

Machine learning based hurricane wind reconstruction

Qidong Yang*

Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY

Courant Institute of Mathematical Sciences, New York University, New York, NY

Chia-Ying Lee

Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY

Michael K. Tippett

Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY

Daniel R. Chavas

Department of Earth, Atmospheric and Planetary Sciences, Purdue University, West Lafayette, IN

Thomas R. Knutson

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, NJ

*Corresponding author: Qidong Yang, qy2216@columbia.edu

ABSTRACT

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

31 **1. Introduction**

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

48 Popular methods for generating the azimuthally symmetric field include empirical models such
49 as Holland et al. (2010) and Willoughby et al. (2006), and theory-based ones, such as Chavas
50 et al. (2015). These three models, referred as Holland10, Willoughby06, and Chavas15 hereafter,
51 all provide a radial profile of the symmetric TC wind that sharply increases from storm center to
52 the radius of maximum wind speed (R_{max}), and then gradually decreases outward to merge with
53 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, 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 down-wind 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 down-shear to left-of-shear with increasing shear magnitude, which is consistent with the findings of Klotz and Jiang (2017), who used global rain-corrected scat-

78 terometer wind data. Additionally, the blocking action from an anticyclone near a TC can cause a
79 mesoscale condition analogous to the wall effect for a vortex (Olfateh et al. 2017), which impacts
80 TC wind structure to certain degree. Land-sea roughness contrast (Wong and Chan 2007), air-sea
81 interaction (Lee and Chen 2012, 2014), as well as the interactions between TCs and midlatitude
82 circulations (Komaromi and Doyle 2018) can influence TC surface wind structure as well.

83 In recent years, machine learning (ML) has made incremental progress in geophysical sciences.
84 Several recent works have applied ML methods in TC studies. Racah et al. (2017) and Kim et al.
85 (2019) utilized convolutional neural networks to detect TCs with weather model variables such
86 as integrated water vapor. Our recent work, Yang et al. (2020), used a long-short-term memory
87 model to predict TC rapid intensification with a set of storm and environment conditions as inputs
88 and gives performance comparable to operational forecasts. Loridan et al. (2017) combined super-
89 vised (quantile regression forest) and unsupervised (principal component analysis) ML methods to
90 model TC wind fields, suggesting that complex wind field asymmetries can possibly be captured.
91 The successes of these ML applications stem in part from the increasing amount of observed
92 and simulated data and the modeling capacity provided by ML algorithms, which allow for ap-
93 proximation of any nonlinear relationship. These works thus inspire us to apply advanced ML
94 methods to the problem of improving modeled wind fields by adding physical dependencies that are
95 missing in existing parametric modeling approaches. Specifically, we propose an XGBoost-based
96 reconstruction of the symmetric and asymmetric components of the TC surface wind field based
97 on environment and storm features as predictors. First proposed by Chen and Guestrin (2016),
98 XGBoost is a decision tree-based ensemble predictive algorithm. The model uses an optimization
99 method called gradient boosting (Friedman 2001) that adjusts decision trees (Quinlan 1986) to
100 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

¹⁰² achieves outstanding performance on structured datasets and can be trained very fast with parallel
¹⁰³ processing.

¹⁰⁴ The remainder of this paper is organized as follows. The observed storm, environment, and the
¹⁰⁵ HWIND data used to train and test the XGBoost models are described in section 2. Feasibility of
¹⁰⁶ the HWIND data for wind field reconstruction model development is evaluated in section 3. The
¹⁰⁷ model development details are specified in section 4. The performance of the reconstruction model
¹⁰⁸ is discussed in section 5, along with a wind field reconstruction case study. Section 6 provides
¹⁰⁹ further analysis on the relationship between wind field asymmetries and storm and environment
¹¹⁰ variables, which play a crucial role in the reconstruction model. We summarize and discuss our
¹¹¹ findings in section 7.

¹¹² 2. Data and methods

¹¹³ a. Tropical cyclones and their environment

¹¹⁴ Here we develop the wind reconstruction model using the same set of storm and environment
¹¹⁵ variables as used in Yang et al. (2020). The storm variables are derived from the National Hurricane
¹¹⁶ Center (NHC; Landsea and Franklin 2013) best-track data, and the environment variables are taken
¹¹⁷ from the monthly European Centre for Medium-Range Weather Forecasts interim reanalysis (ERA-
¹¹⁸ Interim; Dee et al. 2011) database. The monthly environment variables are then linearly interpolated
¹¹⁹ to storm time stamps. We use monthly environmental variables because our intended application is
¹²⁰ TC wind risk assessment in a changing climate using a statistical-dynamical downscaling approach
¹²¹ that combines large-scale climate information from dynamical models or reanalysis with statistical
¹²² models to produce storm-scale features (e.g., Lee et al. 2020). In this application, we need to
¹²³ generate wind fields for thousands of synthetic storms over a long period (e.g., 1951-2100), and

124 with climate forcing from many global models. Using monthly environmental variables reduces
125 the data storage burden. This also removes the need to filter modeled or observed storms from
126 environmental variables. Of course, this approach will not capture changes that are due to sub-
127 monthly environmental variability.

128 A total of 30 variables (8 storm variables and 22 environment variables) are used in this work.
129 Storm variables describe the storm status at the time of a wind field snapshot and include quantities
130 such as storm maximum wind speed, storm center location, and storm translation (Table 1).
131 Environment variables, computed from a 0.25×0.25 degree global grid, depict the large-scale
132 conditions in the neighborhood of the storm and include vertical wind shear, relative humidity,
133 outflow temperature, high-level divergence, and so on as listed in Tables 2 and 3. We represent
134 storm translation and vertical wind shear by their zonal and meridional components (i.e., SH_z and
135 SH_m for shear and MT_z and MT_m for storm motion) to avoid the inherent ambiguity of angular
136 variables which can make their use in ML models cumbersome. Some environment variables are
137 averaged over a small disk centered on the storm and others are averaged over a large annulus.
138 Disk-averaged variables are averaged over the area within 500 km from storm center while annulus-
139 averaged ones are averaged over a ring-shaped area with inner radius of 200 km and outer radius
140 of 800 km from storm center. The former averaging method is intended to capture environment
141 conditions around storm center while variables averaged over a large annulus aim to describe
142 surrounding conditions in which the storms are embedded.

143 TC intensity and storm structure are not unrelated — strong storms are more symmetric than
144 weaker ones (Klotz and Jiang 2016, 2017; Sun et al. 2019). While those aforementioned variables
145 are originally derived for TC intensity prediction (Lee et al. 2015, 2016; Yang et al. 2020), they
146 are suitable for storm structure prediction as well. For example, vertical wind shear is typically
147 a negative factor for intensification by tilting the storm vortex and transporting dry air into the

¹⁴⁸ inner core. A known impact of the above vertical wind shear process on TC structure is that the
¹⁴⁹ strongest convection focuses in the downshear-left quadrant. Uhlhorn et al. (2014) showed that
¹⁵⁰ this convective asymmetry may result in surface wind asymmetries when the shear is sufficient.
¹⁵¹ The work by Knaff et al. (2017) also illustrates a successful example of predicting storm structure
¹⁵² using intensity related variables. Besides, as suggested in Loridan et al. (2017), more complex TC
¹⁵³ environmental features which characterize the atmospheric flow, such as the relative humidity and
¹⁵⁴ upper-level divergence, should also be considered for better surface wind field prediction.

¹⁵⁵ Two additional variables, the maximum azimuthally averaged wind speed (V_{max}) and its cor-
¹⁵⁶ responding radius (R_{max}) that are often used in the parametric wind profiles are included in the
¹⁵⁷ reconstruction model development, too. In this study, we derive V_{max} and R_{max} from azimuthally
¹⁵⁸ averaged HWIND profiles and regard them as known. In practice, V_{max} can be estimated by a
¹⁵⁹ linear fit of storm maximum wind speed (S_{max}) provided by best track data, and R_{max} can be
¹⁶⁰ obtained through an empirical equation from Knaff et al. (2015), theory-based model from Chavas
¹⁶¹ and Lin (2016), or a combination of theory and data (Chavas and Knaff 2020). A recent study by
¹⁶² Knaff and Chavas (2021) notes that estimating R_{max} without direct reconnaissance measurements
¹⁶³ is challenging.

¹⁶⁴ *b. HWIND*

¹⁶⁵ The observed TC wind fields are from the Hurricane Research Division hurricane surface wind
¹⁶⁶ analysis system [HWIND - now Risk Management Solutions (RMS) HWind as per the RMS web-
¹⁶⁷ site] (e.g., Powell et al. 1998; DiNapoli et al. 2012). These wind fields are produced by interpolating
¹⁶⁸ and smoothing wind speed observations from multiple platforms including stepped-frequency mi-
¹⁶⁹ crowave radiometers, satellites, GPS dropsondes, ships, buoys, and land-based observation stations.
¹⁷⁰ We use HWIND data within radii ranging from 0 to 300 km from storm center because radii up to

¹⁷¹ 300 km are generally large enough to cover a storm's main structure. Data at normalized radius
¹⁷² (radius divided by R_{max}) are linearly interpolated. Here R_{max} denotes the radius where the HWIND
¹⁷³ azimuthally averaged wind speed maximum achieved. From 2000 to 2014¹, the dataset contains
¹⁷⁴ 1539 Atlantic hurricane surface (10 m) wind speed field snapshots from 112 hurricanes. Since the
¹⁷⁵ size of the HWIND dataset is limited and ML-based models have a high demand for training data,
¹⁷⁶ a higher than usual proportion of the dataset is used for model training: 93.75% of the HWIND
¹⁷⁷ storms (105 storms, 1410 wind field snapshots) are randomly chosen and used as training set. The
¹⁷⁸ remaining 6.25% of the storms (7 storms, 129 wind field snapshots) are saved as an independent
¹⁷⁹ testing set for model performance evaluation. In terms of the number of wind field snapshots, over
¹⁸⁰ 8% of them are used for performance evaluation.

¹⁸¹ An additional experiment with a second kind of training-testing set split (by wind field snapshots)
¹⁸² was conducted (not shown). In this alternative approach, we randomly chose 92% of snapshots
¹⁸³ as the training set and the rest as the testing set. The evaluation results from this splitting method
¹⁸⁴ were even more promising. However, wind fields from the same storm are correlated in time
¹⁸⁵ because similar surface winds affecting storm-scale and environmental conditions will continue
¹⁸⁶ over multiple snapshots and result in similar snapshots for the same storm. Splitting that data
¹⁸⁷ by snapshot may unfairly put testing set answers into the training set, i.e., the testing data are
¹⁸⁸ not independent of the training data. Consequently, the alternative splitting approach leads to an
¹⁸⁹ over-fitting on training data, and the resultant reconstruction model will have weak generalization
¹⁹⁰ ability on truly independent data. Therefore, we choose splitting by storms rather than splitting by
¹⁹¹ wind field snapshots.

¹Data after 2014 are not publicly available

192 *c. Skill scores*

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

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

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

200 **3. A feasibility test of wind fields in HWIND**

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

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

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

229 4. Model development

230 a. Reconstruction model design

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

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

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

$$\text{azimuthal average} - \text{parametric profile} = \text{residual profile}$$

$$\approx \sum_{n=1}^{\infty} A_n J_0(\lambda_n \frac{r_p}{R_{max}}) \mathbf{1}_{(r_p \leq R_{max})} + \sum_{n=1}^{\infty} B_n J_0(\lambda_n \frac{R_u - r_p}{R_u - R_{max}}) \mathbf{1}_{(r_p \geq R_{max})}, \quad (3)$$

238 where the $J_0(\cdot)$ is the zero-order Bessel function and λ_n is the n th positive root solving $J_0(\cdot) = 0$; $\mathbf{1}$
 239 is the indicator function which is one when its argument is true and zero otherwise. A_n and B_n are
 240 the coefficients associated with the inner and outer series of Bessel functions. The coefficients are
 241 obtained by area-weighted least squares fitting to the HWIND data. The functional dependence of
 242 the outer series on r_p means that the Bessel function is reversed and shifted so that it is zero at
 243 $r_p = R_{max}$. The reason we use separate series of Bessel functions to fit residual profile before and
 244 after R_{max} is to ensure that the fitted residual profile is zero at R_{max} . Thus, by construction, the
 245 Bessel function corrected parametric profile matches the targeted azimuthal average at R_{max} .

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

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

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

250 as:

$$H_a(m, n) = N_{m,n} J_m(\lambda_{m,n} r) \cos(m\theta), \quad (5)$$

$$H_b(m, n) = N_{m,n} J_m(\lambda_{m,n} r) \sin(m\theta),$$

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

256 *b. Symmetric field approximation*

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

267 These wind profile models take as inputs the maximum wind speed (V_{max}) and its corresponding
268 radius (R_{max}) to generate radial profiles of azimuthally averaged wind speed from storm center to
269 outer radii. Additional profile shape adjusting parameters (e.g., peripheral wind observation and
270 its associated radius) can also be provided to improve the fit. In this work, all three models take in

271 V_{max} and R_{max} ; and the Willoughby06 and Chavas15 models additionally depend on storm center
272 latitude and Coriolis parameter, respectively. In Holland10, external pressure, central pressure,
273 peripheral wind radius, and peripheral wind observation are set to 1005 hPa, 950 hPa, 300 km, and
274 17 m/s, respectively, according to the profile baseline in Holland et al. (2010). As for Chavas15,
275 the drag coefficient is obtained following Donelan (2004); the ratio of the exchange coefficients of
276 enthalpy and momentum is computed by a quadratic fit from Chavas et al. (2015); an empirical
277 adjustment suggested in Chavas et al. (2015) is applied to profile eye; and radiative-subsidence rate
278 is fixed to 2×10^{-3} m/s. Chavas15 can also be simplified by setting Coriolis parameter constant to
279 5×10^{-5} s⁻¹. It changes resultant parametric profiles, but does not affect the finally approximated
280 symmetric fields after Bessel function correction.

281 Profile examples from these three models are illustrated in Fig. 3. By construction, all three match
282 the observed wind speed at R_{max} . The Holland10 and Chavas15 models have a better fit inside
283 R_{max} than the Willoughby06 model, but they overestimate wind speed outside R_{max} . The distance
284 outside R_{max} makes up roughly two thirds of the whole modeling radius range. Consequently, in a
285 mean square error sense (not area-weighted), the Willoughby06 model performs relatively better
286 than the other two in this example. Figure 4 further shows the root mean square error (RMSE)
287 of the three wind profile models over the HWIND dataset. Willoughby06 and Chavas15 perform
288 similarly for 1–3 R_{max} ; Chavas15 is better for 0.5–1 R_{max} and Willoughby06 is better beyond 3
289 R_{max} . We note that TC wind damage happens disproportionately due to the strongest winds which
290 occur near R_{max} . The Chavas15 model outperforms both the Willoughby06 and Holland10 models
291 in terms of RMSE from 0 to 3 R_{max} . Nevertheless, as mentioned previously, two series of Bessel
292 functions serve to further improve the representation of the symmetric wind component. Thus
293 what we require is only a profile model that gives a reasonable starting point, and all three models

294 explored here are adequate for this purpose. In the rest of the paper, we will use the Willoughby06
295 model as an example.

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

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

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

309 We also tried to use only Bessel functions to approximate wind profiles (with no parametric
310 profile involved). In Fig. 4, the yellow curve represents RMSE from profiles fitted by a single
311 Bessel function series (eight terms) for the full radial range and the cyan curve is from profiles
312 fitted by two series of Bessel functions (four terms for $r_p \leq R_{max}$ and four terms for $r_p \geq R_{max}$).
313 The yellow curve shows rather large RMSE. In contrast, the cyan curve is quite close to the magenta

314 curve. However, it does not imply that parametric profiles are not necessary. In practice, the Bessel
 315 function coefficients from symmetric models need to be predicted by XGBoost. The benefit of
 316 including the parametric profiles is that the wind reconstruction model is always given a reasonable
 317 profile to start with. Even though XGBoost gives poor Bessel function coefficient predictions, the
 318 final result will not be totally unreasonable. Therefore, a parametric profile combined with the two
 319 Bessel series is still the optimal choice.

320 *c. Asymmetric field approximation*

321 Next, we consider the approximation of the asymmetric field by truncating the eigenfunction
 322 series in Eq. 4. As M and N increase, higher frequency asymmetries are added, and truncation error
 323 decreases (Fig. 5). The truncation error decreases slowly after $M = 3$ and $N = 4$; the asymmetric
 324 modes that appear most useful in approximating the asymmetric fields are $\text{Asymmetry}_{m,n}$ with
 325 $m \in \{1, 2, 3\}$, $n \in \{1, 2, 3, 4\}$.

326 The asymmetric field approximation then can be defined as:

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

327 Substitution using Eq. 4 and 5 shows that

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

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

$$M_{m,n} = \sqrt{a_{m,n}^2 + b_{m,n}^2}, \quad P_{m,n} = \frac{1}{m} \text{atan2} \left(\frac{b_{m,n}}{a_{m,n}} \right). \quad (9)$$

329 In particular, $\text{atan2}(\cdot)$ is the four-quadrant inverse tangent. Figure 6 shows that the three asym-
 330 metries with largest magnitude in the HWIND dataset are $\text{Asymmetry}_{1,1}$, $\text{Asymmetry}_{1,2}$, and
 331 $\text{Asymmetry}_{2,1}$.

332 *d. The complete reconstruction model*

333 The reconstructed wind is the sum of the symmetric and asymmetric wind approximations. There
334 are two sets of unknowns: Bessel function coefficients (A_n and B_n) in the symmetric approximation
335 and eigenfunction coefficients ($a_{m,n}$ and $b_{m,n}$) in the asymmetric approximation. Together there are
336 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
337 $n = 1, 2, 3, 4$. Each coefficient is predicted separately by an XGBoost model trained with the input
338 variables listed in Tables 1, 2 and 3. The hyper-parameters of XGBoost models (e.g., learning rate
339 and max depth of trees) are tuned on the training set using 5-fold cross validation. The chosen
340 hyper-parameters are listed in Table 4. Then the XGBoost models are trained on the complete
341 training set. The performance of the reconstruction model is lastly evaluated on the independent
342 testing set.

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

347 **5. Performance evaluation**

348 *a. Coefficient prediction*

349 The MSESS values for each predicted coefficient are listed in Table 5. The coefficients associated
350 with the symmetric component (A_n and B_n) are predicted more accurately than those associated
351 with the asymmetric component ($a_{m,n}$ and $b_{m,n}$). Lower wavenumber asymmetries (smaller m and
352 n) are more accurately predicted than higher wavenumber ones (bigger m and n).

353 *b. Case studies*

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

373 *c. Reconstructed surface wind*

374 We compute the MSESS for the reconstructions of the symmetric, asymmetric, and full wind
375 fields (Fig. 8). Climatologies used for the reference forecast in Equation 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, up-wind-right and down-wind-left in motion coordinates, and down-shear-left and up-shear-right in shear coordinates. These patterns vary when different testing sets are used. Figs. 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

399 turbulent nature of moist convection. It is, however, the most skillful region for the XGBoost-based
400 wind reconstruction model.

401 **6. Model interpretation**

402 Now we analyze the dependence of the symmetric and asymmetric winds fields on predictors to
403 gain insights about the physical controls of TC surface wind structures. To do so, we developed a
404 sequence of diagnostic XGBoost models, which are developed the same way as the reconstruction
405 XGBoost models, but with predictors added one at a time ordered by how much each predictor
406 reduced model's loss function (i.e. the difference between prediction and observation), going
407 from largest reduction to least reduction. For the symmetric field, we focused on the first terms
408 of Bessel function series, $A_1 J_0(\lambda_n r)$ and $B_1 J_0(\lambda_n r)$, which correct the parametric profile. For
409 the asymmetric field, we focused on three asymmetric modes with relatively large magnitudes:
410 Asymmetry_{1,1}, Asymmetry_{1,2}, and Asymmetry_{2,1}.

411 *a. Symmetric field*

412 As input variables are added, the performance of the diagnostic models improves and eventually
413 stabilizes. We define essential variables (marked in red in Figs. 9 and 10) as those that provide
414 the bulk of the model improvement. Note there are variables in black between two variables in
415 red. It suggests that the variables in black are related with their preceding variables in red (i.e.
416 already recognized as essential), hence they appear relatively unessential. For example, V_{max}
417 (black) locates between R_{max} (red) and $\overline{T_{200_d}}$ (red) in Fig. 9b. It implies V_{max} is correlated with
418 R_{max} . As a result, they achieve similar loss function reduction and are ordered consecutively. It
419 is not hard to imagine that the performance of diagnostic XGBoost with only R_{max} included is
420 close to the performance with both V_{max} and R_{max} included due to these two variables' correlation.

421 Although outflow temperature ($\overline{T_{200_d}}$) does not achieve loss function reduction as much as V_{max} , it
 422 brings new information into diagnostic XGBoost. Thus, when it is included, it increases diagnostic
 423 XGBoost performance more than V_{max} . Therefore, V_{max} is illustrated as unessential relative to
 424 R_{max} and $\overline{T_{200_d}}$.

425 In Fig. 9a, for the inner eye coefficient A_1 , nearly all the diagnostic model improvement comes
 426 from the first variable V_{max} , hence V_{max} is the only essential variable for predicting A_1 and provides
 427 most of the information needed to correct the Willoughby06 inner eyewall profile. Further analysis
 428 also shows that there exists a rather strong positive linear relationship between V_{max} and A_1 , which
 429 reflects the key role that V_{max} plays in setting inner-eyewall wind structure. In the case of outer eye
 430 coefficient B_1 (Fig. 9b), the essential variables are: R_{max} , $\overline{T_{200_d}}$, $\frac{dS}{dt}$, $\overline{RH_{h_a}}$, and lat . This collection
 431 is much more diverse and includes information from various sources such as storm intensity change
 432 ($\frac{dS}{dt}$), inner core size (R_{max}) and location (lat), and outflow temperature ($\overline{T_{200_d}}$).

433 b. Asymmetry magnitude

434 Next, we use coefficients predicted by diagnostic models to construct asymmetries. In Fig. 10,
 435 there is a general reduction of R^2 with increasing m and n , which agrees with the conclusion
 436 from Table 5 that the predictability of higher wavenumber asymmetries is lower than that of lower
 437 wavenumber ones.

438 Figs. 10a–c demonstrate how the magnitudes of the dominant asymmetries ($M_{1,1}$, $M_{1,2}$, and $M_{2,1}$)
 439 relate with input variables. The R^2 curve of $M_{1,1}$ roughly stops increasing after $\overline{T_{200_d}}$ suggesting
 440 that the essential variables for predicting $M_{1,1}$ are: MT , lat , MT_z , $\overline{RH_{l_a}}$, and $\overline{T_{200_d}}$. In particular,
 441 storm translation variables (MT and MT_z) contribute over half of the performance improvement
 442 by all essential variables. This behavior is not surprising since storm translation is a determining
 443 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.

457 c. Asymmetry phase

458 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.

Figs. 10d–f show the R^2 analysis on asymmetry earth phase. The R^2 curve of $P_{1,1}$ levels off at the 3rd 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

490 direction relative to translation direction (SH_{θ}^{MT}). In particular, MT_{θ} and SH_{θ} are defined in earth
491 coordinates ranging from -180 to 180 degrees such that 0, 90, 180 (or -180), and -90 degrees
492 represent east, north, west, and south, respectively. Note that we also shifted $P_{1,1}$ into motion and
493 shear coordinates to facilitate phase-predictor relationship illustration.

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

504 In terms of translation direction (MT_{θ}) in Fig. 11d, D1 denotes cases from the lowest quartile of
505 values of MT_{θ} in which storms move towards east-northeast (MT_{θ} mean of 31 degrees), and D4
506 denotes cases from the highest quartile of values of MT_{θ} where storms move towards west-northwest
507 (MT_{θ} mean of 167 degrees). Even though from D1 to D4 storm motion heading changes from
508 east to west counter-clockwise, $P_{1,1}$ is always peaked at right-of-motion. Storm moving direction
509 is again shown to be the dominant variable controlling $P_{1,1}$ in motion coordinates. In contrast,
510 storm-moving earth-relative direction is not a dominant variable controlling $P_{1,1}$. Figure 11(e)
511 shows that $P_{1,1}$ rotates from left-of-shear to down-shear with shear heading varying from south
512 (-92 degrees) to east-northeast (24 degrees) counter-clockwise. Again, this pattern is proven to be
513 actually driven by storm motion direction by Fig. 11f, which demonstrates that $P_{1,1}$ concentrates at

514 right-of-motion regardless of shear direction. That is, the shear direction relative to the translation
515 direction covers a wide range of different angles across the percentiles (see red arrows), yet the
516 wind field asymmetry $P_{1,1}$ (in motion coordinates) is consistently to the right of motion regardless
517 of the shear direction relative to translation direction.

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

524 7. Discussions and conclusions

525 A downscaling hurricane wind reconstruction method using XGBoost is developed and evaluated
526 in this study with HWIND data over the period 2000 to 2014. This method divides the wind field into
527 symmetric and asymmetric components and models them separately. The symmetric component
528 is approximated by a parametric wind profile along with two series of Bessel functions. The
529 asymmetric component is expanded in a Laplacian eigenfunction series. Then XGBoost models
530 are trained with storm and environment features as inputs to predict the coefficients associated with
531 the Bessel functions and eigenfunctions.

532 We experimented with three existing parametric wind profile models: Holland10, Willoughby06,
533 and Chavas15. The root mean square error (RMSE) between Willoughby06 and azimuthally
534 averaged HWIND profiles, with RMSE maximum around 4, was the lowest of the three parametric
535 wind profile models in the range of 1–8 R_{max} . Chavas15 is the better performing one with RMSE
536 maximum less than 4 over the range of 0.5–3 R_{max} , where winds are relatively strong. Since the

537 Bessel function series were able to correct all three parametric models sufficiently well, the final
538 performance did not depend on which model was used.

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

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

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

575 There are a number of other aspects worthy of further exploration in this work. First, the
576 environment variables used for model development here are monthly averaged which are less
577 informative than the higher temporal resolution data used in operational forecasts. However,
578 no obvious performance improvement was achieved in this study by switching to higher temporal
579 resolution environment fields contained in datasets such as Statistical Hurricane Intensity Prediction
580 Scheme (SHIPS; DeMaria et al. 2005) reanalysis database (not shown). To assess completely
581 how the temporal resolution of environment variables impacts model performance, more detailed
582 experiments are necessary. Second, the HWIND dataset is relatively small and only available for
583 Atlantic hurricanes. ML algorithms, however, place high demands on the training data. Larger
584 datasets might result in better performance. QuickSCAT satellite-estimated wind field data (Draper

585 and Long 2002, 2004), synthetic aperture radar (SAR) wind speed data (e.g., Mouche et al. 2019;
586 Combot et al. 2020), and the multi-platform tropical cyclone surface wind analysis ((MTCSWA)
587 Knaff et al. 2011) may be a good option because they are global data and have high temporal
588 frequency. Lastly, the wind fields contained in HWIND are smoothly interpolated, meaning that
589 they are a truncated version of the true wind. Thus, to capture higher wavenumber asymmetries,
590 realistic wind field simulations from higher resolution global climate simulations or regional
591 prediction systems (Knutson et al. 2013; Roberts et al. 2020) might be useful.

592 *Acknowledgments.* C.-Y. Lee thanks the support from Vetlesen foundation to Lamont-Doherty
593 Earth Observatory. D. R. Chavas is supported by NSF grants 1826161 and 1648681. We thank
594 Dr. Kun Gao for his valuable comments during the GFDL internal review. We would also like to
595 express our appreciation to Dr. John Knaff and two anonymous reviewers for their constructive
596 comments during the peer-review process.

597 *Data availability statement.*

598 References

- 599 Chang, D., S. Amin, and K. Emanuel, 2020: Modeling and Parameter Estimation of Hur-
600 ricane Wind Fields with Asymmetry. *J. Appl. Meteor. Climatol.*, **59** (4), 687–705, doi:
601 10.1175/jamc-d-19-0126.1, URL <https://doi.org/10.1175%2Fjamc-d-19-0126.1>.
- 602 Chavas, D. R., and J. A. Knaff, 2020: A simple physical-statistical model to predict the hurricane
603 radius of maximum wind from outer size. *AGU 2020 Fall Meeting*.
- 604 Chavas, D. R., and N. Lin, 2016: A Model for the Complete Radial Structure of the Tropical
605 Cyclone Wind Field. Part II: Wind Field Variability. *Journal of the Atmospheric Sciences*, **73** (8),
606 3093–3113, doi:10.1175/jas-d-15-0185.1, URL <https://doi.org/10.1175%2Fjas-d-15-0185.1>.

- 607 Chavas, D. R., N. Lin, and K. Emanuel, 2015: A Model for the Complete Radial Structure of the
608 Tropical Cyclone Wind Field. Part I: Comparison with Observed Structure. *J. Atmos. Sci.*, **72** (9),
609 3647–3662, doi:10.1175/jas-d-15-0014.1, URL <https://doi.org/10.1175%2Fjas-d-15-0014.1>.
- 610 Chen, T., and C. Guestrin, 2016: XGBoost: A Scalable Tree Boosting System. *Proceedings of*
611 *the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*,
612 ACM, doi:10.1145/2939672.2939785, URL <https://doi.org/10.1145%2F2939672.2939785>.
- 613 Combot, C., A. Mouche, J. Knaff, Y. Zhao, Y. Zhao, L. Vinour, Y. Quilfen, and B. Chapron, 2020:
614 Extensive High-Resolution Synthetic Aperture Radar (SAR) Data Analysis of Tropical Cyclones:
615 Comparisons with SFMR Flights and Best Track. *Monthly Weather Review*, **148** (11), 4545–
616 4563, doi:10.1175/MWR-D-20-0005.1, URL <https://doi.org/10.1175/MWR-D-20-0005.1>.
- 617 Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance
618 of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137** (656), 553–597, URL <https://doi.org/10.1002/qj.828>.
- 620 DeMaria, M., M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan, 2005: Further Improvements
621 to the Statistical Hurricane Intensity Prediction Scheme (SHIPS). *Wea. Forecasting*, **20** (4),
622 531–543, doi:10.1175/waf862.1, URL <https://doi.org/10.1175%2Fwaf862.1>.
- 623 DiNapoli, S. M., M. A. Bourassa, and M. D. Powell, 2012: Uncertainty and Intercalibration
624 Analysis of H*Wind. *Journal of Atmospheric and Oceanic Technology*, **29** (6), 822–833, doi:
625 10.1175/jtech-d-11-00165.1, URL <https://doi.org/10.1175%2Fjtech-d-11-00165.1>.
- 626 Done, J. M., M. Ge, G. J. Holland, I. Dima-West, S. Phibbs, G. R. Saville, and Y. Wang,
627 2020: Modelling global tropical cyclone wind footprints. *Natural Hazards and Earth Sys-*

628 *tem Sciences*, **20** (2), 567–580, doi:10.5194/nhess-20-567-2020, URL <https://doi.org/10.5194%2Fnhess-20-567-2020>.

630 Donelan, M. A., 2004: On the limiting aerodynamic roughness of the ocean in very strong
631 winds. *Geophysical Research Letters*, **31** (18), doi:10.1029/2004gl019460, URL <https://doi.org/10.1029%2F2004gl019460>.

633 Draper, D., and D. Long, 2004: Simultaneous wind and rain retrieval using SeaWinds data. *IEEE*
634 *Transactions on Geoscience and Remote Sensing*, **42** (7), 1411–1423, doi:10.1109/tgrs.2004.
635 830169, URL <https://doi.org/10.1109%2Ftgrs.2004.830169>.

636 Draper, D. W., and D. G. Long, 2002: An assessment of SeaWinds on QuikSCAT wind retrieval.
637 *Journal of Geophysical Research: Oceans*, **107** (C12), 5–1–5–14, doi:10.1029/2002jc001330,
638 URL <https://doi.org/10.1029%2F2002jc001330>.

639 Emanuel, K., 2004: Tropical cyclone energetics and structure. *Atmospheric Turbu-*
640 *lence and Mesoscale Meteorology*, Cambridge University Press, 165–192, doi:10.1017/
641 cbo9780511735035.010, URL <https://doi.org/10.1017%2Fcbo9780511735035.010>.

642 Emanuel, K., S. Ravela, E. Vivant, and C. Risi, 2006: A Statistical Deterministic Approach to
643 Hurricane Risk Assessment. *Bulletin of the American Meteorological Society*, **87** (3), 299–314,
644 doi:10.1175/bams-87-3-299, URL <https://doi.org/10.1175%2Fbams-87-3-299>.

645 Emanuel, K., and R. Rotunno, 2011: Self-Stratification of Tropical Cyclone Outflow. Part I: Impli-
646 cations for Storm Structure. *J. Atmos. Sci.*, **68** (10), 2236–2249, doi:10.1175/jas-d-10-05024.1,
647 URL <https://doi.org/10.1175%2Fjas-d-10-05024.1>.

- 648 Friedman, J. H., 2001: Greedy function approximation: A gradient boosting machine. *The Annals*
649 *of Statistics*, **29** (5), 1189 – 1232, doi:10.1214/aos/1013203451, URL <https://doi.org/10.1214/aos/1013203451>.
- 650
- 651 Geiger, T., K. Frieler, and A. Levermann, 2016: High-income does not protect against hurricane
652 losses. *Environ. Res. Lett.*, **11** (8), 084012, doi:10.1088/1748-9326/11/8/084012, URL <https://doi.org/10.1088%2F1748-9326%2F11%2F8%2F084012>.
- 653
- 654 Hall, T. M., and S. Jewson, 2007: Statistical modelling of North Atlantic tropical cyclone tracks.
655 *Tellus*, **59A**, doi:10.3402/tellusa.v59i4.15017, URL <https://doi.org/10.3402%2Ftellusa.v59i4.15017>.
- 656
- 657 Holland, G. J., 1980: An Analytic Model of the Wind and Pressure Profiles in Hurricanes. *Mon.*
658 *Wea. Rev.*, **108** (8), 1212–1218, doi:10.1175/1520-0493(1980)108<1212:AAMOTW>2.0.CO;
659 2, URL [https://doi.org/10.1175/1520-0493\(1980\)108<1212:AAMOTW>2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108<1212:AAMOTW>2.0.CO;2).
- 660
- 661 Holland, G. J., J. I. Belanger, and A. Fritz, 2010: A Revised Model for Radial Profiles of
662 Hurricane Winds. *Mon. Wea. Rev.*, **138** (12), 4393–4401, doi:10.1175/2010mwr3317.1, URL
<https://doi.org/10.1175%2F2010mwr3317.1>.
- 663
- 664 Jing, R., and N. Lin, 2020: An Environment-Dependent Probabilistic Tropical Cyclone Model.
665 *Journal of Advances in Modeling Earth Systems*, **12** (3), doi:10.1029/2019ms001975, URL
<https://doi.org/10.1029%2F2019ms001975>.
- 666
- 667 Jordan, M. R., and C. A. Clayson, 2008: A new approach to using wind speed for prediction of
668 tropical cyclone generated storm surge. *Geophysical Research Letters*, **35** (13), doi:10.1029/
2008gl033564, URL <https://doi.org/10.1029%2F2008gl033564>.

- 669 Kim, S., H. Kim, J. Lee, S. Yoon, S. E. Kahou, K. Kashinath, and M. Prabhat, 2019: Deep-
670 Hurricane-Tracker: Tracking and Forecasting Extreme Climate Events. *2019 IEEE Winter Con-*
671 *ference on Applications of Computer Vision (WACV)*, IEEE, doi:10.1109/wacv.2019.00192, URL
672 <https://doi.org/10.1109%2Fwacv.2019.00192>.
- 673 Klausmann, A., 2014: Analysis of Hurricane Irene's Wind Field Using the Advanced Research
674 Weather Research and Forecast (WRF-ARW) Model. *Journal of Marine Science and Engineer-*
675 *ing*, **2** (1), 33–45, doi:10.3390/jmse2010033, URL <https://doi.org/10.3390%2Fjmse2010033>.
- 676 Klotz, B. W., and H. Jiang, 2016: Global composites of surface wind speeds in tropical cyclones
677 based on a 12-year scatterometer database. *Geophysical Research Letters*, **43** (19), 10,480–
678 10,488, doi:10.1002/2016gl071066, URL <https://doi.org/10.1002%2F2016gl071066>.
- 679 Klotz, B. W., and H. Jiang, 2017: Examination of Surface Wind Asymmetries in Tropical Cyclones.
680 Part I: General Structure and Wind Shear Impacts. *Monthly Weather Review*, **145** (10), 3989–
681 4009, doi:10.1175/mwr-d-17-0019.1, URL <https://doi.org/10.1175%2Fmwr-d-17-0019.1>.
- 682 Knaff, J. A., and D. R. Chavas, 2021: Efforts to Estimate the Radius of Maximum Winds in
683 Tropical Cyclones. *AMS 101 Annual Meeting*.
- 684 Knaff, J. A., M. DeMaria, D. A. Molenar, C. R. Sampson, and M. G. Seybold, 2011: An Automated,
685 Objective, Multiple-Satellite-Platform Tropical Cyclone Surface Wind Analysis. *Journal of*
686 *Applied Meteorology and Climatology*, **50** (10), 2149–2166.
- 687 Knaff, J. A., S. P. Longmore, R. T. DeMaria, and D. A. Molenar, 2015: Improved Tropical-
688 Cyclone Flight-Level Wind Estimates Using Routine Infrared Satellite Reconnaissance. *Journal*
689 *of Applied Meteorology and Climatology*, **54** (2), 463–478, doi:10.1175/jamc-d-14-0112.1, URL
690 <https://doi.org/10.1175%2Fjamc-d-14-0112.1>.

- 691 Knaff, J. A., C. R. Sampson, and G. Chirokova, 2017: A Global Statistical–Dynamical Tropical
692 Cyclone Wind Radii Forecast Scheme. *Weather and Forecasting*, **32** (2), 629–644, doi:10.1175/
693 waf-d-16-0168.1, URL <https://doi.org/10.1175%2Fwaf-d-16-0168.1>.
- 694 Knutson, T. R., and Coauthors, 2013: Dynamical Downscaling Projections of Twenty-First-
695 Century Atlantic Hurricane Activity: CMIP3 and CMIP5 Model-Based Scenarios. *Journal*
696 *of Climate*, **26** (17), 6591–6617, doi:10.1175/jcli-d-12-00539.1, URL <https://doi.org/10.1175%2Fjcli-d-12-00539.1>.
- 697 Komaromi, W. A., and J. D. Doyle, 2018: On the Dynamics of Tropical Cyclone and Trough Inter-
698 actions. *Journal of the Atmospheric Sciences*, **75** (8), 2687–2709, doi:10.1175/jas-d-17-0272.1,
699 URL <https://doi.org/10.1175%2Fjas-d-17-0272.1>.
- 700 Landsea, C. W., and J. L. Franklin, 2013: Atlantic Hurricane Database Uncertainty and Presentation
701 of a New Database Format. *Mon. Wea. Rev.*, **141** (10), 3576–3592, doi:10.1175/mwr-d-12-00254.
702 1, URL <https://doi.org/10.1175%2Fmwr-d-12-00254.1>.
- 703 Lee, C.-Y., S. J. Camargo, A. H. Sobel, and M. K. Tippett, 2020: Statistical–Dynamical Down-
704 scaling Projections of Tropical Cyclone Activity in a Warming Climate: Two Diverging Genesis
705 Scenarios. *Journal of Climate*, **33** (11), 4815–4834.
- 706 Lee, C.-Y., and S. S. Chen, 2012: Symmetric and Asymmetric Structures of Hurricane
707 Boundary Layer in Coupled Atmosphere–Wave–Ocean Models and Observations. *Journal*
708 *of the Atmospheric Sciences*, **69** (12), 3576–3594, doi:10.1175/jas-d-12-046.1, URL <https://doi.org/10.1175%2Fjas-d-12-046.1>.
- 709 Lee, C.-Y., and S. S. Chen, 2014: Stable Boundary Layer and Its Impact on Tropical Cyclone
710 Structure in a Coupled Atmosphere–Ocean Model. *Monthly Weather Review*, **142** (5), 1927–

- 713 1944, doi:10.1175/mwr-d-13-00122.1, URL <https://doi.org/10.1175%2Fmwr-d-13-00122.1>.
- 714 715 Lee, C.-Y., M. K. Tippett, S. J. Camargo, and A. H. Sobel, 2015: Probabilistic Multiple Linear
Regression Modeling for Tropical Cyclone Intensity. *Mon. Wea. Rev.*, **143** (3), 933–954, doi:
10.1175/mwr-d-14-00171.1, URL <https://doi.org/10.1175%2Fmwr-d-14-00171.1>.
- 716 717 718 719 Lee, C.-Y., M. K. Tippett, A. H. Sobel, and S. J. Camargo, 2016: Rapid intensification and the bimodal distribution of tropical cyclone intensity. *Nat. Commun.*, **7** (1), doi: 10.1038/ncomms10625, URL <https://doi.org/10.1038%2Fncomms10625>.
- 720 721 722 Lee, C.-Y., M. K. Tippett, A. H. Sobel, and S. J. Camargo, 2018: An Environmentally Forced Tropical Cyclone Hazard Model. *Journal of Advances in Modeling Earth Systems*, **10** (1), 223–241, doi:10.1002/2017ms001186, URL <https://doi.org/10.1002%2F2017ms001186>.
- 723 724 725 Lin, N., and D. Chavas, 2012: On hurricane parametric wind and applications in storm surge modeling. *Journal of Geophysical Research: Atmospheres*, **117** (D9), n/a–n/a, doi:10.1029/2011jd017126, URL <https://doi.org/10.1029%2F2011jd017126>.
- 726 727 728 Loridan, T., R. P. Crompton, and E. Dubossarsky, 2017: A Machine Learning Approach to Modeling Tropical Cyclone Wind Field Uncertainty. *Mon. Wea. Rev.*, **145** (8), 3203–3221, doi:10.1175/mwr-d-16-0429.1, URL <https://doi.org/10.1175%2Fmwr-d-16-0429.1>.
- 729 730 731 732 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. *Journal of Applied Meteorology and Climatology*, **54** (3), 624–642, doi:10.1175/jamc-d-14-0095.1, URL <https://doi.org/10.1175%2Fjamc-d-14-0095.1>.

733 Mei, W., C. Pasquero, and F. Primeau, 2012: The effect of translation speed upon the intensity
734 of tropical cyclones over the tropical ocean. *Geophysical Research Letters*, **39** (7), n/a–n/a,
735 doi:10.1029/2011gl050765, URL <https://doi.org/10.1029%2F2011gl050765>.

736 Mouche, A., B. Chapron, J. Knaff, Y. Zhao, B. Zhang, and C. Combot, 2019: Copolarized
737 and Cross-Polarized SAR Measurements for High-Resolution Description of Major Hurri-
738 cane Wind Structures: Application to Irma Category 5 Hurricane. *Journal of Geophysical*
739 *Research: Oceans*, **124** (6), 3905–3922, doi:10.1029/2019jc015056, URL [https://doi.org/10.](https://doi.org/10.1029%2F2019jc015056)
740 1029%2F2019jc015056.

741 Mueller, K. J., M. DeMaria, J. A. Knaff, J. P. Kossin, and T. H. V. Haar, 2006: Objective Estimation
742 of Tropical Cyclone Wind Structure from Infrared Satellite Data. *Weather and Forecasting*,
743 **21** (6), 990 – 1005, doi:10.1175/WAF955.1, URL <https://doi.org/10.1175/WAF955.1>.

744 Needham, H. F., and B. D. Keim, 2014: Correlating Storm Surge Heights with Tropical Cyclone
745 Winds at and before Landfall. *Earth Interactions*, **18** (7), 1–26, doi:10.1175/2013ei000527.1,
746 URL <https://doi.org/10.1175%2F2013ei000527.1>.

747 Olfateh, M., D. P. Callaghan, P. Nielsen, and T. E. Baldock, 2017: Tropical cyclone wind field
748 asymmetry-Development and evaluation of a new parametric model. *Journal of Geophysical Re-*
749 *search: Oceans*, **122** (1), 458–469, doi:10.1002/2016jc012237, URL <https://doi.org/10.1002%2F2016jc012237>.

751 Peduzzi, P., B. Chatenoux, H. Dao, A. D. Bono, C. Herold, J. Kossin, F. Mouton, and O. Nordbeck,
752 2012: Global trends in tropical cyclone risk. *Nat. Clim. Change*, **2** (4), 289–294, doi:10.1038/nclimate1410, URL <https://doi.org/10.1038%2Fnclimate1410>.

- 754 Powell, M. D., S. H. Houston, L. R. Amat, and N. Morisseau-Leroy, 1998: The HRD real-time hur-
755 ricane wind analysis system. *Journal of Wind Engineering and Industrial Aerodynamics*, **77-78**,
756 53–64, doi:10.1016/s0167-6105(98)00131-7, URL <https://doi.org/10.1016%2Fs0167-6105%2898%2900131-7>.
- 757
- 758 Quinlan, J. R., 1986: Induction of Decision Trees. *Mach. Learn.*, **1 (1)**, 81–106, doi:10.1023/A:
759 1022643204877, URL <https://doi.org/10.1023/A:1022643204877>.
- 760 Racah, E., C. Beckham, T. Maharaj, S. Kahou, Prabhat, and C. Pal, 2017: ExtremeWeather: A
761 large-scale climate dataset for semi-supervised detection, localization, and understanding of
762 extreme weather events. *NIPS*.
- 763 Roberts, M. J., and Coauthors, 2020: Impact of Model Resolution on Tropical Cyclone Simulation
764 Using the HighResMIP-PRIMAVERA Multimodel Ensemble. *Journal of Climate*, **33 (7)**, 2557–
765 2583, doi:10.1175/jcli-d-19-0639.1, URL <https://doi.org/10.1175%2Fjcli-d-19-0639.1>.
- 766 Shapiro, L. J., 1983: The Asymmetric Boundary layer Flow Under a Translating Hurricane. *Jour-*
767 *768 nal of the Atmospheric Sciences*, **40 (8)**, 1984–1998, doi:10.1175/1520-0469(1983)040<1984:
TABLFU>2.0.CO;2, URL [https://doi.org/10.1175/1520-0469\(1983\)040<1984:TABLFU>2.0.](https://doi.org/10.1175/1520-0469(1983)040<1984:TABLFU>2.0.)
769 CO;2.
- 770 Sun, Z., B. Zhang, J. A. Zhang, and W. Perrie, 2019: Examination of Surface Wind Asymmetry
771 in Tropical Cyclones over the Northwest Pacific Ocean Using SMAP Observations. *Remote*
772 *Sensing*, **11 (22)**, 2604, doi:10.3390/rs11222604, URL <https://doi.org/10.3390%2Fr11222604>.
- 773 Ueno, M., and K. Bessho, 2011: A Statistical Analysis of Near-Core Surface Wind Asymmetries
774 in Typhoons Obtained from QuikSCAT Data. *Journal of the Meteorological Society of Japan.*
775 *Ser. II*, **89 (3)**, 225–241, doi:10.2151/jmsj.2011-304.

- 776 Uhlhorn, E. W., B. W. Klotz, T. Vukicevic, P. D. Reasor, and R. F. Rogers, 2014: Observed
777 Hurricane Wind Speed Asymmetries and Relationships to Motion and Environmental Shear.
778 *Mon. Wea. Rev.*, **142** (3), 1290–1311, doi:10.1175/mwr-d-13-00249.1, URL <https://doi.org/10.1175%2Fmwr-d-13-00249.1>.
- 780 Willoughby, H. E., R. W. R. Darling, and M. E. Rahn, 2006: Parametric Representation of the
781 Primary Hurricane Vortex. Part II: A New Family of Sectionally Continuous Profiles. *Mon. Wea.
782 Rev.*, **134** (4), 1102–1120, doi:10.1175/mwr3106.1, URL <https://doi.org/10.1175%2Fmwr3106.1>.
783 1.
- 784 Willoughby, H. E., and M. E. Rahn, 2004: Parametric Representation of the Primary Hurricane
785 Vortex. Part I: Observations and Evaluation of the Holland (1980) Model. *Mon. Wea. Rev.*,
786 **132** (12), 3033–3048, doi:10.1175/mwr2831.1, URL <https://doi.org/10.1175%2Fmwr2831.1>.
- 787 Wong, M. L. M., and J. C. L. Chan, 2007: Modeling the Effects of Land–Sea Roughness Contrast
788 on Tropical Cyclone Winds. *Journal of the Atmospheric Sciences*, **64** (9), 3249–3264, doi:
789 10.1175/jas4027.1, URL <https://doi.org/10.1175%2Fjas4027.1>.
- 790 Wood, V. T., L. W. White, H. E. Willoughby, and D. P. Jorgensen, 2013: A New Parametric Tropical
791 Cyclone Tangential Wind Profile Model. *Monthly Weather Review*, **141** (6), 1884–1909, doi:
792 10.1175/mwr-d-12-00115.1, URL <https://doi.org/10.1175%2Fmwr-d-12-00115.1>.
- 793 Yang, Q., C.-Y. Lee, and M. K. Tippett, 2020: A Long Short-Term Memory Model for Global Rapid
794 Intensification Prediction. *Wea. Forecasting*, **35** (4), 1203–1220, doi:10.1175/waf-d-19-0199.1,
795 URL <https://doi.org/10.1175%2Fwaf-d-19-0199.1>.

796 LIST OF TABLES

797	Table 1. Storm variables.	39
798	Table 2. Environment variables. Variables presented here are derived at storm center, 799 averaged over a disk, or averaged over an annulus. Disk-averaged variables are 800 averaged over the area within 500 km from storm center while annulus-averaged 801 ones are averaged over a ring-shaped area with inner radius of 200 km and outer 802 radius of 800 km from storm center.	40
803	Table 3. Environment variables (continued).	41
804	Table 4. XGBoost model configuration example	42
805	Table 5. MSESS for predicted coefficients by XGBoost models on testing set.	43

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
$\frac{dS}{dt}$	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

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{PI}_{pd}	Potential intensity in minimum sea-level pressure, disk-averaged
PI_s	Potential intensity in maximum wind speed at storm center
\overline{PI}_{sd}	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 a 100 km storm-centered, disk averaged
T_{200}	Temperature at 200hPa 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 mb at storm center in earth coordinates
\overline{SH}_{za}	SH_z , annulus-averaged
SH_m	Vertical wind shear (meridional) between 850 and 200 mb at storm center in earth coordinates
\overline{SH}_{ma}	SH_m , annulus-averaged
D_{200}	Divergence field at 200hPa
\overline{D}_{200_a}	D_{200} , annulus-averaged
$\frac{dPI_p}{dt}$	Change in potential intensity of minimum sea-level pressure over 12 hours

TABLE 3: Environment variables (continued).

Notation	Definition
RH_l	Low level relative humidity (850 to 700 hPa) at storm center
\overline{RH}_{l_a}	RH_l , annulus-averaged
RH_h	High level relative humidity (500 to 300 hPa)
\overline{RH}_{h_a}	RH_h , annulus-averaged
SLM	Storm location mark: land or ocean
$P - S$	$\overline{PI_{sd}} - S_{max}$

TABLE 4: XGBoost model configuration example

Hyper-parameter	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

TABLE 5: MSESS for predicted coefficients by XGBoost models on testing set.

(a) MSESS for A_n and B_n

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

(b) MSESS for $a_{m,n}$

MSESS	$n = 1$	$n = 2$	$n = 3$	$n = 4$
$m = 1$	0.50	0.11	0.44	0.13
$m = 2$	0.04	0.04	0.00	0.01
$m = 3$	-0.02	0.00	0.00	0.01

(c) MSESS for $b_{m,n}$

MSESS	$n = 1$	$n = 2$	$n = 3$	$n = 4$
$m = 1$	0.09	0.06	0.08	0.08
$m = 2$	-0.42	-0.04	-0.05	0.00
$m = 3$	-0.02	0.00	0.01	0.00

806 LIST OF FIGURES

807	Fig. 1.	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.	46
811	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 motion (a-c) and shear (d-f) coordinates. In motion coordinates, up is down-wind (the storm translation direction, DW); down is up-wind (UW); left is left-of-wind (LW); and right is right-of-wind (RW). For the shear coordinates, up is down-shear (the vertical wind shear direction, DS); down is up-shear (US); left is left-of-shear (LS); and right is right-of-shear (RS). (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. (c) and (f) are the differences between (a) and (b), and (d) and (e).	47
821	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.	48
825	Fig. 4.	The root mean square error (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.	49
830	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 panels (a) and (b), respectively.	50
833	Fig. 6.	Magnitude of asymmetries at m and n	51
834	Fig. 7.	A reconstruction case study of one wind field from Hurricane Issac at 1600 UTC August 28, 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) is the HWIND truth, (b) is the approximation, and (c) is the reconstruction. The second row (d-f) shows errors from the approximation (against the truth), the reconstruction (against the truth), and the XGBoost prediction space (against the approximation).	52
842	Fig. 8.	Mean square error skill score (MSESS) of XGBoost reconstructed wind from testing set on normalized radius. (a) and (b) are the MSESS of the reconstructed symmetric fields comparing to symmetric field climatology and Willoughby06 parametric fields. (c-e) are the MSESS of the reconstructed asymmetric fields comparing to the HWIND climatology in earth, motion, and shear coordinates. (f-h) are the same as (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.	53

Fig. 9. 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. 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 10-fold cross validation and the vertical bars denote the variance of the validation. Along x-axis only 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. 54

Fig. 11. $P_{1,1}$ 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 to 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.

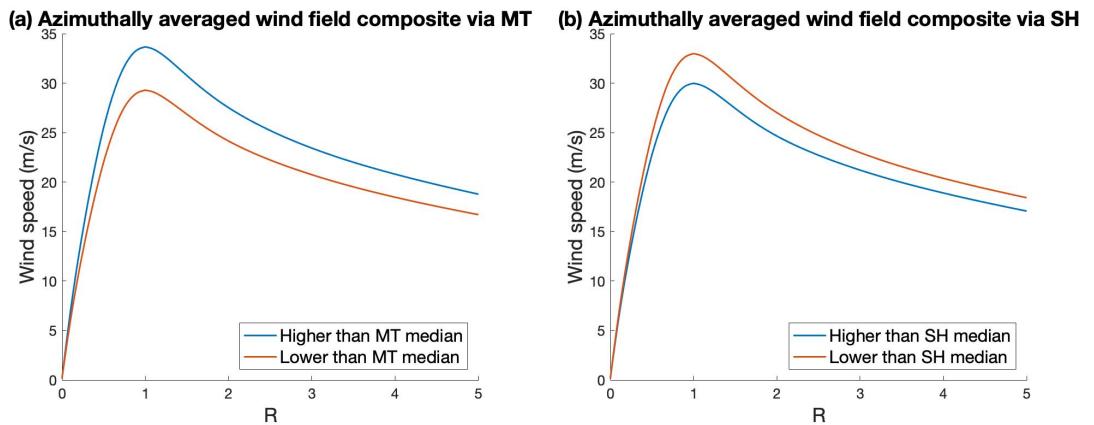


FIG. 1: Radial wind speed profile composites of the azimuthally averaged HWIND wind fields. The x-axis shows radii normalized by radius of maximum wind speed (R_{max}). In (a) and (b), blue (orange) curves are composites with translation speed MT and shear magnitude SH greater (less) than the respective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

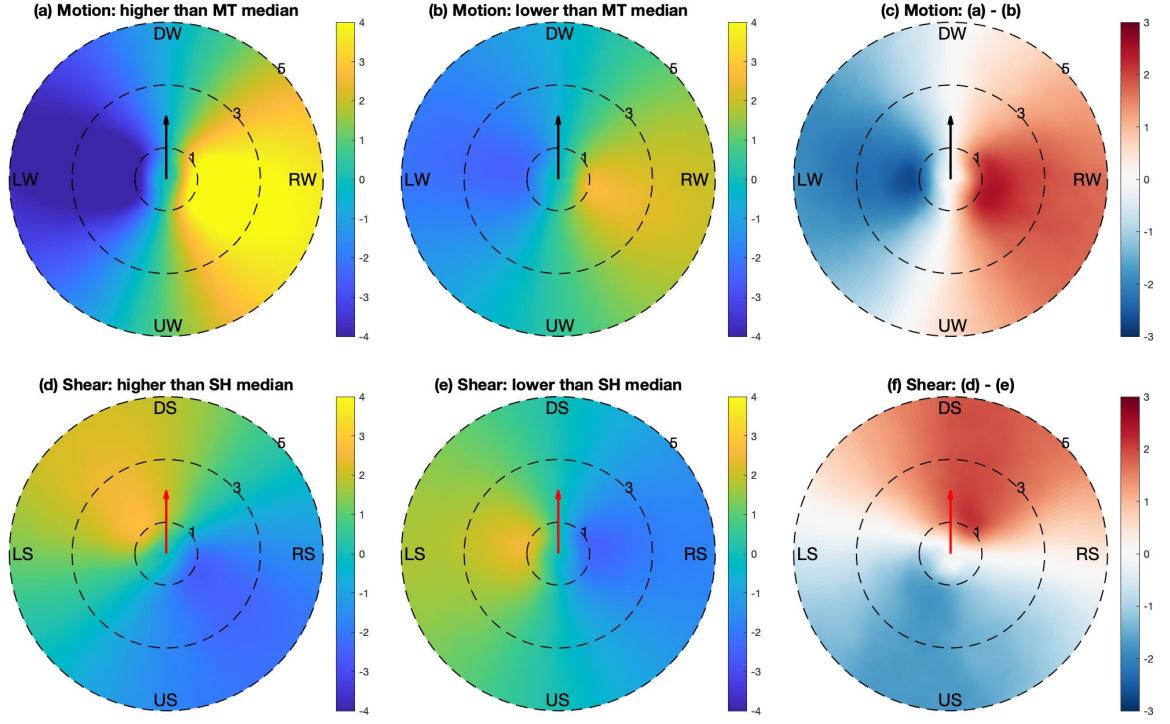


FIG. 2: HWIND asymmetric wind field composites based on storm translation speed MT and vertical wind shear magnitude SH . The dashed circle and the number beside it represent the normalized radius. The black and red arrows denote storm motion and wind shear directions. The composites are constructed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

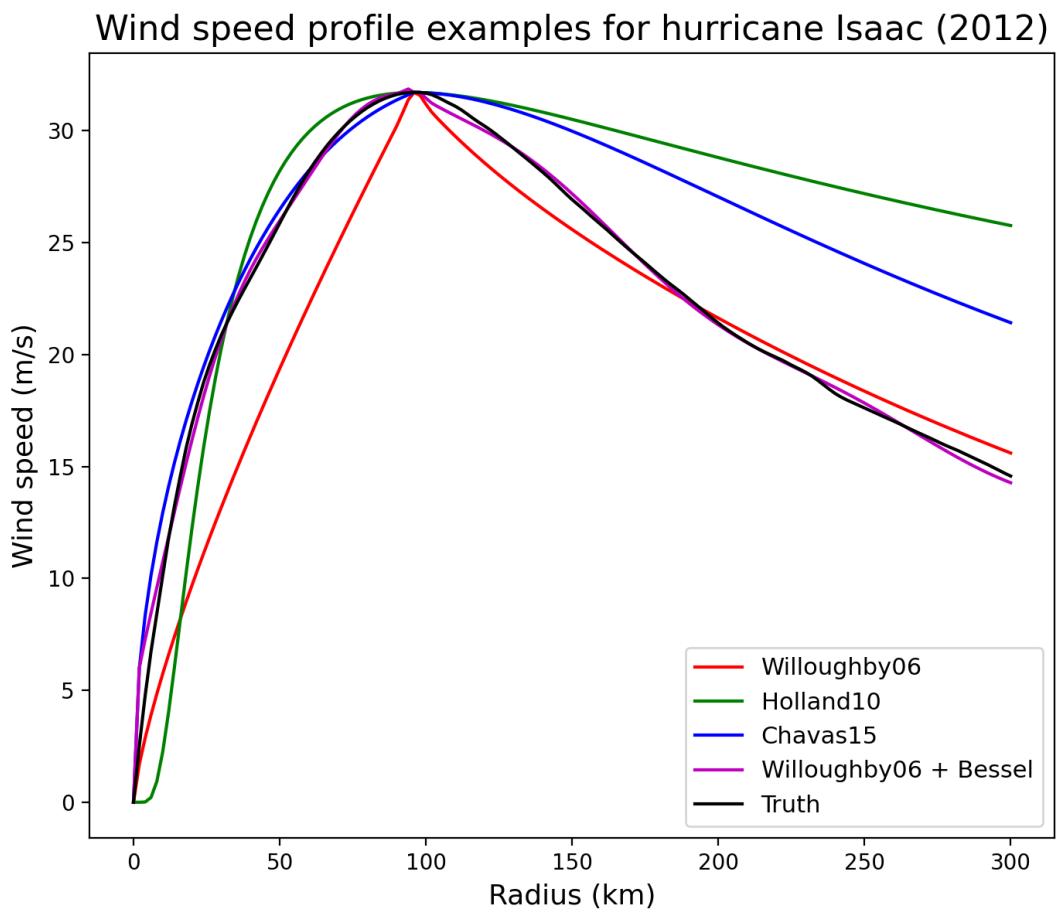


FIG. 3: An azimuthally averaged wind profile of Hurricane Isaac (2012) from HWIND (black), and the corresponding simulated profiles from Willoughby06, Holland10, and Chavas15 denoted by the red, green, and blue curves, respectively. The magenta curve is the profile approximated by Willoughby06 along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

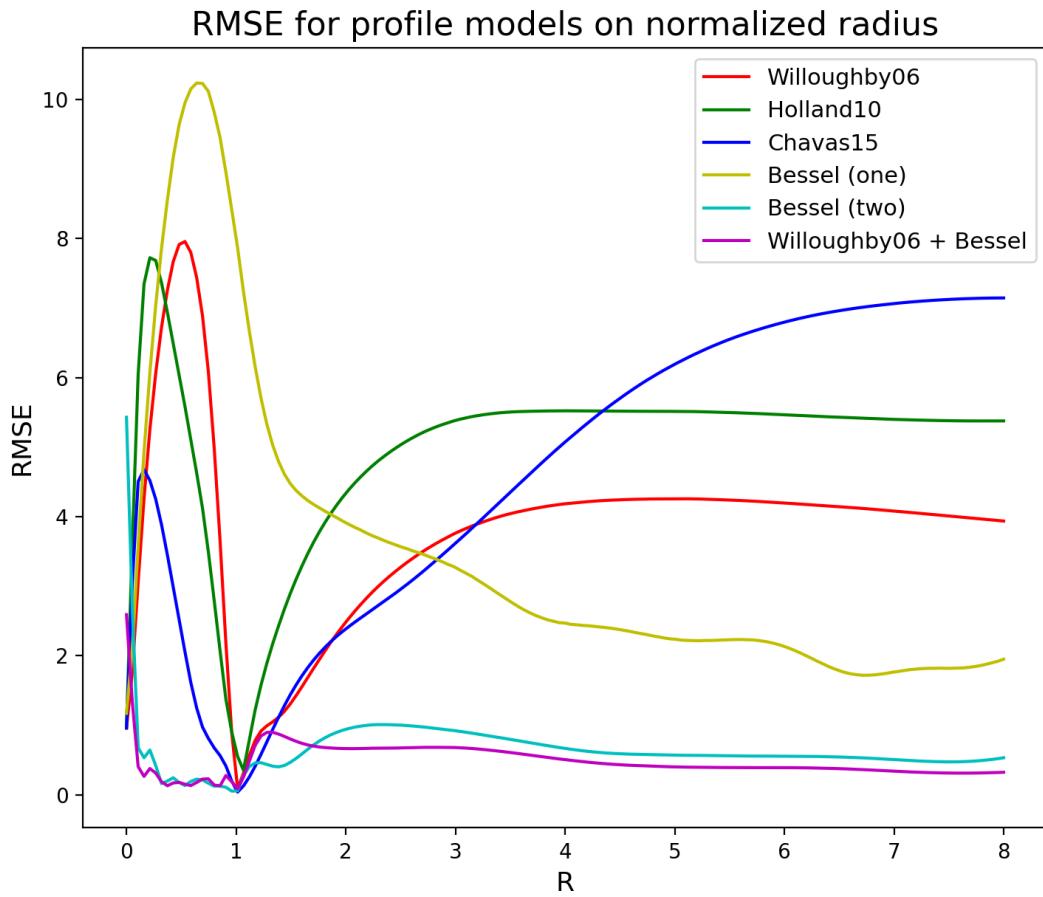


FIG. 4: The root mean square error (RMSE) of the Willoughby06 (red), Holland10 (green), and Chavas15 (blue) models over HWIND dataset in normalized radius. The yellow (cyan) curve is the RMSE from the profiles fitted by a single (two) series of Bessel functions; while the magenta curve is from the profiles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

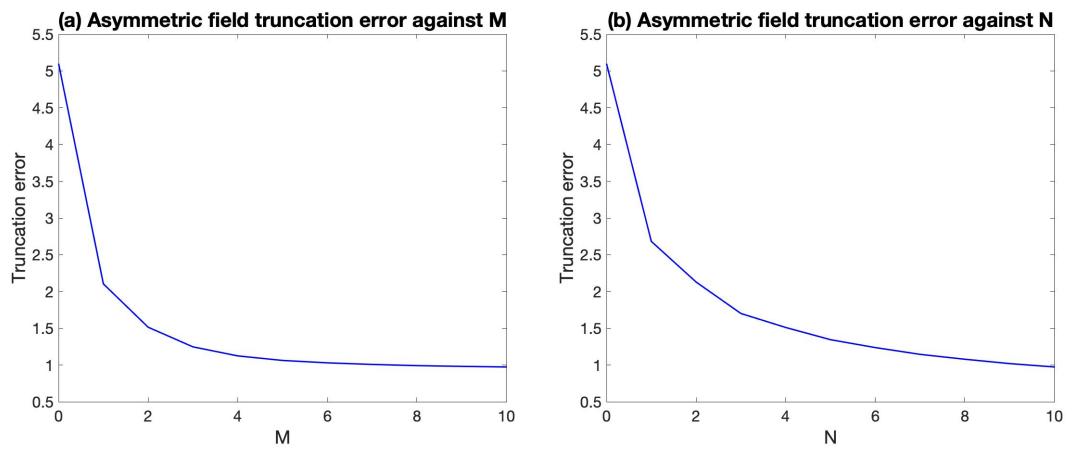


FIG. 5: The truncation error of asymmetric fields with increasing (a) M and (b) N . The M and N are the upper limits of the sums in Eq. 4. N and M are kept as constants as 10 in panels (a) and (b), respectively.

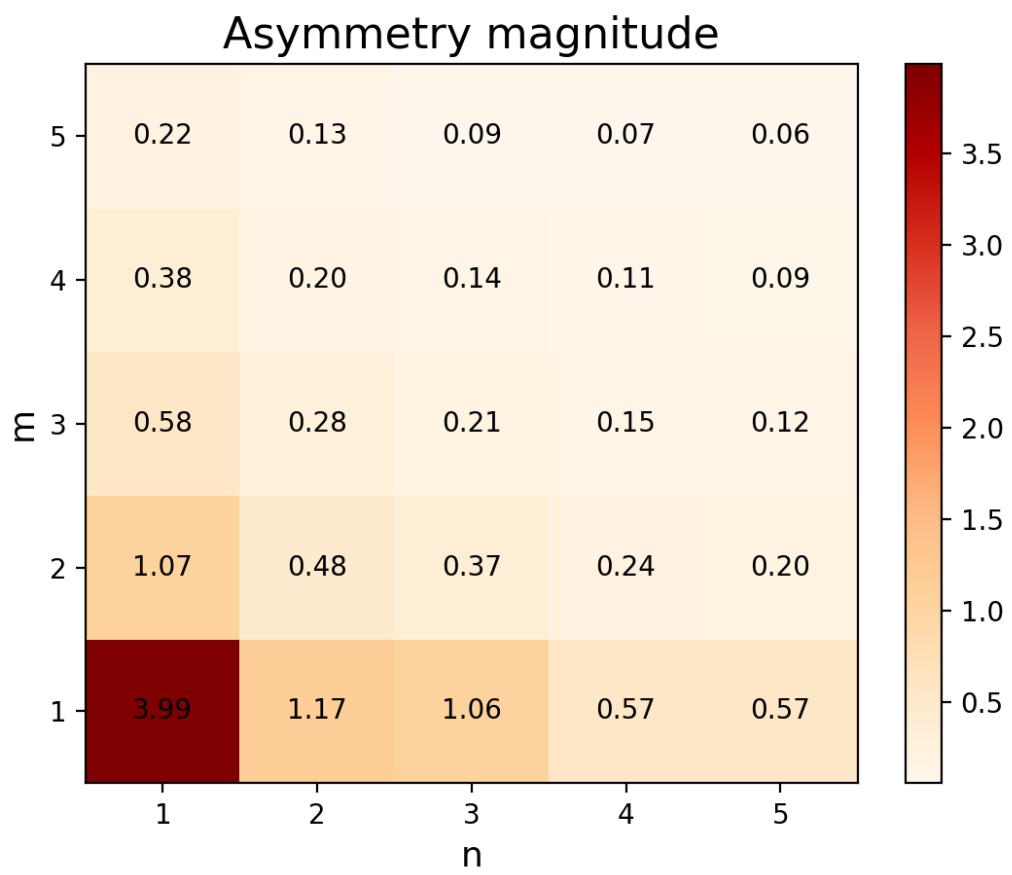


FIG. 6: Magnitude of asymmetries at m and n .

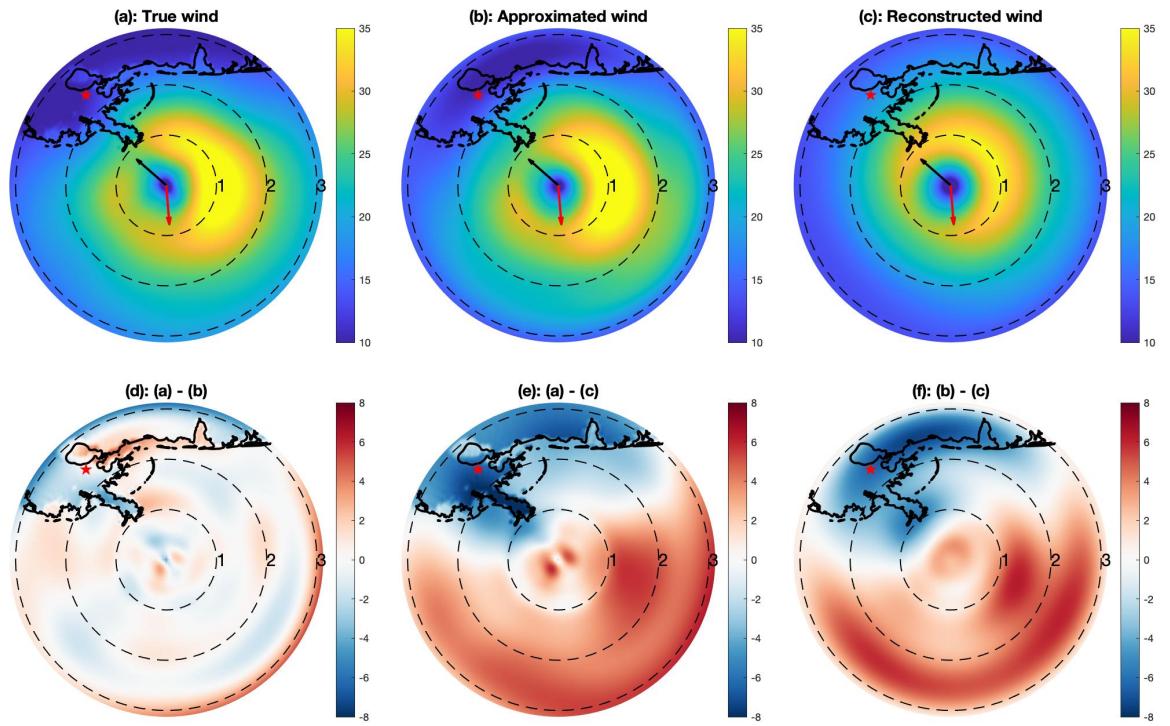


FIG. 7: A reconstruction case study of one wind field from Hurricane Isaac at 1600 UTC August 28, 2012. Plotted wind fields are in earth coordinates with radii up to 300 km from storm center, and the dashed circle and the small number beside it indicate the locations of 1, 2, and 3 R_{max} . Land coastline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

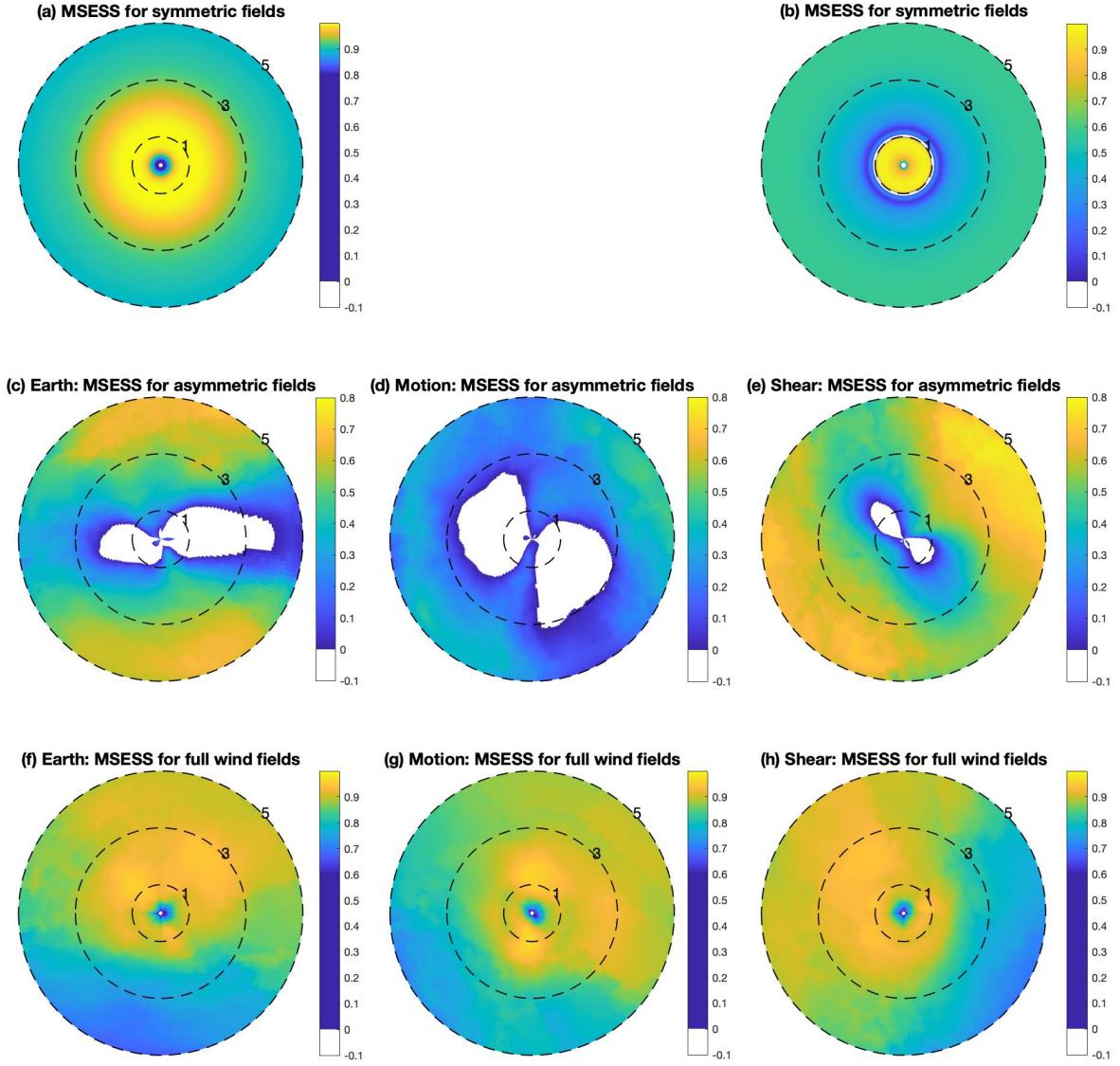


FIG. 8: Mean square error skill score (MSESS) of XGBoost reconstructed wind from testing set on normalized radius. (a) and (b) are the MSESS of the reconstructed symmetric fields comparing to symmetric field climatology and Willoughby06 parametric fields. (c-e) are the MSESS of the reconstructed asymmetric fields comparing to the HWIND climatology in earth, motion, and shear coordinates. (f-h) are the same as (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.

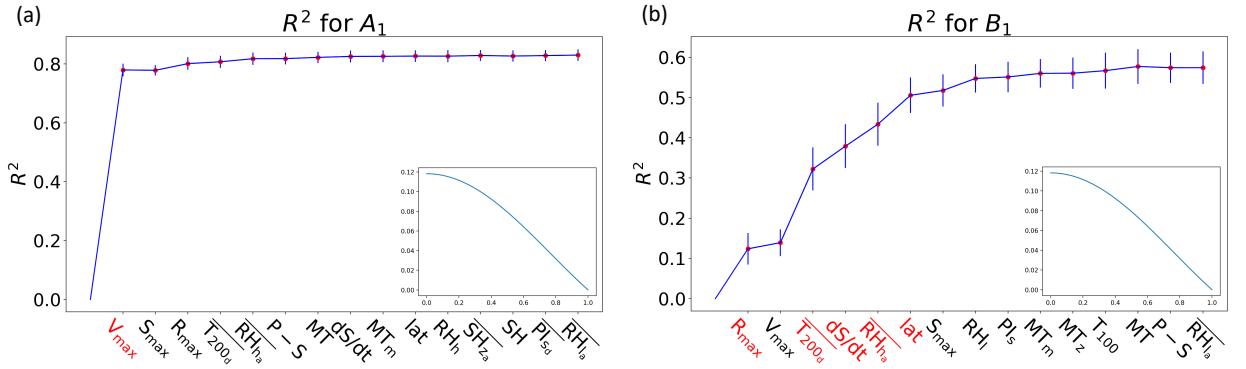


FIG. 9: R^2 of Bessel function coefficients A_1 and B_1 which are responsible for parametric wind profile correction inside and outside the R_{max} , respectively. R^2 is calculated from predictions of the diagnostic XGBoost models that are trained and run with inclusion of one more predictor at a time. The red dots represent mean R^2 from 10-fold cross validation and the vertical bars denote the variance of the validation. Along x-axis only 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.

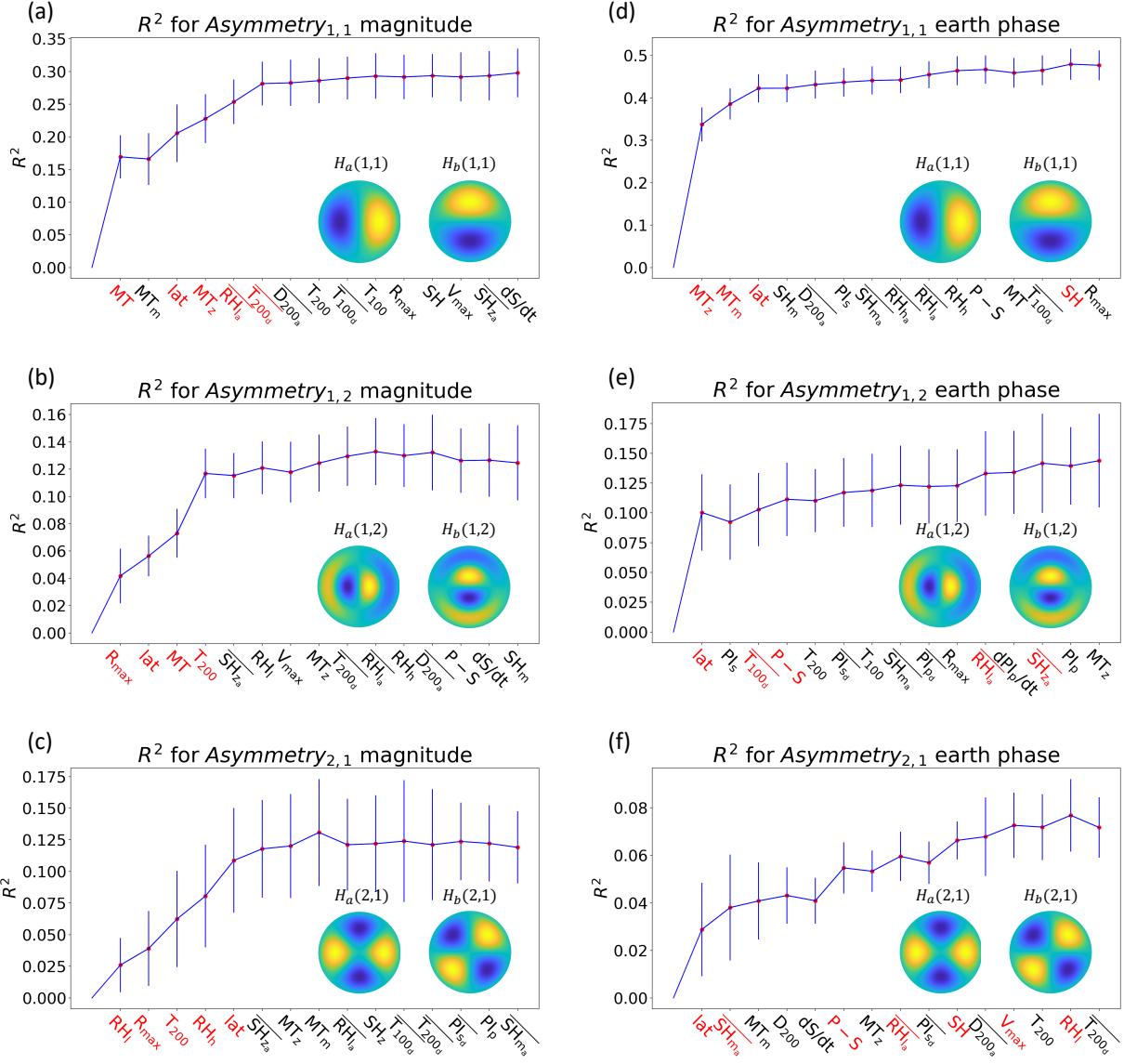


FIG. 10: Similar to Fig. 9 but for asymmetry magnitude (a-c) and phase (d-f) which are constructed from the coefficients predicted by the diagnostic XGBoost models. Note that a different vertical axis scaling is used for each panel in the figure. Panel (a) shows the dominant contribution of storm translation 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.

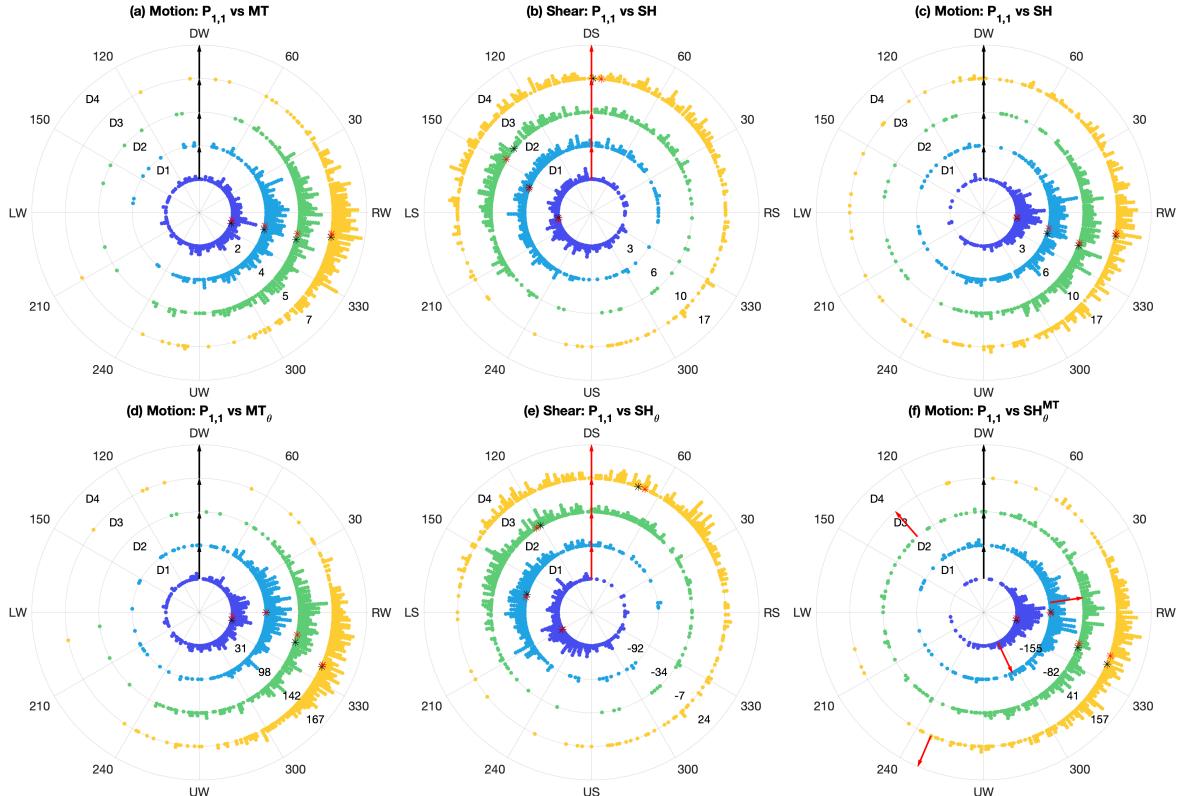


FIG. 11: $P_{1,1}$ is the phase of the asymmetric wind component for wavenumber one in the azimuthal and radial directions. This figure shows histograms of $P_{1,1}$ conditional on a variable of interest as indicated in each panel's title. The panel title also shows which coordinate the $P_{1,1}$ is 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 to 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.