with translation speed MT and shear magnitude SH greater (less) than the respective medians from the sample data.

and data (Chavas and Knaff 2022). A recent study by Knaff and Chavas (2021) notes that estimating R_{max} without direct reconnaissance measurements is challenging.

b. HWIND

The observed TC wind fields are from the Hurricane Research Division hurricane surface wind analysis system [HWIND, now Risk Management Solutions (RMS) HWind as per the RMS website] (e.g., Powell et al. 1998; DiNapoli et al. 2012). These wind fields are produced by interpolating and smoothing wind speed observations from multiple platforms including stepped-frequency microwave radiometers, satellites, GPS dropsondes, ships, buoys, and land-based observation stations. We use HWIND data within radii ranging from 0 to 300 km from storm center because radii up to 300 km are generally large enough to cover a storm's main structure. Data at normalized radius (radius divided by R_{max}) are linearly interpolated. Here R_{max} denotes the radius where the HWIND azimuthally averaged wind speed maximum achieved. From 2000 to 2014,[1] the dataset contains 1539 Atlantic hurricane surface (10 m) wind speed field snapshots from 112 hurricanes. Since the size of the HWIND dataset is limited and ML-based models have a high demand for training data, a higher than usual proportion of the dataset is used for model training: 93.75% of the HWIND storms (105 storms, 1410 wind field snapshots) are randomly chosen and used as training set. The remaining 6.25% of the storms (7 storms, 129 wind field snapshots) are saved as an independent testing set for model performance evaluation. In terms of the number of wind field snapshots, over 8% of them are used for performance evaluation.

An additional experiment with a second kind of training-testing set split (by wind field snapshots) was conducted (not shown). In this alternative approach, we randomly chose 92% of snapshots as the training set and the rest as the testing set. The evaluation results from this splitting method were even more promising. However, wind fields from the same storm are correlated in time because similar surface winds affecting storm-scale and environmental conditions will continue over multiple snapshots and result in similar snapshots for the same storm. Splitting that data by snapshot may unfairly put testing set answers into the training set, i.e., the testing data are not independent of the training data. Consequently, the alternative splitting approach leads to an overfitting on training data, and the resultant reconstruction model will have weak generalization ability on truly independent data. Therefore, we choose splitting by storms rather than splitting by wind field snapshots.

[1] Data after 2014 are not publicly available.

c. Skill scores

Model performance is evaluated on the independent testing dataset by computing the mean square error skill score (MSESS) of various predicted quantities. The MSESS of N forecasts $\hat{y}_i$ with verifying observations y_i is

$$\mathrm{MSESS} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - y_{\mathrm{ref}})^2}, \tag{1}$$

where y_{ref} is a reference forecast. Here we use either the climatological value (average) of y_i in the training dataset or the value predicted by a parametric wind profile model. Positive MSESS indicates that the reconstruction model has skill greater than the reference forecast; negative MSESS indicates otherwise. The maximum possible value of MSESS is 1, meaning perfect predictions.

3. A feasibility test of wind fields in HWIND

The fact that HWIND fields are an interpolated estimate of the reality and are smoother than simulations from high-resolution weather prediction models (Klausmann 2014; Done et al. 2020) raises the question of whether HWIND captures known relationships with the two dominant factors: storm translation and vertical shear.

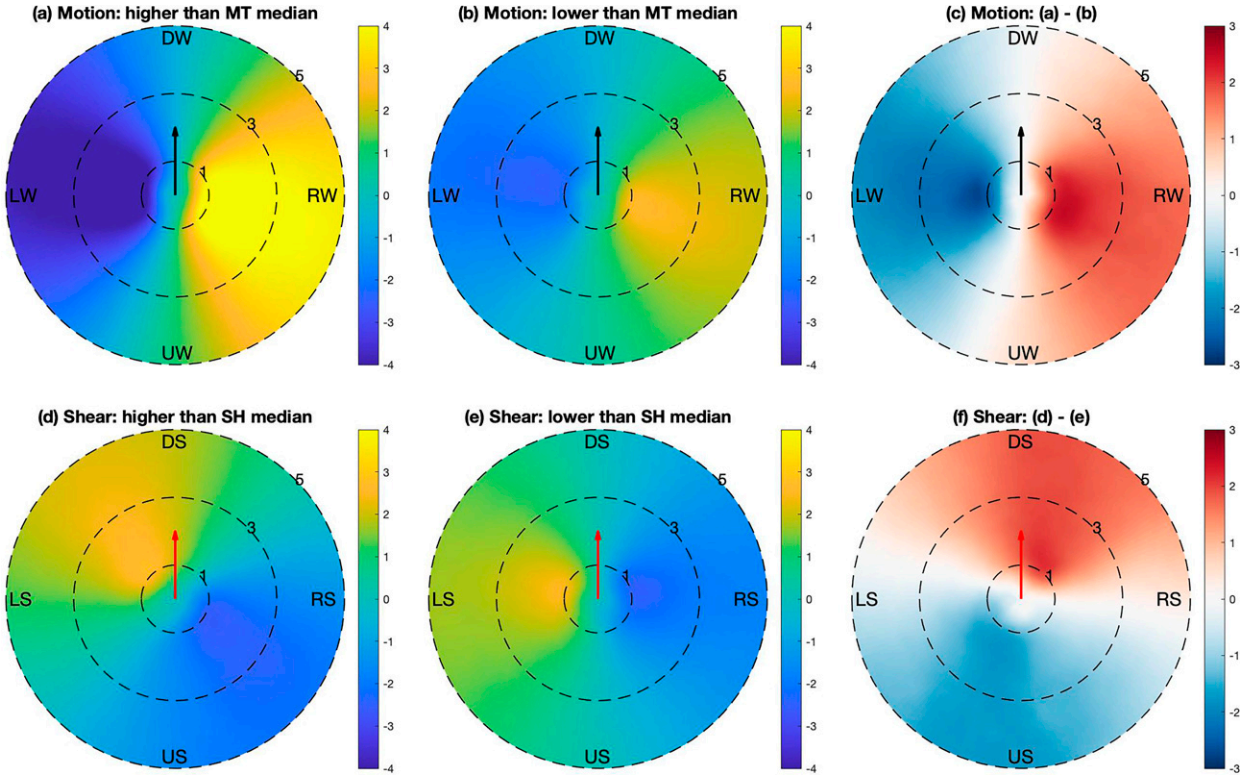

FIG. 2. HWIND asymmetric wind field composites based on storm translation speed MT and vertical wind shear magnitude SH. The dashed circle and the number beside it represent the normalized radius. The black and red arrows denote storm motion and wind shear directions. The composites are constructed in (a)–(c) motion and (d)–(f) shear coordinates. In motion coordinates, up is downwind (the storm translation direction, DW), down is upwind (UW), left is left-of-wind (LW), and right is right-of-wind (RW). For the shear coordinates, up is downshear (the vertical wind shear direction, DS), down is upshear (US), left is left-of-shear (LS), and right is right-of-shear (RS). Panels (a) and (b) show composites with MT greater and lower than the MT median while (d) and (e) are for those with SH greater and lower than the SH median. Panels (c) and (f) are the differences between (a) and (b) and (d) and (e).

Figure 1 shows that the wind speed composite of azimuthally averaged HWIND is larger for larger translation magnitude (MT), which is in agreement with the study by Mei et al. (2012), who showed a strong positive correlation between storm intensity and motion. On the other hand, HWIND data for storms in more sheared environments (SH greater than the median) tend to have lower symmetric wind speeds because of the negative impact of vertical wind shear on storm intensity as discussed previously.

The asymmetry composites of four cases (fast versus slow motion, and high versus low shear) are shown in Fig. 2 in motion and shear coordinates, by which we mean the asymmetries are shown relative to the storm motion and shear directions. Here, fast versus slow motion cases are analyzed in the motion coordinates, and the high versus low shear cases are studied in the shear coordinates. There is a right-of-motion and left-of-shear asymmetry maximum in Figs. 2a,b and Figs. 2d,e, respectively, which is in agreement with findings from Ueno and Bessho (2011). The magnitude of asymmetries, not the phase of asymmetries, in HWIND increases with translation speed magnitude (Figs. 2a–c), which differs from Uhlhorn et al. (2014), who show the storm translation speed impacts the phase of the asymmetries but not the magnitude. It is possible that the analysis assumptions used in HWIND or its input data (such as flight-level wind) lead to such dependence. Interestingly, Uhlhorn et al. (2014) showed that flight-level wind asymmetry magnitude increases with the storm translation speed. This dependence was also found in Mueller et al. (2006), who showed that there is an improvement in the estimation of flight-level wind using infrared satellite data when storm motion is considered. Figures 2d–f reveal no strong connection between asymmetry magnitude and vertical shear magnitude (SH), but do show the phase of the asymmetries rotates from left-of-shear to downshear-left with increasing SH. Figures 1 and 2 indicate that HWIND symmetric and asymmetric wind fields depend realistically on storm motion and shear, and that it is reasonable to use HWIND in developing a wind reconstruction model.

4. Model development

a. *Reconstruction model design*

The surface wind field is the sum of a symmetric and an asymmetric component:

$$\text{wind field} = \text{symmetric field} + \text{asymmetric field}. \quad (2)$$

The symmetric field is approximated by a parametric wind profile model (e.g., Chavas15) along with two series of Bessel functions. The Bessel functions serve as parametric profile corrections by fitting to the residual between the azimuthally averaged wind speed and the parametric profile. One series of Bessel functions is for the inner region $r_p \le R_{\max}$, and the other one is for the outer region $r_p \ge R_{\max}$; r_p is the profile radius ranging from 0 to R_u (i.e., 300 km). This approximation is summarized in the following equation:

$$\text{azimuthal average} - \text{parametric profile} = \text{residual profile} \approx \sum_{n=1}^{\infty} A_n J_0\left(\lambda_n \frac{r_p}{R_{\max}}\right)\mathbf{1}(r_p \le R_{\max}) + \sum_{n=1}^{\infty} B_n J_0\left(\lambda_n \frac{R_u - r_p}{R_u - R_{\max}}\right)\mathbf{1}(r_p \ge R_{max}), \quad (3)$$

where the $J_0(\cdot)$ is the zero-order Bessel function and λ_n is the nth positive root solving $J_0(\cdot) = 0$; $\mathbf{1}$ is the indicator function, which is one when its argument is true and zero otherwise. The terms A_n and B_n are the coefficients associated with the inner and outer series of Bessel functions. The coefficients are obtained by area-weighted least squares fitting to the HWIND data. The functional dependence of the outer series on r_p means that the Bessel function is reversed and shifted so that it is zero at $r_p = R_{\max}$. The reason we use separate series of Bessel functions to fit residual profile before and after $R_{\max}$ is to ensure that the fitted residual profile is zero at $R_{\max}$. Thus, by construction, the Bessel function corrected parametric profile matches the targeted azimuthal average at $R_{\max}$.

The asymmetric field accounts for asymmetries induced by the storm and its ambient environment. We decompose it using Laplacian eigenfunctions on the unit disk:

$$\begin{aligned}\text{asymmetric field} &\approx \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \text{Asymmetry}_{m,n} \\ &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} a_{m,n} H_a(m,n) + b_{m,n} H_b(m,n),\end{aligned} \quad (4)$$

where $a_{m,n}$ and $b_{m,n}$ are the coefficients associated with the eigenfunctions $H_a(m, n)$ and $H_b(m, n)$. The coefficients are computed using area-weighted least squares and the eigenfunctions are defined as

$$\begin{aligned} H_a(m,n) &= N_{m,n} J_m(\lambda_{m,n} r)\cos(m\theta), \\ H_b(m,n) &= N_{m,n} J_m(\lambda_{m,n} r)\sin(m\theta), \end{aligned} \quad (5)$$

in which $N_{m,n}$ denotes the normalization factor such that $\int\int H(m,n)^2 r dr d\theta = 1$; r and θ are the unit radius and azimuth angle in Earth coordinates; $J_m(\cdot)$ is the mth-order Bessel function of the first kind, and $\lambda_{m,n}$ is the nth positive root of $J_m(\cdot) = 0$. The index m corresponds to the wavenumber in the azimuthal direction, and the index n plays a similar role but in the radial direction [see $H_a(m, n)$ and $H_b(m, n)$ plot examples in Fig. 10].

b. Symmetric field approximation

To approximate the symmetric field, we first explore three commonly used parametric wind profile models: the Holland10, Willoughby06, and Chavas15 models. The Holland10 model is an improved version of Holland (1980) which tends to overestimate wind decay with radius and leads to an underestimation in wind speeds at storm's outer regions (Willoughby and Rahn 2004). The Willoughby06 model is composed of three segments of piecewise-continuous wind profiles: the inner eyewall segment where wind speed increases in proportion to a power of radius, the outer eyewall segment which is designed to decay exponentially with radial e-folding distance, and a radially varying polynomial ramp function that concatenates the first two segments. The Chavas15 model consists of two theoretical solutions to the structure at the top of the boundary layer in the inner ascending (Emanuel and Rotunno 2011) and outer descending (Emanuel 2004) region.

These wind profile models take as inputs the maximum wind speed ($V_{\max}$) and its corresponding radius ($R_{\max}$) to generate radial profiles of azimuthally averaged wind speed from storm center to outer radii. Additional profile shape adjusting parameters (e.g., peripheral wind observation and its associated radius) can also be provided to improve the fit. In this work, all three models take in $V_{\max}$ and $R_{\max}$, and the Willoughby06 and Chavas15 models additionally depend on storm center latitude and Coriolis parameter, respectively. In Holland10, external pressure, central pressure, peripheral wind radius, and peripheral wind observation are set to 1005 hPa, 950 hPa, 300 km, and 17 m s^{-1}, respectively, according to the profile baseline in Holland et al. (2010). As for Chavas15, the drag coefficient is obtained following Donelan (2004); the ratio of the exchange coefficients of enthalpy and momentum is computed by a quadratic fit from Chavas et al. (2015); an empirical adjustment suggested in Chavas et al. (2015) is applied to profile eye; and radiative-subsidence rate is fixed to 2×10^{-3} m s^{-1}. Chavas15 can also be simplified by setting Coriolis parameter constant to 5×10^{-5} s^{-1}. It changes resultant parametric profiles, but does not affect the finally approximated symmetric fields after Bessel function correction.

Profile examples from these three models are illustrated in Fig. 3. By construction, all three match the observed wind speed at $R_{\max}$. The Holland10 and Chavas15 models have a better fit inside $R_{\max}$ than the Willoughby06 model, but they overestimate wind speed outside $R_{\max}$. The distance outside $R_{\max}$ makes up roughly two thirds of the whole modeling radius range. Consequently, in a mean square error sense (not area weighted), the Willoughby06 model performs relatively better than the other two in this example. Figure 4 further shows the root-mean-square error (RMSE) of the three wind profile models over the HWIND dataset. Willoughby06 and Chavas15 perform similarly for 1–3 $R_{\max}$;

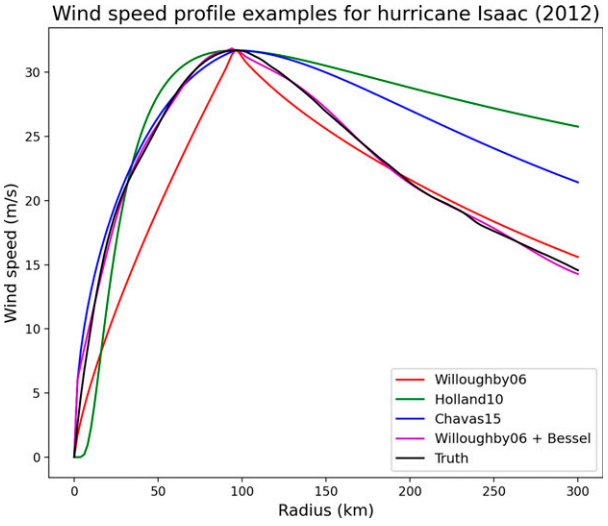

FIG. 3. An azimuthally averaged wind profile of Hurricane Isaac (2012) from HWIND (black), and the corresponding simulated profiles from Willoughby06, Holland10, and Chavas15 denoted by the red, green, and blue curves, respectively. The magenta curve is the profile approximated by Willoughby06 along with two series of Bessel functions.

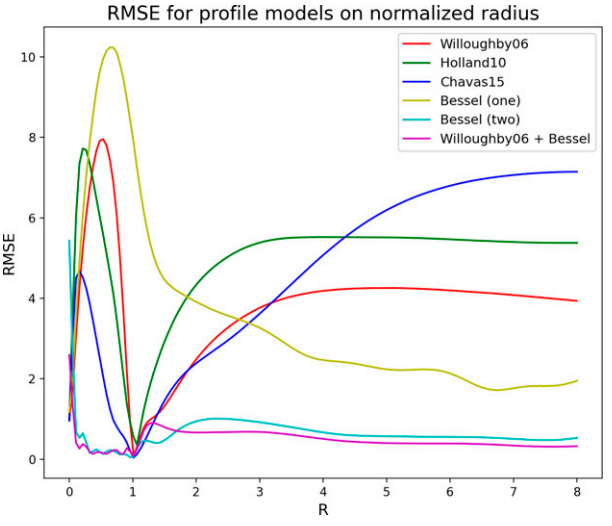

FIG. 4. The RMSE of the Willoughby06 (red), Holland10 (green), and Chavas15 (blue) models over HWIND dataset in normalized radius. The yellow (cyan) curve is the RMSE from the profiles fitted by a single (two) series of Bessel functions, while the magenta curve is from the profiles approximated by the Willoughby06 model along with two series of Bessel functions.

Chavas15 is better for 0.5–1 $R_{\max}$ and Willoughby06 is better beyond 3 $R_{\max}$. We note that TC wind damage happens disproportionately due to the strongest winds which occur near $R_{\max}$. The Chavas15 model outperforms both the Willoughby06 and Holland10 models in terms of RMSE from 0 to 3 $R_{\max}$. Nevertheless, as mentioned previously, two series of Bessel functions serve to further improve the representation of the symmetric wind component. Thus, what we require is only a profile model that gives a reasonable starting point, and all three models explored here are adequate for this purpose. In the rest of the paper, we will use the Willoughby06 model as an example.

The other ingredients of the symmetric wind modeling here are the two series of Bessel functions [Eq. (3)] for modeling the residual symmetric profile inside and outside $R_{\max}$. In practice, we must truncate these two series. Our experiments show that the residual profiles (either before or after $R_{\max}$) can be approximated sufficiently well using the first four terms. Thus, the azimuthally averaged wind speed approximation formula is

$$\text{azimuthal average approx.} \equiv \sum_{n=1}^{4} A_n J_0\left(\lambda_n \frac{r_p}{R_{\max}}\right)\mathbf{1}(r_p \le R_{\max}) + \sum_{n=1}^{4} B_n J_0\left(\lambda_n \frac{R_u - r_p}{R_u - R_{\max}}\right)\mathbf{1}(r_p \ge R_{\max}) + \text{parametric profile}. \quad (6)$$

Then the symmetric field can be produced by unfolding (i.e., repeating) the estimated profile azimuthally. The effectiveness of this truncation is verified in Figs. 3 and 4. In Fig. 3, the profile simulated by Eq. (6) (magenta curve) is close to the observed profile (black curve) both before and after $R_{\max}$. The RMSE of Eq. (6) in Fig. 4 (magenta curve) has the lowest value roughly throughout the whole normalized radius range. Note that the Bessel function coefficients used in these two figures are obtained from fitting to observations (as opposed to being predicted as we will discuss later). In other words, the demonstrated results show the error due to truncation, which represents a lower bound on the prediction error.

We also tried to use only Bessel functions to approximate wind profiles (with no parametric profile involved). In Fig. 4, the yellow curve represents RMSE from profiles fitted by a single Bessel function series (eight terms) for the full radial range and the cyan curve is from profiles fitted by two series of Bessel functions (four terms for $r_p \le R_{\max}$ and four terms for $r_p \ge R_{\max}$). The yellow curve shows rather large RMSE. In contrast, the cyan curve is quite close to the magenta curve. However, it does not imply that parametric profiles are not necessary. In practice, the Bessel function coefficients from symmetric models need to be predicted by XGBoost. The benefit of including the parametric profiles is that the wind reconstruction model is always given a reasonable profile to start with. Even though XGBoost gives poor Bessel function coefficient predictions, the final result will not be totally unreasonable. Therefore, a parametric profile combined with the two Bessel series is still the optimal choice.

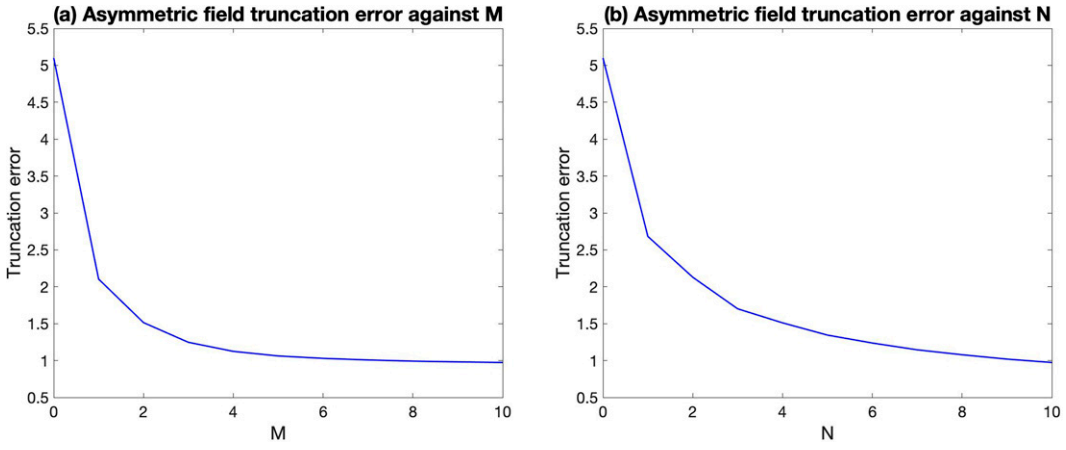

FIG. 5. The truncation error of asymmetric fields with increasing (a) M and (b) N. The M and N are the upper limits of the sums in Eq. (4); N and M are kept as constants as 10 in (a) and (b), respectively.

c. *Asymmetric field approximation*

Next, we consider the approximation of the asymmetric field by truncating the eigenfunction series in Eq. (4). As M and N increase, higher-frequency asymmetries are added, and truncation error decreases (Fig. 5). The truncation error decreases slowly after $M = 3$ and $N = 4$; the asymmetric modes that appear most useful in approximating the asymmetric fields are $\text{Asymmetry}_{m,n}$ with $m \in \{1, 2, 3\}$, $n \in \{1, 2, 3, 4\}$.

The asymmetric field approximation then can be defined as

$$\text{asymmetric field approx.} \equiv \sum_{m=1}^{3}\sum_{n=1}^{4} a_{m,n}H_a(m,n) + b_{m,n}H_b(m,n). \tag{7}$$

Substitution using Eqs. (4) and (5) shows that

$$\begin{aligned}\text{Asymmetry}_{m,n} &= N_{m,n}J_m(\lambda_{m,n}r)\left[a_{m,n}\cos(m\theta) + b_{m,n}\sin(m\theta)\right] \\ &= M_{m,n}N_{m,n}J_m(\lambda_{m,n}r)\cos\left[m(\theta - P_{m,n})\right],\end{aligned} \tag{8}$$

where the magnitude $M_{m,n}$ and phase $P_{m,n}$ are

$$M_{m,n} = \sqrt{a_{m,n}^2 + b_{m,n}^2},\ P_{m,n} = \frac{1}{m}\,\text{atan2}\frac{b_{m,n}}{a_{m,n}}. \tag{9}$$

In particular, atan2(·) is the four-quadrant inverse tangent. Figure 6 shows that the three asymmetries with largest magnitude in the HWIND dataset are $\text{Asymmetry}_{1,1}$, $\text{Asymmetry}_{1,2}$, and $\text{Asymmetry}_{2,1}$.

d. *The complete reconstruction model*

The reconstructed wind is the sum of the symmetric and asymmetric wind approximations. There are two sets of unknowns: Bessel function coefficients (A_n and B_n) in the symmetric approximation and eigenfunction coefficients ($a_{m,n}$ and $b_{m,n}$) in the asymmetric approximation. Together there are 32 coefficients to be determined: A_n and B_n for n = 1,2,3,4, and $a_{m,n}$ and $b_{m,n}$ for $m = 1, 2, 3$ and $n = 1, 2, 3, 4$. Each coefficient is predicted separately by an XGBoost model trained with the input variables listed in Tables 1 and 2. The hyperparameters of XGBoost models (e.g., learning rate and max depth of trees) are tuned on the training set using five-fold cross validation. The chosen hyperparameters are listed in Table 3. Then the XGBoost models are trained on the complete training set. The performance of the reconstruction model is last evaluated on the independent testing set.

It is emphasized that in this study, the wind reconstruction model is developed using HWIND fields in Earth coordinates with radii up to 300 km from storm center. Normalized radius grid as well as motion and shear coordinates are only used for analyzing and evaluating the input data and the simulated results.

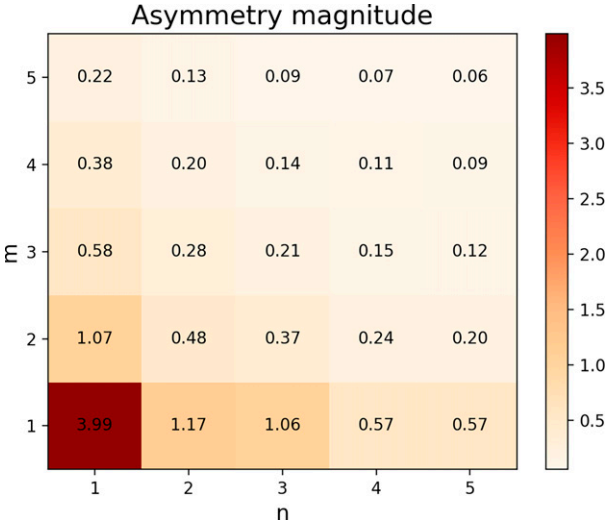

FIG. 6. Magnitude of asymmetries at m and n.

TABLE 3. XGBoost model configuration example.

Hyperparameter	Value
Learning rate	0.1
The number of estimators	50
Maximum depth of a tree	2
Minimum sum of instance weight (Hessian) needed in a child	3
Gamma	2
Subsample ratio of the training instances	0.78
Subsample ratio of columns when constructing each tree	0.77
L1 regularization term on weights	1.11
L2 regularization term on weights	1

5. Performance evaluation

a. Coefficient prediction

The MSESS values for each predicted coefficient are listed in Table 4. The coefficients associated with the symmetric component (A_n and B_n) are predicted more accurately than those associated with the asymmetric component ($a_{m,n}$ and $b_{m,n}$). Lower wavenumber asymmetries (smaller m and n) are more accurately predicted than higher wavenumber ones (bigger m and n).

b. Case studies

We illustrate the reconstruction model using a wind field snapshot from a storm that is not included in the training set, Hurricane Isaac (2012), as a case study. We first compare the HWIND observation (Fig. 7a) with the optimal reconstructed wind field (Fig. 7b), which is when the coefficients in Eqs. (6) and (8) are computed from the HWIND data directly. This comparison shows that the optimal reconstruction captures the main characteristics of the HWIND data, including the magnitude and the location of the peak wind. The difference between the HWIND observation and the optimal reconstruction is the approximation error (Fig. 7d) which contains length scales and structures that the reconstruction model cannot resolve. The XGBoost reconstructed wind (Fig. 7c) is, by subjective observation, a satisfactory match to the HWIND observation, but with wind speed underestimation around wind peak. The largest positive differences (red) between the XGBoost reconstructed wind and the HWIND observation are to the southeast (upwind) of the storm center (Fig. 7e), and the largest negative differences (blue) are over land to the northwest (downwind). The errors over land suggest that the input variables do not provide enough information to capture abrupt wind speed decreases over land due to land–sea roughness contrast. Differences between the XGBoost reconstructed wind and the optimal reconstructed wind are similar in pattern and magnitude but spatially smoothed (Fig. 7f). Overall, the proposed model is able to reconstruct wind fields with acceptable error in the open sea with a caveat of tending to overestimate wind speed over land. It is because the sharp transition from smooth ocean to rougher land will be smoothed out by the model.

TABLE 4. MSESS for predicted coefficients by XGBoost models on testing set.

MSESS	$n = 1$	$n = 2$	$n = 3$	$n = 4$
A_n	0.78	0.85	0.19	0.33
B_n	0.31	0.49	0.54	0.61
$a_{1,n}$	0.50	0.11	0.44	0.13
$a_{2,n}$	0.04	0.04	0.00	0.01
$a_{3,n}$	−0.02	0.00	0.00	0.01
$b_{1,n}$	0.09	0.06	0.08	0.08
$b_{2,n}$	−0.42	−0.04	−0.05	0.00
$b_{3,n}$	−0.02	0.00	0.01	0.00

c. Reconstructed surface wind

We compute the MSESS for the reconstructions of the symmetric, asymmetric, and full wind fields (Fig. 8). Climatologies used for the reference forecast in Eq. (1) are calculated from the training set. For the symmetric fields, we used the symmetric field climatology from HWIND (Fig. 8a) and Willoughby06 parametric fields (Fig. 8b) as the reference forecast, respectively. Basically all the MSESS values are positive with respect to climatology, and values greater than 0.9 concentrate within 3 $R_{\max}$ (Fig. 8a). In Fig. 8b, MSESS values are positive almost everywhere on the disk except at the center (white dot) and $R_{\max}$ (white ring). To match the ground truth, we manually set the reconstructed symmetric fields to have zero wind speed at storm center, which results in the white dot at disk center. The white ring at $R_{\max}$ is due to the Willoughby06 wind profiles matching $V_{\max}$ at $R_{\max}$, hence it is impossible to improve performance at $R_{\max}$. The highest MSESS values are within $R_{\max}$, where the Bessel function series in the symmetric model corrects the Willoughby06 profile's inner eye and the XGBoost models predict the corresponding coefficients accurately. In Figs. 8c–e, the same reconstructed asymmetric fields are aligned in Earth, motion, and shear coordinates (i.e., corresponding the disk upward represents Earth north, translation direction, and shear direction) and the MSESS is computed with respect to the climatology in the corresponding coordinate system. The MSESS values are positive at most areas of the disks. MSESS values are lowest in motion coordinates, which indicates that the climatology in motion coordinates explains relatively more variability. The low and negative MSESS values (the dark blue and the white space) tend to concentrate at east and west in Earth coordinates, upwind-right and downwind-left in motion coordinates, and downshear-left and upshear-right in shear coordinates. These patterns vary when different testing sets are used. Figures 8f–h show the MSESS of the reconstructed full wind fields in the three coordinate systems using corresponding full wind field climatologies as the reference field. Most of the areas inside 3 $R_{\max}$ on the three disks have MSESS values greater than 0.8 (yellow). This is the region where winds are strong, and severe TC wind damages usually occur, and also the most challenging area for parametric models due to the turbulent nature of moist convection. It is, however, the most skillful region for the XGBoost-based wind reconstruction model.

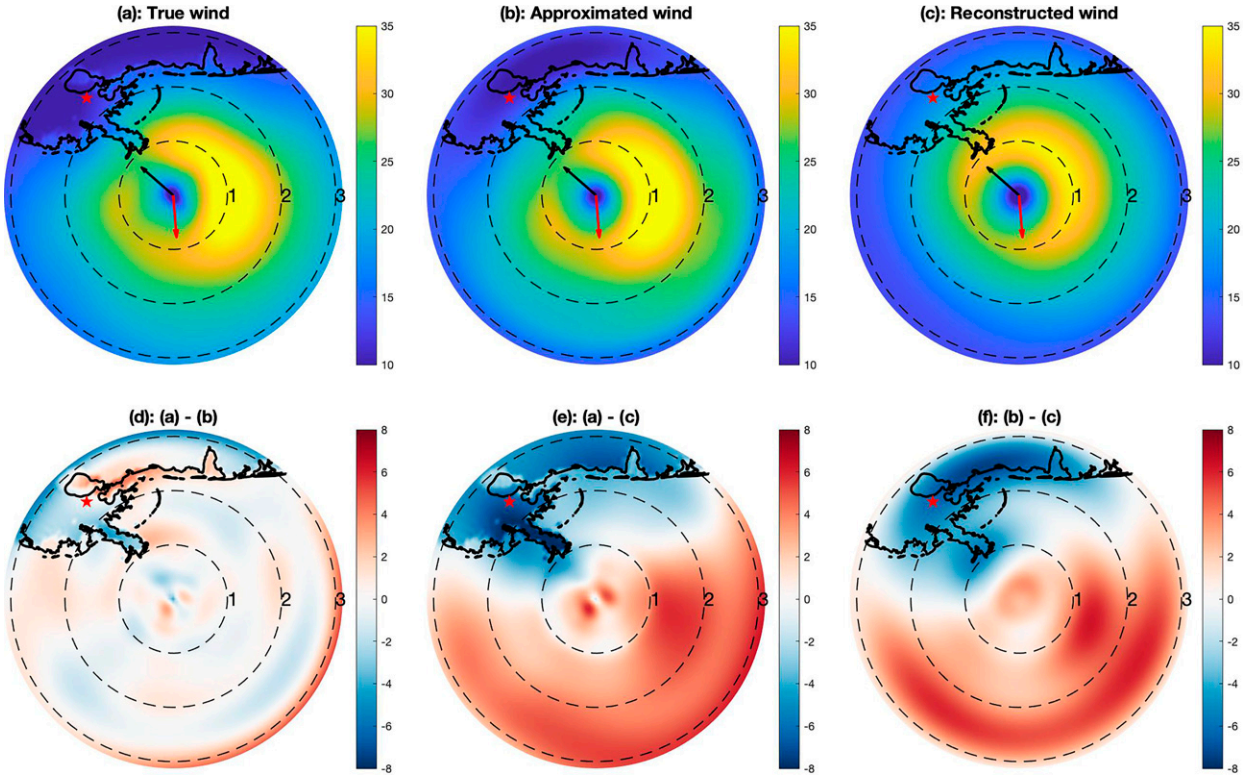

FIG. 7. A reconstruction case study of one wind field from Hurricane Isaac at 1600 UTC 28 Aug 2012. Plotted wind fields are in Earth coordinates with radii up to 300 km from storm center, and the dashed circle and the small number beside it indicate the locations of 1, 2, and 3 $R_{\max}$. Land coastline is also plotted, and the red star marks New Orleans, LA. Black and red arrows indicate the translation and shear directions, respectively. (a) HWIND truth, (b) approximation, and (c) reconstruction. (d)–(f) Errors from the approximation (against the truth), the reconstruction (against the truth), and the XGBoost prediction space (against the approximation).

6. Model interpretation

Now we analyze the dependence of the symmetric and asymmetric winds fields on predictors to gain insights about the physical controls of TC surface wind structures. To do so, we developed a sequence of diagnostic XGBoost models, which are developed the same way as the reconstruction XGBoost models, but with predictors added one at a time ordered by how much each predictor reduced model's loss function (i.e., the difference between prediction and observation), going from largest reduction to least reduction. For the symmetric field, we focused on the first terms of Bessel function series, $A_1J_0(\lambda_n r)$ and $B_1J_0(\lambda_n r)$, which correct the parametric profile. For the asymmetric field, we focused on three asymmetric modes with relatively large magnitudes: Asymmetry$_{1,1}$, Asymmetry$_{1,2}$, and Asymmetry$_{2,1}$.

a. *Symmetric field*

As input variables are added, the performance of the diagnostic models improves and eventually stabilizes. We define essential variables (marked in red in Figs. 9 and 10) as those that provide the bulk of the model improvement. Note there are variables in black between two variables in red. It suggests that the variables in black are related with their preceding variables in red (i.e., already recognized as essential), hence they appear relatively unessential. For example, $V_{\max}$ (black) locates between $R_{\max}$ (red) and $\overline{T_{200_d}}$ (red) in Fig. 9b. It implies $V_{\max}$ is correlated with $R_{\max}$. As a result, they achieve similar loss function reduction and are ordered consecutively. It is not hard to imagine that the performance of diagnostic XGBoost with only $R_{\max}$ included is close to the performance with both $V_{\max}$ and $R_{\max}$ included due to these two variables' correlation. Although outflow temperature ($\overline{T_{200_d}}$) does not achieve loss function reduction as much as $V_{\max}$, it brings new information into diagnostic XGBoost. Thus, when it is included, it increases diagnostic XGBoost performance more than $V_{\max}$. Therefore, $V_{\max}$ is illustrated as unessential relative to $R_{\max}$ and $\overline{T_{200_d}}$.

In Fig. 9a, for the inner eye coefficient A_1, nearly all the diagnostic model improvement comes from the first variable $V_{\max}$, hence $V_{\max}$ is the only essential variable for predicting A_1 and provides most of the information needed to correct the Willoughby06 inner eyewall profile. Further analysis also shows that there exists a rather strong positive linear relationship between $V_{\max}$ and A_1, which reflects the key role that $V_{\max}$ plays in setting inner-eyewall wind structure. In the case of outer eye coefficient B_1 (Fig. 9b), the essential variables are $R_{\max}$, $\overline{T_{200_d}}$, dS/dt, $\overline{\mathrm{RH}_{h_a}}$, and lat. This collection is much

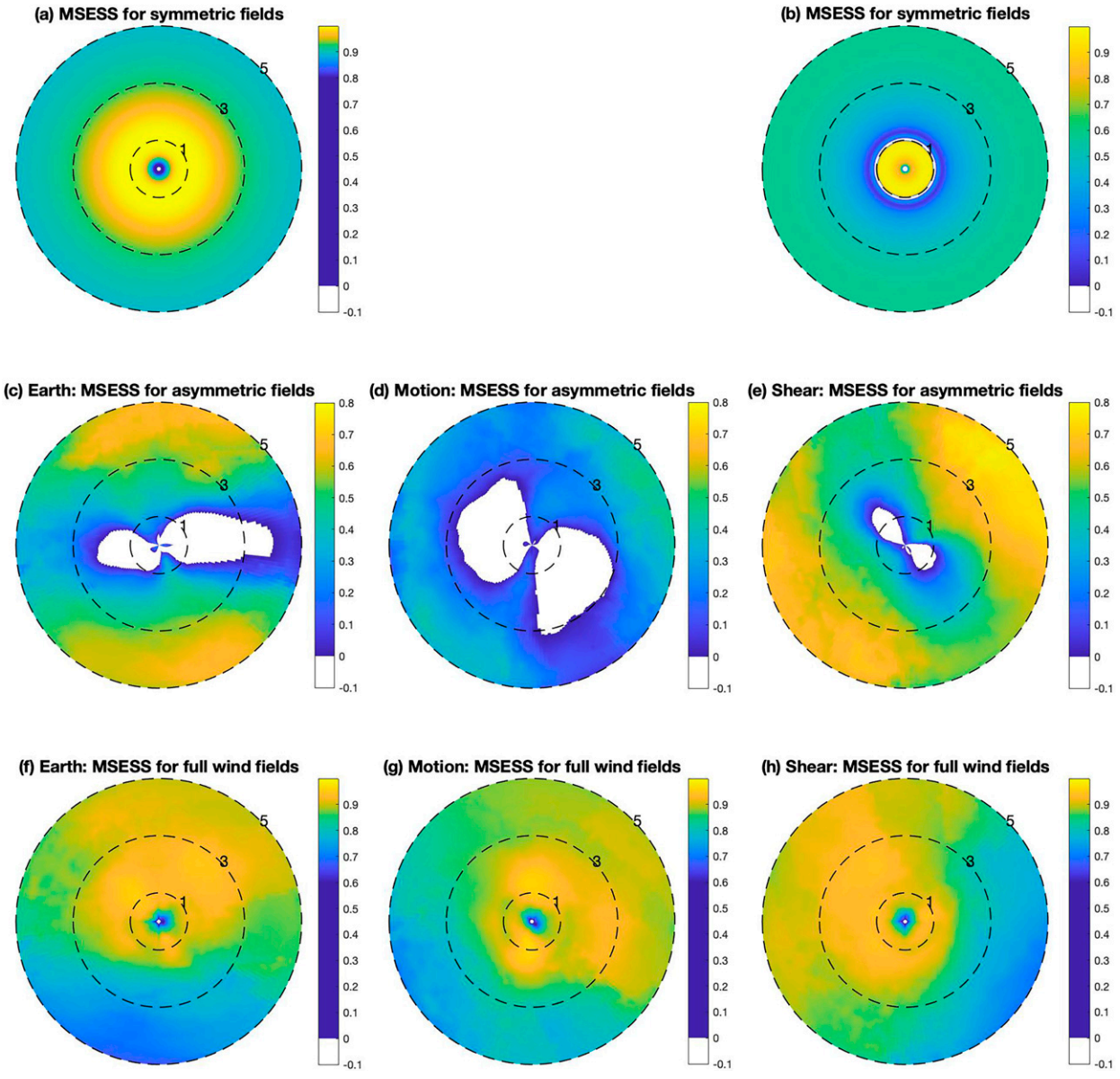

FIG. 8. Mean square error skill score (MSESS) of XGBoost reconstructed wind from testing set on normalized radius. (a),(b) The MSESS of the reconstructed symmetric fields comparing to symmetric field climatology and Willoughby06 parametric fields. (c)–(e) The MSESS of the reconstructed asymmetric fields comparing to the HWIND climatology in Earth, motion, and shear coordinates. (f)–(h) As in (c)–(e), but for the full (symmetric + asymmetric) wind fields. Larger positive MSESS values indicate where the XGBoost reconstruction has relatively larger skill compared to the HWIND climatology in the indicated reference coordinates.

more diverse and includes information from various sources such as storm intensity change (dS/dt), inner core size ($R_{\max}$) and location (lat), and outflow temperature ($\overline{T_{200_d}}$).

b. *Asymmetry magnitude*

Next, we use coefficients predicted by diagnostic models to construct asymmetries. In Fig. 10, there is a general reduction of R^2 with increasing m and n, which agrees with the conclusion from Table 4 that the predictability of higher wavenumber asymmetries is lower than that of lower wavenumber ones.

Figures 10a–c demonstrate how the magnitudes of the dominant asymmetries ($M_{1,1}$, $M_{1,2}$, and $M_{2,1}$) relate with input variables. The R^2 curve of $M_{1,1}$ roughly stops increasing after $\overline{T_{200_d}}$ suggesting that the essential variables for predicting $M_{1,1}$ are MT, lat, MT_z, $\overline{\mathrm{RH}_{l_a}}$, and $\overline{T_{200_d}}$. In particular, storm translation variables (MT and MT_z) contribute over half of the performance improvement by all essential variables. This behavior is not surprising since storm translation is a

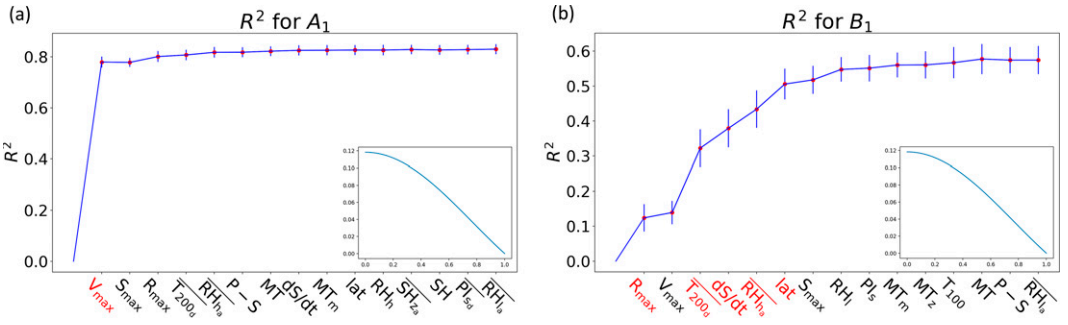

FIG. 9. The R^2 of Bessel function coefficients A_1 and B_1, which are responsible for parametric wind profile correction inside and outside the R_{max}, respectively. The R^2 is calculated from predictions of the diagnostic XGBoost models that are trained and run with inclusion of one more predictor at a time. The red dots represent mean R^2 from tenfold cross validation and the vertical bars denote the variance of the validation. Along the x axis only the first 15 variables are presented, and those in red are regarded as essential. The Bessel function associated with A_1 and B_1 is also plotted at each panel's corner.

determining factor on wind field asymmetry (e.g., Ueno and Bessho 2011; Sun et al. 2019). Including humidity and temperature variables ($\overline{RH_{l_a}}$ and $\overline{T_{200_d}}$) supports the suggestions from Loridan et al. (2017) that the inclusion of additional environmental conditions may have positive impact on surface wind estimation. The appearance of lat may be explained by its climatological connection with environmental conditions such as temperature and humidity. Essential variables for $M_{1,2}$ are R_{max}, lat, MT, and T_{200}. MT and lat are present here roughly for the same reason as in the case of $M_{1,1}$. R_{max} contributes storm size information. Storm center temperature (T_{200}) has a strong connection to storm intensity and storm center latitude. For $M_{2,1}$, the important variables are: RH_l, R_{max}, T_{200}, RH_h, and lat, which all have been recognized as important in the previous cases of $M_{1,1}$ and $M_{2,1}$.

To sum up, the main storm features affecting asymmetry magnitude are storm location, inner core size, and translation speed, of which translation speed plays the dominant role. Environmental variables related to humidity and temperature also contribute to the asymmetry magnitude prediction to some degree.

c. *Asymmetry phase*

1) R^2 ANALYSIS

Asymmetry is also parameterized by phase [$P_{m,n}$, Eq. (8)], that is, how each asymmetry rotates in response to a storm or environment variable. It can be defined in Earth, motion, and shear coordinates (relative to north, storm translation, and wind shear directions), unlike the magnitude which is the same across coordinate systems. Here we stick with the Earth coordinates as our wind reconstruction model, and follow the same procedure as the last section to analyze how asymmetry phase in Earth coordinates relates with input variables.

Figures 10d–f show the R^2 analysis on asymmetry Earth phase. The R^2 curve of $P_{1,1}$ levels off at the third variable. In particular, the first variable, MT_z, brings in the biggest increase of R^2 indicating zonal direction components of storm motion contain most of the information necessary to predict $P_{1,1}$. Other essential variables for $P_{1,1}$ are MT_m, lat, and SH. It is noteworthy that in terms of wind shear–related variables (e.g., SH_z, SH_m, and SH), only SH is barely recognized as an essential predictor. It indicates that wind shear does not affect $P_{1,1}$ as strongly as reported in previous studies (e.g., Uhlhorn et al. 2014), or wind shear is highly related with storm motion. When it comes to $P_{1,2}$, five variables: lat, $\overline{T_{100_d}}$, $P-S$, $\overline{RH_{l_a}}$, and $\overline{SH_{z_a}}$, are essential. Besides wind shear, storm location (lat), temperature ($\overline{T_{100_d}}$), and humidity ($\overline{RH_{l_a}}$) variables are also identified as important. It is not surprising considering they have already been shown to be connected to asymmetry magnitude in the last section. As for the inclusion of intensity variable ($P-S$), it can be explained by its dependency on wind shear. For $P_{2,1}$, the R^2 curve reveals that the important variables are lat, $\overline{SH_{m_a}}$, $P-S$, $\overline{RH_{l_a}}$, SH, V_{max}, and RH_l. This time one more storm intensity variable (V_{max}), which has strong correlation with $P-S$, is included.

In summary, storm center latitude is important for the prediction of all three phases. $P_{1,1}$ is mainly controlled by storm motion. For $P_{1,2}$ and $P_{2,1}$, more diverse information is needed for their prediction including wind shear, storm intensity, temperature, and humidity.

2) CIRCULAR DISTRIBUTION ANALYSIS

Next, we conduct analysis on circular distributions of asymmetry phases to show how they vary with the predictors identified in the R^2 analysis. The circular distribution analysis on $P_{1,2}$ and $P_{2,1}$ does not reveal any strong predictor correlation, and only results for $P_{1,1}$ (i.e., the phase of the asymmetric wind component for wavenumber one in the azimuthal and radial directions) are presented here.

Figure 11 demonstrates circular distributions of $P_{1,1}$ conditional on storm translation speed magnitude (MT), direction (MT_θ), vertical wind shear magnitude (SH), direction (SH_θ), and shear direction relative to translation direction (SH_θ^{MT}). In particular, MT_θ and SH_θ are defined in Earth coordinates ranging from $-180°$ to $180°$ such that $0°$, $90°$, $180°$ (or $-180°$), and $-90°$ represent east, north, west, and south, respectively. Note that we also shifted $P_{1,1}$ into motion and shear coordinates to facilitate phase-predictor relationship illustration.

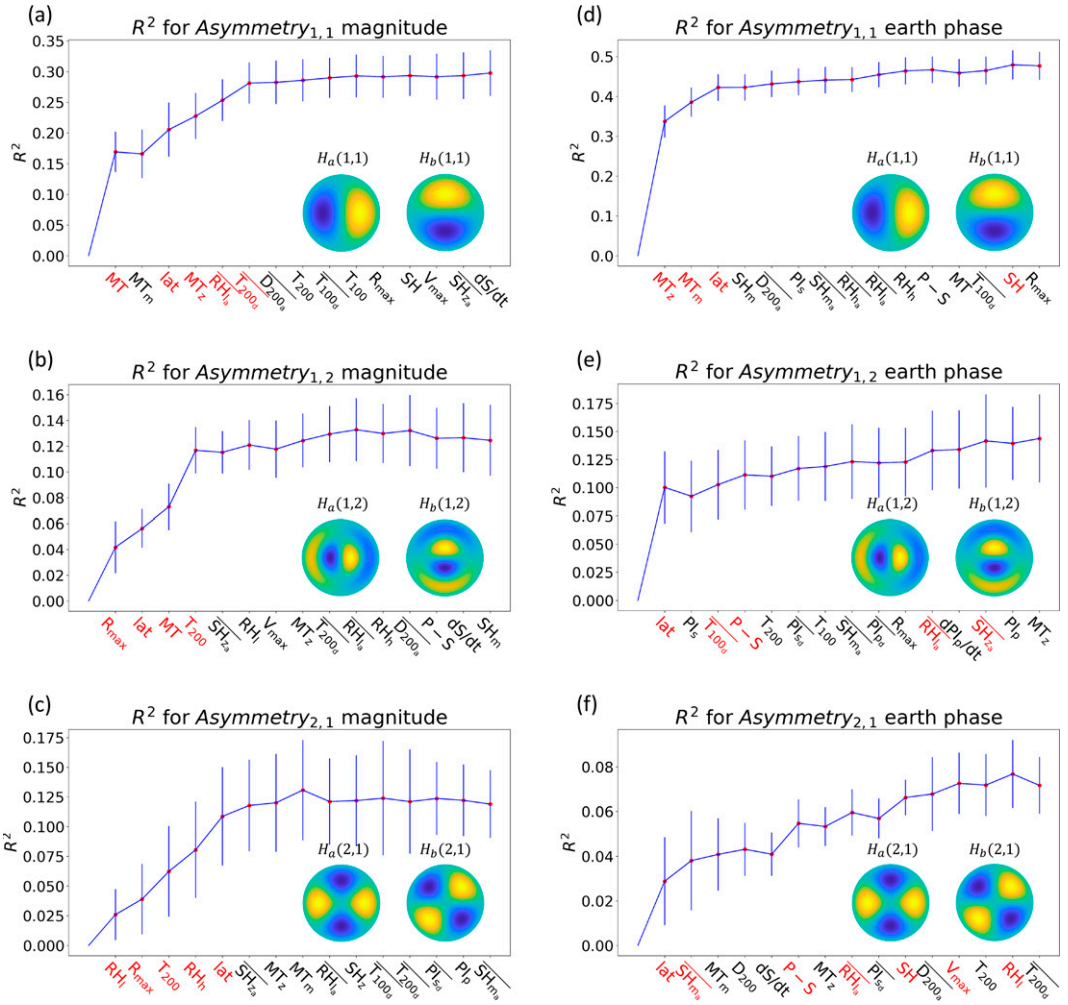

FIG. 10. As in Fig. 9, but for asymmetry (a)–(c) magnitude and (d)–(f) phase, which are constructed from the coefficients predicted by the diagnostic XGBoost models. Note that a different vertical axis scaling is used for each panel in the figure. Panel (a) shows the dominant contribution of storm translation speed magnitude (MT) to the explained variance (R^2) for magnitude of the asymmetric TC wind component for wavenumber one in the azimuthal and radial directions. Panel (d) shows that the zonal component of storm motion (MT_z) is the most important factor for predicting the phase of the wavenumber one asymmetry in the azimuthal and radial directions. The eigenfunctions [$H_a(m, n)$ and $H_b(m, n)$] of each asymmetry are also shown at the corner of each panel.

In Fig. 11a, we analyze $P_{1,1}$ conditional on translation speed magnitude (MT). D1 represents cases from the lowest quartile of values of MT (i.e., slowest moving storms) and D4 represents cases from the highest quartile of values of MT (i.e., fastest moving storms). The mean MT of D1 cases is 2 m s^{-1} and that of D4 is 7 m s^{-1}, as indicated by small black numbers on the lower-right part of the panel. From D1 to D4, storm moving speed increases, but $P_{1,1}$ stays at right-of-motion all the time, which suggests that MT does not affect $Asymmetry_{1,1}$'s rotation relative to storm motion direction. Figure 11b shows that $P_{1,1}$ rotates from left-of-shear to downshear clockwise as wind shear magnitude (SH) increases. However, in Fig. 11c, $P_{1,1}$ remains at right-of-motion regardless of the increment of SH, which implies that the pattern observed in Fig. 11b is primary contributed by storm motion direction.

In terms of translation direction (MT_θ) in Fig. 11d, D1 denotes cases from the lowest quartile of values of MT_θ in which storms move toward east-northeast (MT_θ mean of 31°), and D4 denotes cases from the highest quartile of values of MT_θ where storms move toward west-northwest (MT_θ mean of 167°). Even though from D1 to D4 storm motion heading changes from east to west counterclockwise, $P_{1,1}$ is always peaked at right-of-motion. Storm moving direction is again shown to be the dominant variable controlling $P_{1,1}$ in motion coordinates. In contrast, storm-moving Earth-relative direction is not a dominant variable controlling $P_{1,1}$. Figure 11e shows that $P_{1,1}$ rotates from left-of-shear to downshear with

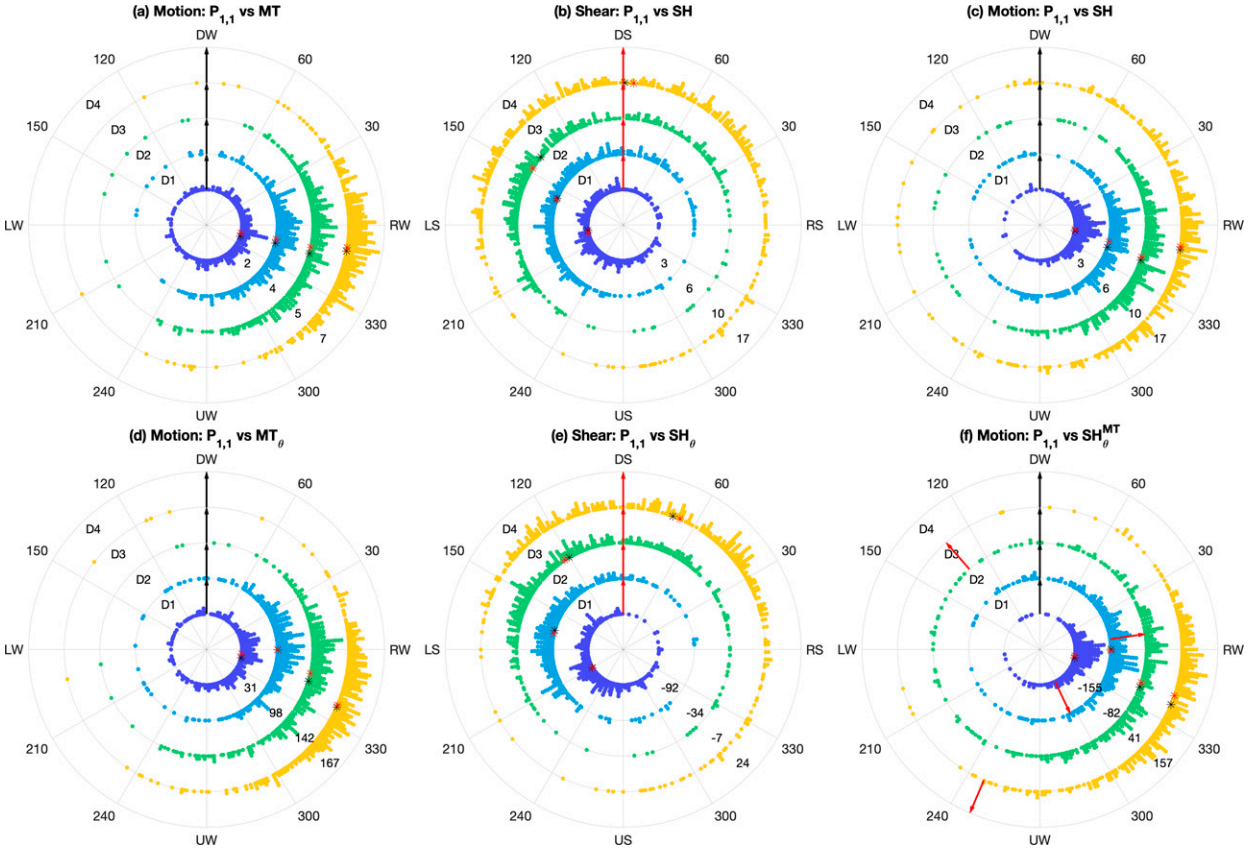

FIG. 11. The $P_{1,1}$ is the phase of the asymmetric wind component for wavenumber one in the azimuthal and radial directions. This figure shows histograms of $P_{1,1}$ conditional on a variable of interest as indicated in each panel's title. The panel title also shows which coordinate the $P_{1,1}$ is in (i.e., motion or shear coordinates). D1 (purple) shows a histogram of $P_{1,1}$ conditional on a variable of interest at its 0–25 percentiles. D2 (blue), D3 (green) and D4 (yellow) are for the 25–50, 50–75, and 75–100 percentiles, respectively. At the lower-right corner of each panel, we list the mean of the variable of interest at each percentile interval. For each interval we also mark the corresponding $P_{1,1}$ mean and medium with black and red stars. Shear and storm motion direction of each panel (if applicable) are marked with red and black arrows. See text for examples and further details.

shear heading varying from south (−92°) to east-northeast (24°) counterclockwise. Again, this pattern is proven to be actually driven by storm motion direction by Fig. 11f, which demonstrates that $P_{1,1}$ concentrates at right-of-motion regardless of shear direction. That is, the shear direction relative to the translation direction covers a wide range of different angles across the percentiles (see red arrows), yet the wind field asymmetry $P_{1,1}$ (in motion coordinates) is consistently to the right of motion regardless of the shear direction relative to translation direction.

To sum up, $P_{1,1}$ derived from HWIND dataset locates at right-of-motion all the time regardless of the variation of shear magnitude, shear direction, translation speed magnitude, and translation's Earth relative direction. The observed rotation patterns of $P_{1,1}$ in shear coordinates can also be partially explained by storm motion direction. Therefore, storm motion direction is the dominant predictor controlling the rotation of $\text{Asymmetry}_{1,1}$. However, when a storm is nearly stationary, the secondary dominant predictor, wind shear, plays the most important role.

7. Discussion and conclusions

A downscaling hurricane wind reconstruction method using XGBoost is developed and evaluated in this study with HWIND data over the period 2000–14. This method divides the wind field into symmetric and asymmetric components and models them separately. The symmetric component is approximated by a parametric wind profile along with two series of Bessel functions. The asymmetric component is expanded in a Laplacian eigenfunction series. Then XGBoost models are trained with storm and environment features as inputs to predict the coefficients associated with the Bessel functions and eigenfunctions.

We experimented with three existing parametric wind profile models: Holland10, Willoughby06, and Chavas15. The RMSE between Willoughby06 and azimuthally averaged HWIND profiles, with RMSE maximum around 4, was the lowest of the three parametric wind profile models in the range of 1–8 $R_{\max}$. Chavas15 is the better performing one with RMSE maximum less than 4 over the range of 0.5–3

R_{max}, where winds are relatively strong. Since the Bessel function series were able to correct all three parametric models sufficiently well, the final performance did not depend on which model was used.

Reconstruction performance was assessed in terms of mean square error skill score (MSESS) computed on an independent subset of the data (i.e., testing set) with respect to a reference reconstruction. For symmetric fields, when the reference reconstruction was the parametric profile, the XGBoost symmetric reconstruction had substantially smaller errors than the parametric wind profile model, especially within R_{max} giving MSESS up to 0.8. The XGBoost symmetric reconstruction achieved higher MSESS when symmetric climatology was used as the reference reconstruction. In the case of the asymmetric wind component, the model is skillful in reconstructing low-wavenumber asymmetries but not high-wavenumber ones. The model's error is mainly concentrated to the east and west in Earth coordinates, upwind-right and downwind-left in motion coordinates, and downshear-left and upshear-right in shear coordinates. In terms of full wind fields, using climatology as reference reconstruction, the model has positive MSESS roughly over all the areas, but MSESS is particularly high in the range from 0.5 to 3 R_{max}, where severe TC damage concentrates. However, some details of the performance patterns observed here may be specific to the relatively small number of storms in the testing dataset.

In addition to wind reconstruction, the model is also a tool for analyzing relationships between wind asymmetries (magnitude $M_{m,n}$ and phase $P_{m,n}$) and predictors. Asymmetry magnitude is mainly affected by storm location, inner core size, and translation speed, while asymmetry phase is mostly controlled by vertical wind shear and storm motion direction. Further asymmetry phase circular distribution analysis showed that the wavenumber-1 asymmetry ($P_{1,1}$) rotates from left-of-shear to downshear clockwise when shear gets stronger or shear heading moves from south to east. However, $P_{1,1}$ always concentrates to right of storm motion no matter how its associated predictor varies. In other words, the rotation of $P_{1,1}$ is dominantly controlled by storm motion direction.

Ultimately, we would like to apply this XGBoost-based wind reconstruction model to generate wind fields for synthetic storm tracks generated by statistical–dynamical downscaling models (e.g., Lee et al. 2018, 2020), which output storm track data along with other environmental conditions. Our approach is different from other ML-based wind models such as the one developed by Loridan et al. (2017) who, except for the storm ambient pressure, used only storm structure parameters including those hard to obtain from a typical synthetic storm event set, such as the angle of the maximum surface wind. Our approach is also quite different from another recently developed wind reconstruction model that uses a wind profile model and a numerical TC boundary layer model (Done et al. 2020). Done et al.'s (2020) approach is more physically based in that it considers physics of the TC boundary layers and does not rely so heavily on data. It is also a better tool for estimating TC winds over land because it directly accounts for terrain effects. However, our model is computationally more efficient by omitting the details in the boundary layer physics. In future work, it would be of interest to compare our results to Done et al.'s (2020) and to take into account the terrain effects for predicting surface winds over land.

There are a number of other aspects worthy of further exploration in this work. First, the environment variables used for model development here are monthly averaged which are less informative than the higher temporal resolution data used in operational forecasts. However, no obvious performance improvement was achieved in this study by switching to higher temporal resolution environment fields contained in datasets such as Statistical Hurricane Intensity Prediction Scheme (SHIPS; DeMaria et al. 2005) reanalysis database (not shown). To assess completely how the temporal resolution of environment variables impacts model performance, more detailed experiments are necessary. Second, the HWIND dataset is relatively small and only available for Atlantic hurricanes. ML algorithms, however, place high demands on the training data. Larger datasets might result in better performance. QuickSCAT satellite-estimated wind field data (Draper and Long 2002, 2004), synthetic aperture radar (SAR) wind speed data (e.g., Mouche et al. 2019; Combot et al. 2020), and the multiplatform tropical cyclone surface wind analysis (MTCSWA; Knaff et al. 2011) may be a good option because they are global data and have high temporal frequency. Last, the wind fields contained in HWIND are smoothly interpolated, meaning that they are a truncated version of the true wind. Thus, to capture higher wavenumber asymmetries, realistic wind field simulations from higher-resolution global climate simulations or regional prediction systems (Knutson et al. 2013; Roberts et al. 2020) might be useful.

Acknowledgments. C.-Y. Lee thanks the support from Vetlesen foundation to Lamont-Doherty Earth Observatory. D. R. Chavas is supported by NSF Grants 1826161 and 1648681. We thank Dr. Kun Gao for his valuable comments during the GFDL internal review. We would also like to express our appreciation to Dr. John Knaff and two anonymous reviewers for their constructive comments during the peer-review process. The authors would also like to thank Dr. Adam Sobel and Columbia University Initiative on Extreme Weather and Climate for their support.

Data availability statement. ERA-Interim data are available at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim. National Hurricane Center best-track data are available at https://www.nhc.noaa.gov/data/#hurdat. HWIND data are available at https://www.rms.com/event-response/hwind.

REFERENCES

Chang, D., S. Amin, and K. Emanuel, 2020: Modeling and parameter estimation of hurricane wind fields with asymmetry. *J. Appl. Meteor. Climatol.*, **59**, 687–705, https://doi.org/10.1175/JAMC-D-19-0126.1.

Chavas, D. R., and N. Lin, 2016: A model for the complete radial structure of the tropical cyclone wind field. Part II: Wind field

variability. *J. Atmos. Sci.*, **73**, 3093–3113, https://doi.org/10.1175/JAS-D-15-0185.1.

——, and J. A. Knaff, 2022: A simple model for predicting the tropical cyclone radius of maximum wind from outer size. *Wea. Forecasting*, https://doi.org/10.1175/WAF-D-21-0103.1, in press.

——, N. Lin, and K. Emanuel, 2015: A model for the complete radial structure of the tropical cyclone wind field. Part I: Comparison with observed structure. *J. Atmos. Sci.*, **72**, 3647–3662, https://doi.org/10.1175/JAS-D-15-0014.1.

Chen, T., and C. Guestrin, 2016: XGBoost: A scalable tree boosting system. *KDD '16: Proc. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, San Francisco, CA, ACM, 785–794, https://doi.org/10.1145/2939672.2939785.

Combot, C., A. Mouche, J. Knaff, Y. Zhao, Y. Zhao, L. Vinour, Y. Quilfen, and B. Chapron, 2020: Extensive high-resolution synthetic aperture radar (SAR) data analysis of tropical cyclones: Comparisons with SFMR flights and best track. *Mon. Wea. Rev.*, **148**, 4545–4563, https://doi.org/10.1175/MWR-D-20-0005.1.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

DeMaria, M., M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan, 2005: Further improvements to the Statistical Hurricane Intensity Prediction Scheme (SHIPS). *Wea. Forecasting*, **20**, 531–543, https://doi.org/10.1175/WAF862.1.

DiNapoli, S. M., M. A. Bourassa, and M. D. Powell, 2012: Uncertainty and intercalibration analysis of H*wind. *J. Atmos. Oceanic Technol.*, **29**, 822–833, https://doi.org/10.1175/JTECH-D-11-00165.1.

Done, J. M., M. Ge, G. J. Holland, I. Dima-West, S. Phibbs, G. R. Saville, and Y. Wang, 2020: Modelling global tropical cyclone wind footprints. *Nat. Hazards Earth Syst. Sci.*, **20**, 567–580, https://doi.org/10.5194/nhess-20-567-2020.

Donelan, M. A., 2004: On the limiting aerodynamic roughness of the ocean in very strong winds. *Geophys. Res. Lett.*, **31**, L18306, https://doi.org/10.1029/2004GL019460.

Draper, D. W., and D. G. Long, 2002: An assessment of SeaWinds on QuikSCAT wind retrieval. *J. Geophys. Res.*, **107**, 3212, https://doi.org/10.1029/2002JC001330.

——, and ——, 2004: Simultaneous wind and rain retrieval using SeaWinds data. *IEEE Trans. Geosci. Remote Sens.*, **42**, 1411–1423, https://doi.org/10.1109/TGRS.2004.830169.

Emanuel, K., 2004: Tropical cyclone energetics and structure. *Atmospheric Turbulence and Mesoscale Meteorology*, Cambridge University Press, 165–192, https://doi.org/10.1017/CBO9780511735035.010.

——, and R. Rotunno, 2011: Self-stratification of tropical cyclone outflow. Part I: Implications for storm structure. *J. Atmos. Sci.*, **68**, 2236–2249, https://doi.org/10.1175/JAS-D-10-05024.1.

——, S. Ravela, E. Vivant, and C. Risi, 2006: A statistical deterministic approach to hurricane risk assessment. *Bull. Amer. Meteor. Soc.*, **87**, 299–314, https://doi.org/10.1175/BAMS-87-3-299.

Friedman, J. H., 2001: Greedy function approximation: A gradient boosting machine. *Ann. Stat.*, **29**, 1189–1232, https://doi.org/10.1214/aos/1013203451.

Geiger, T., K. Frieler, and A. Levermann, 2016: High-income does not protect against hurricane losses. *Environ. Res. Lett.*, **11**, 084012, https://doi.org/10.1088/1748-9326/11/8/084012.

Hall, T. M., and S. Jewson, 2007: Statistical modelling of North Atlantic tropical cyclone tracks. *Tellus*, **59A**, 5970529, https://doi.org/10.3402/tellusa.v59i4.15017.

Holland, G. J., 1980: An analytic model of the wind and pressure profiles in hurricanes. *Mon. Wea. Rev.*, **108**, 1212–1218, https://doi.org/10.1175/1520-0493(1980)108<1212:AAMOTW>2.0.CO;2.

——, J. I. Belanger, and A. Fritz, 2010: A revised model for radial profiles of hurricane winds. *Mon. Wea. Rev.*, **138**, 4393–4401, https://doi.org/10.1175/2010MWR3317.1.

Jing, R., and N. Lin, 2020: An environment-dependent probabilistic tropical cyclone model. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001975, https://doi.org/10.1029/2019MS001975.

Jordan, M. R., and C. A. Clayson, 2008: A new approach to using wind speed for prediction of tropical cyclone generated storm surge. *Geophys. Res. Lett.*, **35**, L13802, https://doi.org/10.1029/2008GL033564.

Kim, S., H. Kim, J. Lee, S. Yoon, S. E. Kahou, K. Kashinath, and M. Prabhat, 2019: Deep-hurricane-tracker: Tracking and forecasting extreme climate events. *2019 IEEE Winter Conf. on Applications of Computer Vision (WACV)*, Waikoloa, HI, Institute of Electrical and Electronics Engineers, 1761–1769, https://doi.org/10.1109/WACV.2019.00192.

Klausmann, A., 2014: Analysis of Hurricane Irene's wind field using the advanced research Weather Research and Forecast (WRF-ARW) Model. *J. Mar. Sci. Eng.*, **2**, 33–45, https://doi.org/10.3390/jmse2010033.

Klotz, B. W., and H. Jiang, 2016: Global composites of surface wind speeds in tropical cyclones based on a 12-year scatterometer database. *Geophys. Res. Lett.*, **43**, 10 480–10 488, https://doi.org/10.1002/2016GL071066.

——, and ——, 2017: Examination of surface wind asymmetries in tropical cyclones. Part I: General structure and wind shear impacts. *Mon. Wea. Rev.*, **145**, 3989–4009, https://doi.org/10.1175/MWR-D-17-0019.1.

Knaff, J. A., and D. R. Chavas, 2021: Efforts to estimate the radius of maximum winds in tropical cyclones. *Fourth Special Symp. on Tropical Meteorology and Tropical Cyclones*, Online, Amer. Meteor. Soc., 12.3, https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/379349.

——, M. DeMaria, D. A. Molenar, C. R. Sampson, and M. G. Seybold, 2011: An automated, objective, multiple-satellite-platform tropical cyclone surface wind analysis. *J. Appl. Meteor. Climatol.*, **50**, 2149–2166, https://doi.org/10.1175/2011JAMC2673.1.

——, S. P. Longmore, R. T. DeMaria, and D. A. Molenar, 2015: Improved tropical-cyclone flight-level wind estimates using routine infrared satellite reconnaissance. *J. Appl. Meteor. Climatol.*, **54**, 463–478, https://doi.org/10.1175/JAMC-D-14-0112.1.

——, C. R. Sampson, and G. Chirokova, 2017: A global statistical–dynamical tropical cyclone wind radii forecast Scheme. *Wea. Forecasting*, **32**, 629–644, https://doi.org/10.1175/WAF-D-16-0168.1.

Knutson, T. R., and Coauthors, 2013: Dynamical downscaling projections of twenty-first-century Atlantic hurricane activity: CMIP3 and CMIP5 model-based scenarios. *J. Climate*, **26**, 6591–6617, https://doi.org/10.1175/JCLI-D-12-00539.1.

Komaromi, W. A., and J. D. Doyle, 2018: On the dynamics of tropical cyclone and trough interactions. *J. Atmos. Sci.*, **75**, 2687–2709, https://doi.org/10.1175/JAS-D-17-0272.1.

Landsea, C. W., and J. L. Franklin, 2013: Atlantic hurricane database uncertainty and presentation of a new database format.

Mon. Wea. Rev., **141**, 3576–3592, https://doi.org/10.1175/MWR-D-12-00254.1.

Lee, C.-Y., and S. S. Chen, 2012: Symmetric and asymmetric structures of hurricane boundary layer in coupled atmosphere–wave–ocean models and observations. *J. Atmos. Sci.*, **69**, 3576–3594, https://doi.org/10.1175/JAS-D-12-046.1.

——, and ——, 2014: Stable boundary layer and its impact on tropical cyclone structure in a coupled atmosphere–ocean model. *Mon. Wea. Rev.*, **142**, 1927–1944, https://doi.org/10.1175/MWR-D-13-00122.1.

——, M. K. Tippett, S. J. Camargo, and A. H. Sobel, 2015: Probabilistic multiple linear regression modeling for tropical cyclone intensity. *Mon. Wea. Rev.*, **143**, 933–954, https://doi.org/10.1175/MWR-D-14-00171.1.

——, ——, A. H. Sobel, and S. J. Camargo, 2016: Rapid intensification and the bimodal distribution of tropical cyclone intensity. *Nat. Commun.*, **7**, 10625, https://doi.org/10.1038/ncomms10625.

——, ——, ——, and ——, 2018: An environmentally forced tropical cyclone hazard model. *J. Adv. Model. Earth Syst.*, **10**, 223–241, https://doi.org/10.1002/2017MS001186.

——, S. J. Camargo, A. H. Sobel, and M. K. Tippett, 2020: Statistical–dynamical downscaling projections of tropical cyclone activity in a warming climate: Two diverging genesis scenarios. *J. Climate*, **33**, 4815–4834, https://doi.org/10.1175/JCLI-D-19-0452.1.

Lin, N., and D. Chavas, 2012: On hurricane parametric wind and applications in storm surge modeling. *J. Geophys. Res.*, **117**, D09120, https://doi.org/10.1029/2011JD017126.

Loridan, T., S. Khare, E. Scherer, M. Dixon, and E. Bellone, 2015: Parametric modeling of transitioning cyclone wind fields for risk assessment studies in the western North Pacific. *J. Appl. Meteor. Climatol.*, **54**, 624–642, https://doi.org/10.1175/JAMC-D-14-0095.1.

——, R. P. Crompton, and E. Dubossarsky, 2017: A machine learning approach to modeling tropical cyclone wind field uncertainty. *Mon. Wea. Rev.*, **145**, 3203–3221, https://doi.org/10.1175/MWR-D-16-0429.1.

Mei, W., C. Pasquero, and F. Primeau, 2012: The effect of translation speed upon the intensity of tropical cyclones over the tropical ocean. *Geophys. Res. Lett.*, **39**, L07801, https://doi.org/10.1029/2011GL050765.

Mouche, A., B. Chapron, J. Knaff, Y. Zhao, B. Zhang, and C. Combot, 2019: Copolarized and cross-polarized SAR measurements for high-resolution description of major hurricane wind Structures: Application to Irma category 5 hurricane. *J. Geophys. Res. Oceans*, **124**, 3905–3922, https://doi.org/10.1029/2019JC015056.

Mueller, K. J., M. DeMaria, J. A. Knaff, J. P. Kossin, and T. H. V. Haar, 2006: Objective estimation of tropical cyclone wind structure from infrared satellite data. *Wea. Forecasting*, **21**, 990–1005, https://doi.org/10.1175/WAF955.1.

Needham, H. F., and B. D. Keim, 2014: Correlating storm surge heights with tropical cyclone winds at and before landfall. *Earth Interact.*, **18**, https://doi.org/10.1175/2013EI000527.1.

Olfateh, M., D. P. Callaghan, P. Nielsen, and T. E. Baldock, 2017: Tropical cyclone wind field asymmetry-development and evaluation of a new parametric model. *J. Geophys. Res. Oceans*, **122**, 458–469, https://doi.org/10.1002/2016JC012237.

Peduzzi, P., B. Chatenoux, H. Dao, A. D. Bono, C. Herold, J. Kossin, F. Mouton, and O. Nordbeck, 2012: Global trends in tropical cyclone risk. *Nat. Climate Change*, **2**, 289–294, https://doi.org/10.1038/nclimate1410.

Powell, M. D., S. H. Houston, L. R. Amat, and N. Morisseau-Leroy, 1998: The HRD real-time hurricane wind analysis system. *J. Wind Eng. Ind. Aerodyn.*, **77–78**, 53–64, https://doi.org/10.1016/S0167-6105(98)00131-7.

Quinlan, J. R., 1986: Induction of decision trees. *Mach. Learn.*, **1**, 81–106, https://doi.org/10.1023/A:1022643204877.

Racah, E., C. Beckham, T. Maharaj, S. Kahou, Prabhat, and C. Pal, 2017: Extreme weather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. *Proc. 31st Int. Conf. on Neural Information Processing Systems*, Long Beach, CA, ACM, 3405–3416, https://dl.acm.org/doi/10.5555/3294996.3295099.

Roberts, M. J., and Coauthors, 2020: Impact of model resolution on tropical cyclone simulation using the HighResMIP-PRIMAVERA multimodel ensemble. *J. Climate*, **33**, 2557–2583, https://doi.org/10.1175/JCLI-D-19-0639.1.

Shapiro, L. J., 1983: The asymmetric boundary layer flow under a translating hurricane. *J. Atmos. Sci.*, **40**, 1984–1998, https://doi.org/10.1175/1520-0469(1983)040<1984:TABLFU>2.0.CO;2.

Sun, Z., B. Zhang, J. A. Zhang, and W. Perrie, 2019: Examination of surface wind asymmetry in tropical cyclones over the northwest Pacific Ocean using SMAP observations. *Remote Sens.*, **11**, 2604, https://doi.org/10.3390/rs11222604.

Ueno, M., and K. Bessho, 2011: A statistical analysis of near-core surface wind asymmetries in typhoons obtained from QuikSCAT Data. *J. Meteor. Soc. Japan. Ser. II*, **89**, 225–241, https://doi.org/10.2151/jmsj.2011-304.

Uhlhorn, E. W., B. W. Klotz, T. Vukicevic, P. D. Reasor, and R. F. Rogers, 2014: Observed hurricane wind speed asymmetries and relationships to motion and environmental shear. *Mon. Wea. Rev.*, **142**, 1290–1311, https://doi.org/10.1175/MWR-D-13-00249.1.

Willoughby, H. E., and M. E. Rahn, 2004: Parametric representation of the primary hurricane vortex. Part I: Observations and evaluation of the Holland (1980) model. *Mon. Wea. Rev.*, **132**, 3033–3048, https://doi.org/10.1175/MWR2831.1.

——, R. W. R. Darling, and M. E. Rahn, 2006: Parametric representation of the primary hurricane vortex. Part II: A new family of sectionally continuous profiles. *Mon. Wea. Rev.*, **134**, 1102–1120, https://doi.org/10.1175/MWR3106.1.

Wong, M. L. M., and J. C. L. Chan, 2007: Modeling the effects of land–sea roughness contrast on tropical cyclone winds. *J. Atmos. Sci.*, **64**, 3249–3264, https://doi.org/10.1175/JAS4027.1.

Wood, V. T., L. W. White, H. E. Willoughby, and D. P. Jorgensen, 2013: A new parametric tropical cyclone tangential wind profile model. *Mon. Wea. Rev.*, **141**, 1884–1909, https://doi.org/10.1175/MWR-D-12-00115.1.

Yang, Q., C.-Y. Lee, and M. K. Tippett, 2020: A long short-term memory model for global rapid intensification prediction. *Wea. Forecasting*, **35**, 1203–1220, https://doi.org/10.1175/WAF-D-19-0199.1.