A Machine Learning Approach to Modeling Tropical Cyclone Wind Field Uncertainty

THOMAS LORIDAN AND RYAN P. CROMPTON
Risk Frontiers, Macquarie University, Sydney, Australia

EUGENE DUBOSSARSKY
Prescient, Sydney, Australia

(Manuscript received 12 November 2016, in final form 22 April 2017)

ABSTRACT

Tropical cyclone (TC) risk assessment models and probabilistic forecasting systems rely on large ensembles to simulate the track trajectories, intensities, and spatial distributions of damaging winds from severe events. Given computational constraints associated with the generation of such ensembles, the representation of TC winds is typically based on very simple parametric formulations. Such models strongly underestimate the full range of TC wind field variability and thus do not allow for accurate representation of the risk profile. With this in mind, this study explores the potential of machine learning algorithms as an alternative to current parametric methods. First, a catalog of high-resolution TC wind simulations is assembled for the western North Pacific using the Weather Research and Forecasting (WRF) Model. The simulated wind fields are then decomposed via principal component analysis (PCA) and a quantile regression forest model is trained to predict the conditional distributions of the first three principal component (PC) weights. With this model, predictions can be made for any quantiles in the distributions of the PC weights thereby providing a way to account for uncertainty in the modeled wind fields. By repeatedly sampling the quantile values, probabilistic maps for the likelihood of attaining given wind speed thresholds can be easily generated. Similarly the inclusion of such a model as part of a TC risk assessment framework can greatly increase the range of wind field patterns sampled, providing a broader view of the threat posed by TC winds.

1. Introduction

Tropical cyclone (TC) winds can wreak havoc on communities by directly damaging properties, crops, and threatening human lives; or indirectly from their contribution to TC-induced storm surge, waves, and flood risk. As a result, considerable resources have been invested in TC risk assessment and warning systems. Risk assessment (e.g., Vickery et al. 2009) and real-time probabilistic forecasting (e.g., DeMaria et al. 2009, 2013; Sampson et al. 2016) typically employ large ensembles of TC simulations in an attempt to sample the full range of potential risks. Monte Carlo methods are often at the core of such systems as they allow for the accounting of uncertainties in the TC track, intensity, and surface wind field structure. Among these three key components, the modeling of the spatial variability in surface wind field estimates is clearly an area warranting further attention.

Given the computing cost associated with the generation of large ensembles, only simple wind formulations can be considered and a trade-off exists between the need to account for a wide range of physical processes shaping the wind field, and the level of complexity that can be practically implemented. Parametric wind models have flourished as a computationally efficient solution to this problem and the standard approach to date has been to model the surface wind field as the interaction between an axisymmetric stationary vortex (e.g., Holland 1980) and the forward motion of the storm (e.g., Georgiou 1985). This results in stronger winds occurring on the right (left) of the moving TC in the Northern (Southern) Hemisphere with the left–right wind ratio increasing as a direct function of the storm translational speed. Most parametric TC wind models are based around this simple motion-induced asymmetry.
Parametric approaches have gradually improved with the inclusion of model-specific coefficients that better represent complex features such as wind field asymmetries, radial decay, or the impact of surface drag on boundary layer winds. Yet the key assumptions remain, and the modeling of spatial variability is limited as a consequence. Although some level of uncertainty can be specified by varying model parameters, the range of structures that these models can cover is narrowly constrained by their core formulation. Recent studies have highlighted such limitations (Uhlhorn et al. 2014; Klotz and Jiang 2016).

Using airborne stepped frequency microwave radiometers (SFMR) to analyze wind fields from 35 hurricanes, Uhlhorn et al. (2014) stress the importance of the environmental flow on the location of the surface wind maximum, a result consistent with Ueno and Kunii (2009), Ueno and Bessho (2011), and Klotz and Jiang (2016). All studies identified the wind shear vector direction relative to the storm heading as a key indicator of asymmetric wind patterns, especially in strongly sheared environments with Uhlhorn et al. (2014) concluding that the horizontal surface wind field in TCs influenced by shear may deviate systematically from the expected motion-induced asymmetric structure.

TCs are typically embedded in strongly sheared environments while undergoing extratropical transition (Jones et al. 2003; Evans and Hart 2008). In the western North Pacific (WNP), where ~40% of TCs transition (Kitabatake 2011), important departures from the idealized MIA framework have been reported (Kitabatake and Fujibe 2009; Loridan et al. 2014, 2015) with up to two-thirds of cases exhibiting strong surface winds to the left-hand side of the forward motion. Departures for extratropical transitioning storms have also been observed in other basins with the wind field of Hurricane Sandy (2013) at landfall a recent example (see http://earthobservatory.nasa.gov/IOTD/view.php?id=79626).

Acknowledging these limitations, and as an alternative to current parametric wind modeling techniques, here we investigate the potential of machine learning (ML) methods, whose modern approaches fall within what Breiman (2001b) refers to as the “algorithmic culture.” The landmark Breiman (2001b) study contrasts two approaches to statistical modeling. The “data modeling culture” spans models that seek to replicate the available data from clearly stated rules specified by the modeler. The parametric wind models described above fall within this first category. The algorithmic culture treats the very complex processes responsible for data generation as a black box and assumes that data are better approximated by relaxing the need for clear and simple model rules. This is this second philosophy we wish to apply to the problem of modeling TC surface wind field uncertainty and as a consequence the models generated will not necessarily be fully interpretable. To confirm the validity of the approach without a full understanding of the rules on which the model is built, and as suggested by Breiman (2001b), a cross-validation procedure is needed in which the model is judged on its ability to simulate data not involved during the development phase.

With very fast runtimes ML algorithms can be efficiently deployed in risk assessment or probabilistic forecasting frameworks. For our purposes a class of models called “quantile regression” is prioritized as this allows modeling of the full conditional distribution of key wind field patterns and therefore provides a very practical way to better account for uncertainty in the TC wind field structure.

As ML techniques are rather new to this field, section 2 first introduces key elements of the methodology and discusses the concepts behind both the random forest algorithm and quantile regression. Section 3 follows with a simple example application of this framework for the WNP, while section 4 summarizes our findings and offers some directions for future development.

2. Methodology

ML algorithms are traditionally split in two groups—unsupervised and supervised learning—and our methodology involves both. Unsupervised learning algorithms identify patterns in the data (hence allowing for dimension reduction) but are not required to predict a predefined target (i.e., the model is not told what to look for). Supervised learning algorithms have a clear target that they attempt to predict.

The philosophy behind supervised learning is that, given a large amount of relevant data, predictive models can be trained without having to explicitly formulate any model rules. In the current context this implies no a priori assumptions about the TC wind field structure or asymmetries. Instead the foundation of the work is shifted toward gathering relevant data from which the model can learn and assimilate key spatial patterns.

As for most predictive modeling problems the first step is to clearly identify how the model is to be used. This includes identifying what the model should predict (model target) and what information is potentially available to make that prediction (input features). In the context of TC wind field modeling the target could be, depending on the application, either the scalar field of wind magnitude around the TC center or the
two-dimensional vector field comprising zonal and meridional wind components. Similarly, the variables available to predict the target field will vary from one application to the next.

In simple risk assessment models the wind magnitude is often used to assess building damage and the pool of input features might be limited to the storm trajectory, speed, intensity, and size. On the other hand if the model is to be used in real time to drive a surge and wave forecasting system, the two-dimensional wind vector field is required and a lot more input information about the state of the atmosphere surrounding the TC and the ocean surface might be needed. These application-dependent considerations are critical as they impact every step of the model building process.

In brief, our methodology is structured as follows:

1) Generate and process the wind model training data and extract potential predictors (see section 2a).
2) Perform a principal component analysis (PCA) on the training data (unsupervised learning). This produces a set of principal components (PCs) and corresponding weights (see section 2b) from which TC wind fields can be reconstructed.
3) Apply supervised learning algorithms to model the weights (see section 2c) and assess model predictions in a cross-validation exercise.
4) Enhance the wind field models using model chains and uncertainty sampling (see section 2d).

Further detail is presented in the following sections.

a. Data collection and processing

Collecting relevant data to train models on can be a challenging task and in the case of TC winds the potential sources are rather scarce. While TCs occur regularly in seven different basins worldwide, only in the North Atlantic (NA) and part of the eastern North Pacific has there been a prolonged effort to collect high-quality observations. The aircraft reconnaissance data available in these basins, combined with routine infrared satellite imagery and other in situ near-surface instruments punctually available have provided the basis for our current understanding of TC wind field structure (Landsea et al. 2004; Powell et al. 1998, 2009; Knaff et al. 2015). These would be the first choice training data sources for studies of the NA.

In other basins, including in the WNP where TCs occur with even larger frequency and higher peak intensities (Chavas and Emanuel 2010) than the NA, the quality of available data is not comparable. Aircraft reconnaissance in the WNP ceased in 1987 (Martin and Gray 1993) and, while satellite-based products such as QuikSCAT (Hoffman and Leidner 2005) or ASCAT (Gelsthorpe et al. 2000) have provided alternatives for surface wind studies, they rarely allow for a complete, reliable picture of the wind field unless thorough quality control procedures are applied (e.g., Stiles et al. 2014).

The Multiplatform Tropical Cyclone Surface Wind Analysis (MTCSWA), which offers the most complete and centralized view of TC winds globally, is another candidate for studies in the WNP but it has its own limitations. In their description of the MTCSWA product, Knaff et al. (2011) provide a good overview of the range of potential sources and their limitations in the context of modeling TC winds, highlighting the issues arising when high winds and heavy precipitation are present. Moreover, the MTCSWA also includes prior assumptions about the TC wind field based on MIA (see Mueller et al. 2006) and is therefore not ideally suited to our study.

As an alternative, we employ high-resolution numerical modeling to develop a training data source. Section 3 describes how the Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008) is used to generate high-resolution TC wind fields for the WNP. Since these are model realizations rather than observations, the data generated will likely suffer from some well-known limitations of the WRF model such as a low bias in storm intensities (Davis et al. 2010). Despite these limitations [see Loridan et al. (2015) for further details] and in the absence of wind field observations, the WRF modeled wind fields are considered the best training data source for our study.

Once the training data are assembled, variables with potential predictive skill are extracted. Relevant features will typically include track and intensity parameters, as well as broad characteristics of the wind field. More complex features characterizing the atmospheric flow can also be considered in the modeling process (e.g., wind shear, upper-level divergence, relative humidity) but their inclusion is conditional on their availability when the model gets deployed.

b. Unsupervised learning and dimension reduction

The number of input features and target variables to model can quickly become very large when spatially gridded fields are involved. Our training data comprise a series of wind field targets, each a snapshot of the wind field at a particular time. The targets are only fully described via knowledge of the wind strength for every grid point in the domain. Several methods exist to reduce the dimensionality of the problem and PCA is introduced here as an example.
Mathematically PCA involves identification of a set of orthogonal vectors (the PCs) allowing decomposition of the target field using a series of coordinates, or weights $W_i(t)$:

$$\text{TARGET}(t) = \text{MFLD} + \sum_{i=1}^{n_{pc}} W_i(t) \text{PC}_i,$$

where MFLD is the average or mean target field and $n_{pc}$ represents the total number of components equal to the number of grid points in the domain. With the PCs ordered by importance (based on the percentage of variance explained), this decomposition enables analysis of the key patterns of variability among the targets. Analysis of the leading PC fields can provide great insight into the potential physical mechanisms driving the variability as was shown in the Knaff et al. (2015) analysis of TC flight level winds.

More importantly, Eq. (1) can be truncated to keep only the leading vectors, providing a practical and computationally efficient way in which PCA can capture most of the variability while reducing the dimensionality of the dataset. The TC wind field models outlined in section 3 employ only the first three PCs and their corresponding weights.

Unsupervised learning algorithms can also help reduce the dimensions of input features when these are represented by gridded fields. In their ML modeling of the Indian summer monsoon rainfall, Saha et al. (2016) showed how another algorithm—a sparse auto-encoder—helped combine potential gridded predictors to enhance predictive skill while reducing their number.

c. Supervised learning and quantile regression

The next step is to select supervised learning algorithms that can learn to predict the TC wind fields reconstructed from the PC weights (the target) given a set of input features. A wide range of algorithms exist and depending on the nature of the target a regression model (when the target is a numeric field) or classification model (when the target is a categorical field) is used. As our objective is to model the PC weights for the wind field the focus here is on regression models.

A number of regression algorithms were tested but only variations of the same algorithm—random forests (Breiman 2001a)—are presented in section 3. The foundation of this method is the decision tree—an algorithm designed to recursively partition the training data into smaller groups that are more homogenous with respect to the target. This is done via a series of nested if–then statements at the nodes or branches of the tree, each time selecting a predictor variable to split the data and a threshold value for the split. The data in the terminal nodes (leaves) of the tree are then used to make predictions, for instance by taking the average of all target values in that subgroup or from simple linear regression.

Decision trees are easily interpretable and thus popular as explanatory models as they allow the modeler to fully follow the steps taken by the algorithm. In atmospheric science they have been used to gain an understanding of underlying physical mechanisms as was illustrated in the Zhang et al. (2013a, b) study of TC track modeling in the WNP. However, decision trees are also known to lack stability and small perturbations in the data can easily lead to very different tree architectures and/or splitting criteria. The accuracy of the predictions they produce are also easily outperformed by other (less interpretable) algorithms (Breiman 2001b).

The random forest algorithm (Breiman 2001a) helps overcome these issues by assembling a large number of decision trees under the following conditions: 1) each tree in the ensemble is given only a bagged version of the training data (i.e., a random subset of the data sampled with replacement), and 2) only a random subset of all input features are evaluated at each node in the trees. The final prediction from the forest is then obtained by averaging the predictions from each tree, greatly reducing the risk of overfitting the training data and improving both model stability and performance. The trade-off when compared to ordinary decision trees is a loss of transparency in the model rules and random forests are therefore better suited for applications where the goal is predictive accuracy rather than interpretability. As a consequence, and as mentioned in introduction, an objective assessment of the model behavior can only be gained from a cross-validation exercise. For that purpose the models in section 3c are repeatedly built using all but one storm while their predictions are evaluated on the left out, unseen storm (leave-one-out evaluation).

Another major strength of the random forest approach is that a portion of the data is left out when training each tree in the ensemble; these “out-of-bag” samples can be used to evaluate tree performance during the training process. In other words the entire dataset can be used for both model training and evaluation simultaneously. Furthermore, to obtain a picture of the importance of a given predictor, the algorithm records the reduction in performance that results from a random shuffle of that predictor’s values across the training records. An importance estimate can then be assessed from the mean reduction in prediction performance (cf. that obtained with the original
predictor) on the out-of-bag samples and averaged over all trees that use that predictor. This allows ranking of all input features in terms of their importance and provides a criteria to reduce the size of the list of inputs.

The Boruta algorithm (Kursa and Rudnicki 2010) (see section 3c) was built for this purpose: it is a wrapper around the random forest algorithm and allows identification of features that are not relevant by comparing their importance score to some “shadow features.” These shadow features are random and their importance scores are therefore only attributable to random fluctuations. Any input feature that does not register an importance score significantly higher than the maximum score from the shadow features is considered unimportant and can be omitted from the set of inputs.

As with most supervised learning methods, when a random forest makes a prediction for a new sample (i.e., unseen data) it is estimating the mean from the distribution of possible values to expect under the set of conditions described by the input features. This estimate is referred to as the conditional mean response. Ideally when assessing risk, a model should also provide information on the distribution of possible outcomes given the conditions; referred to as the conditional response distribution. Methods that directly predict specified quantiles in the distribution of the target variable have been developed to do this and their application to problems in environmental science is a natural fit (Baur et al. 2004; Friederichs and Hense 2007).

The common approach—quantile regression (Koenker and Hallock 2001)—allows the modeler to build linear regression algorithms targeting a specific quantile in the distribution of the response variable. By building multiple quantile regression models for a range of quantiles, one can then approximate the conditional response distribution. However, for this method to properly capture the distribution, multiple models need be built and since these are all constructed independently the method can suffer from what is termed “quantile crossing”: that is, when the prediction for a higher quantile is smaller than that for a lower quantile.

Alternatively, random forests offer a very powerful framework for quantile regression development as the original method can be generalized to approximate any conditional quantile for the response variable. A single model can be produced to predict all quantiles, thereby removing the risk of quantile crossing. In his description of the method, called quantile regression forest, Meinshausen (2006) summarizes the key difference with the traditional random forest algorithm by contrasting the way both methods treat the information in each node of the forest trees. Where random forests only record the mean of the training samples falling in a particular node, the quantile regression forest approach keeps the value of all observations in that node to allow assessment of the conditional distribution.

The quantile regression forest is the algorithm employed (see section 3) to predict the conditional distributions of the first three PC weights of the wind magnitude field ($W_1$, $W_2$, and $W_3$).

d. Model chain and uncertainty sampling

The information available to estimate $W_1$, $W_2$, and $W_3$ will vary depending on the scope and context in which the model is deployed. When first modeling $W_1$ at a given time $t$ [$W_1(t)$] along a TC track, a quantile regression forest can be trained using track, intensity and wind field structure parameters to approximate the conditional distribution of $W_1(t)$. Once $W_1(t)$ is determined it can be used in the modeling of $W_2(t)$ by adding it to the pool of input features to ensure that the

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chaba</td>
<td>2010</td>
</tr>
<tr>
<td>2</td>
<td>Megi</td>
<td>2010</td>
</tr>
<tr>
<td>3</td>
<td>Nida</td>
<td>2009</td>
</tr>
<tr>
<td>4</td>
<td>Melor</td>
<td>2009</td>
</tr>
<tr>
<td>5</td>
<td>Choi-Wan</td>
<td>2009</td>
</tr>
<tr>
<td>6</td>
<td>Kujira</td>
<td>2009</td>
</tr>
<tr>
<td>7</td>
<td>Jangmi</td>
<td>2008</td>
</tr>
<tr>
<td>8</td>
<td>Sinlaku</td>
<td>2008</td>
</tr>
<tr>
<td>9</td>
<td>Nakri</td>
<td>2008</td>
</tr>
<tr>
<td>10</td>
<td>Rammasun</td>
<td>2008</td>
</tr>
<tr>
<td>11</td>
<td>Krosa</td>
<td>2007</td>
</tr>
<tr>
<td>12</td>
<td>Wipha</td>
<td>2007</td>
</tr>
<tr>
<td>13</td>
<td>Nari</td>
<td>2007</td>
</tr>
<tr>
<td>14</td>
<td>Sepat</td>
<td>2007</td>
</tr>
<tr>
<td>15</td>
<td>Usagi</td>
<td>2007</td>
</tr>
<tr>
<td>16</td>
<td>Man-Yi</td>
<td>2007</td>
</tr>
<tr>
<td>17</td>
<td>Fitow</td>
<td>2007</td>
</tr>
<tr>
<td>18</td>
<td>Soulik</td>
<td>2006</td>
</tr>
<tr>
<td>19</td>
<td>Shan-Shan</td>
<td>2006</td>
</tr>
<tr>
<td>20</td>
<td>Ioke</td>
<td>2006</td>
</tr>
<tr>
<td>21</td>
<td>Saomai</td>
<td>2006</td>
</tr>
<tr>
<td>22</td>
<td>Nabi</td>
<td>2005</td>
</tr>
<tr>
<td>23</td>
<td>Sonca</td>
<td>2005</td>
</tr>
<tr>
<td>24</td>
<td>Mawar</td>
<td>2005</td>
</tr>
<tr>
<td>25</td>
<td>Sudal</td>
<td>2004</td>
</tr>
<tr>
<td>26</td>
<td>Megi</td>
<td>2004</td>
</tr>
<tr>
<td>27</td>
<td>Chaba</td>
<td>2004</td>
</tr>
<tr>
<td>28</td>
<td>Songda</td>
<td>2004</td>
</tr>
<tr>
<td>29</td>
<td>Tokage</td>
<td>2004</td>
</tr>
<tr>
<td>30</td>
<td>Stella</td>
<td>1998</td>
</tr>
</tbody>
</table>
predictions are not independent. The distribution of $W_2(t)$ is then conditional on the initial pool of inputs and on $W_1(t)$. The same applies for the modeling of $W_3(t)$ by adding $W_2(t)$ as a predictor during training.

If models are built in such a way the predictions at time $t$ can be generated as follows:

1) Randomly sample a quantile value and determine the corresponding $W_1(t)$ value conditional on input features at time $t$.

2) Randomly sample a quantile value for $W_2(t)$ and determine $W_2(t)$ conditional on all input features involved in step 1 plus the $W_1(t)$ prediction.

3) Randomly sample a quantile value for $W_3(t)$ and determine $W_3(t)$ conditional on all input features involved in step 2 plus $W_2(t)$.

4) Construct a wind field at time $t$ according to Eq. (1) ($n_{pc} = 3$).

The process above can be repeated for a large number of iterations to sample the space of $W_i(t)$ predictions and gain an understanding of the range of wind fields expected for a given TC. This framework could be used, for example, to enhance national TC agency advisory forecasts as it enables a probabilistic prediction of the spatial extent of the wind field to be made. 

Section 3c
illustrates how probabilistic maps for the risk of reaching given wind speed thresholds can be generated from repeated sampling of the modeled conditional distributions as per steps 1–4 above.

For applications where wind field predictions are needed for multiple time steps along a TC track (e.g., for probabilistic wind field forecasts beyond time $t$ or in risk assessment models), the steps listed above can be modified to account for time dependency: at time step $t_1$, $W_1(t)$, $W_2(t)$, and $W_3(t)$ can be added to the pool of predictors in step 1. Including previous predictions ensures that a prediction at $t_1$ is conditional on the state of the wind field at time $t$.

Similarly, once values of $W_1(t_1)$, $W_2(t_1)$, and $W_3(t_1)$ have been sampled they can be used as inputs to the model at time step $t_2$ to generate the next predictions. This forward feeding of the predictions can be repeated throughout the lifetime of the TC to model the wind field at all track points. Weights prior to the previous time step could also be included as predictors; however, adding too many levels of autoregression limits the training data as only targets with several previous steps will be eligible for model building. The model chain with one level of autoregression is used in the analysis below.

### 3. A TC wind field model for the western North Pacific

The models described below employ the concepts presented in section 2 to predict the shape of the surface wind speed field. In particular the models are designed to produce a distribution of potential wind field patterns under a given set of conditions, therefore accounting for uncertainty around the expected outcome.

#### a. WRF Model runs (data collection)

A total of 30 historical WNP TCs were simulated (Table 1) using the WRF Model to create a database of wind field targets characteristic of wind field variability in the basin. The TCs were selected to cover a range of track trajectories, time of year, and environmental conditions. Note, however, that no attempt was made to exactly match the historical events, rather the intention is to simply simulate physically realistic TCs in a similar environment as those observed (see appendix A for model run and WRF modeling framework details). Moreover, in order to homogenize the analysis of wind field variability, only overwater time steps were retained.

Figure 1 provides an overview of the climatology of key input features (Table 2) and highlights the variety of events selected. The WRF Model snapshots analyzed across all events are mostly at latitudes of 15°–35°N (Fig. 1a) and longitudes of 125°–145°E, with the exception of one track that occurred much farther east (Fig. 1c). The bimodal partitioning of the heading angle distribution (Fig. 1g) shows the two most common directions are either northeastern ($H_a \approx 45°$, mathematical convention, see Table 2) or northwestern ($H_a \approx 135°$). In terms of intensity, the vast majority of our dataset consists of Saffir–Simpson hurricane wind scale category 2–3.
storms with only one case bordering on category 4 status (Figs. 1i,k).

Analysis of the remaining parameters shows that the events are most commonly characterized by translational speeds of 18–20 km h\(^{-1}\) (Fig. 1e), radius of maximum winds of 30–40 km (Fig. 1m), and maximum winds located at roughly 90° to the right of the storm heading (Fig. 1o). There is, however, a significant proportion of targets for which the maximum winds are located to the left of the track (i.e., \(A_m < -180°\); Fig. 1o) or along the direction of storm heading (\(A_m \approx 0°\) or \(A_m \approx -180°\)).

The hourly 10-m horizontal wind components are extracted from the WRF Model runs (4-km horizontal resolution) and a catalog of 1580 gridded fields of wind magnitude is assembled. These instantaneous wind speed estimates are taken to be representative of a 1-min sustained wind at 10 m (Davis et al. 2010) and the wind output from the ML models assume this averaging period and height.

To homogenize the dataset, the wind fields are rotated and interpolated into a standard reference grid (see Figs. 2b,d) with dimensions 151 × 151 (i.e., 75 points in each direction starting from the storm center). The convention adopted is to rotate the wind magnitude field so that the wind maximum (\(V_m\)) ends up at point \((X = 0, Y = 25)\) and this gives rise to a target dependent grid resolution of \(R_m/25\) covering a \(6R_m \times 6R_m\) domain.

Figure 2 provides examples of wind targets before (Figs. 2a,c) and after (Figs. 2b,d) rotation and interpolation.
for two very different cases. The wind field from target 310 (Figs. 2a,b) follows a rather typical pattern for a fast-moving TC with a compact area of strong winds and the maximum wind speed located at right angles, to the right of the storm heading direction. Target 11 (Figs. 2c,d) represents a recurving storm at the beginning of the transitioning phase and by contrast has strong winds present on both sides of the track, with the maximum velocity to the left of the storm heading.

As an illustration of the traditional parametric modeling approach, the Kepert (2001) model is implemented with parameter values characteristic of target 310 (Figs. 2a,b). The key model inputs extracted from the WRF simulations are as follows: $P_0 = 1008$ mb, $P_m = 955$ mb, $V_m = 47.5$ m s$^{-1}$, $R_m = 41$ km, $T_s = 35.5$ km h$^{-1}$, and a latitude of 19.85°N, while Holland B is set to 1.3. The reader is referred to appendix B for details regarding implementation, which closely follows that of Khare et al. (2009). The resulting wind field (Fig. 3) is a rather typical example of the type of wind patterns that can be produced from MIA parametric formulations, and although some of the initial shortcomings of the Holland (1980) method have been addressed over the years, the general framework has mostly been conserved (i.e., MIA assumption).

For a case like target 11 (Fig. 2c), the MIA assumption is directly violated as the maximum winds are located to the left of the moving TC, therefore spinning in the opposite direction of storm motion (Northern Hemisphere). Any attempt to model a stationary vortex and then vectorially add the forward motion of the TC would result in stronger winds to the right of the track. The modeler is therefore left with two options: either to try to capture the maximum value to the left by artificially increasing the right-hand side maximum, or focus on the right-hand side maximum alone and neglect the stronger winds to the left of the track. Both cases lead to a strong misrepresentation of the wind speed field and therefore of any concomitant risks associated with it. One of the key aims of the current study is to introduce a method capable of dealing with such real wind fields.

Perhaps it is also important to clarify that the Fig. 3 wind field should not be seen as an attempt to perfectly simulate target 310. Instead it represents the mean wind patterns that would result if averaging the wind fields from a large number of cases where input parameters have the same values (conditional mean response). Therefore, and given the extent of the processes not modeled by parametric formulations, there is no reason to expect the Fig. 3 estimate to match the details of target 310. Another key aim of the current study is to characterize not only the mean but the full response distribution of wind fields to expect given a set of input parameter values.

As the focus of our study is the shape of the wind field, and since the modeling of TC intensity is achieved via another component of the system, each target field is normalized by $V_m$. The resulting training targets are therefore bound between 0 and 1 and Fig. 4a presents the mean normalized wind field over the entire training set [our MFLD in Eq. (1)].

b. Dimension reduction

The normalized wind targets are decomposed using PCA and our focus is on the first three PCs (Figs. 4b–d), which collectively explain over two-thirds of the variability in the data. The first PC (Fig. 4b) exhibits a pattern that can modulate the radial wind decay in the far field, while the second (Fig. 4c) and third (Fig. 4d) PCs represent wavenumber-1-type asymmetries at radii beyond $R_m$. At radii close to $R_m$, these three PCs offer a range of asymmetry patterns allowing for rotation of the regions of strong winds while also modulating the decay inside the eye of the TC.

It is also worth noting that the value of all three PCs at the point $(X = 0, Y = 25)$ is exactly 0 due to the convention selected for the reference grid. This means that they will not alter the value of $V_m$ that is input, which is important for risk assessment and probabilistic forecasting applications as it ensures $V_m$ will be perfectly replicated at the specified location ($R_m, A_m$).

The distributions for the associated weights [see Eq. (1)] for the 1580 targets are shown in Fig. 5. We remind the reader that the objective of the supervised learning exercise discussed in section 3c is to train quantile regression forest models to predict the distributions of the three weights, conditional on a set of input features.

Fig. 3. Wind field from Kepert (2001) parametric model: $P_0 = 955$ mb, $V_m = 47.5$ m s$^{-1}$, $R_m = 41$ km, $T_s = 35.5$ km h$^{-1}$. See appendix B for details.
c. Conditional distribution models

1) PARAMETER IMPORTANCE

The Boruta algorithm is used to assess the importance of a range of input features that includes the following: the variables in Fig. 1; PC weights at time $t$ [W$_1(t)$, W$_2(t)$, and W$_3(t)$]; and PC weights 6 h before the time of prediction, W$_1(t-1)$, W$_2(t-1)$, and W$_3(t-1)$. Table 3 shows the output from the Boruta analysis where all predictors were considered “important” for all three weights as they all score significantly higher than the set of random shadow features.

Table 3 shows that the most important input features when predicting W$_1(t)$ are $R_m$, $dR_m$, $A_m$, and the value of W$_1$ itself 6 h before time $t$ [W$_1(t-1)$]; whereas W$_3(t-1)$, $dT_s$, and $dH_A$ are the least important. For W$_2(t)$, $R_m$ and $A_m$ again rank very high along with W$_2(t-1)$ and $P_o$. Interestingly W$_1(t)$ ranks fifth suggesting a level of dependency between W$_2(t)$ and W$_1(t)$. The most important parameter in the prediction of W$_1(t)$, $dR_m$, is much less important to the prediction of W$_2(t)$, as is W$_1(t-1)$. The 6-hourly change in storm speed ($dT_s$) once again ranks low.

The highest importance score for any predictor across all three weights is produced by $A_m$ in the prediction of the third PC weight. The next most important input feature in the prediction of W$_3(t)$ is the 6-hourly change in the same variable ($dA_m$). Similar to the first two PC weights the results for W$_3(t)$ show a level of dependency between weights with W$_3(t-1)$, W$_1(t)$, and W$_2(t)$ all ranking high as predictors. This confirms the importance of modeling dependency between weights (model chain). Once again $dT_s$ ranks very low and could likely be dropped without any important deterioration in model performance.

To illustrate the value of quantile regression forests a first model is built for W$_1(t)$ using the entire training set

![Fig. 4. (a) Normalized wind field averaged over the 1580 WRF Model targets along with the (b) first, (c) second, and (d) third principal components for the normalized wind magnitude field.](image-url)
(30 storms) and only the 16 input features from Fig. 1. All quantile regression forest model parameters are set to their default values (see Meinshausen 2006). The model then predicts the $W_1(t)$ distribution with all 16 input features set to their mean values. To test the sensitivity of the model two additional model predictions are run with $R_m$ moved to the 25th quantile of all values ($R_m = 20 \text{ km}$) and 75th quantile ($R_m = 142 \text{ km}$).

The predicted cumulative distribution functions (CDFs) are shown in Fig. 6. For the average case (black dots) the model predicts a close to 50% chance of $W_1(t)$ being negative, while for a $R_m$ of 142 km (red dots) this falls to around 15% and for $R_m = 20 \text{ km}$ it rises to $\sim 75\%$. For larger storms the probability of $W_1(t)$ being positive is therefore greater which, when referring to Fig. 4b and Eq. (1), indicates a reduction (relative to the mean profile of Fig. 4a) of mean wind speeds in the far field and an increase within the eye. This is reversed for smaller storms.

Interestingly the impact of the change in $R_m$ is not limited to a simple translation of the CDF for the average case, illustrating the value of simulating the entire response distribution with a quantile regression forest.

Although sensitivity analysis of the kind described above can elucidate some aspects of model behavior and allow some interpretation of the physical mechanisms involved, in general random forest-based methods tend to behave as black box models. And as mentioned in the introduction, confidence in such types of methods comes from careful cross-validation rather than from a complete understanding of model rules. We describe this process below.

### 2) MODELS FOR PREDICTION AT TIME $t$

We next build models for the conditional distributions of $W_1(t)$, $W_2(t)$, and $W_3(t)$ using the 16 input features in Fig. 1 and the model chain approach described in section 2c [i.e., the model for $W_2(t)$ will include $W_1(t)$ as a predictor while the $W_3(t)$ model will have both $W_1(t)$ and $W_2(t)$ as predictors]. The $W_1(t - 1)$, $W_2(t - 1)$, and $W_3(t - 1)$ are omitted here as we first assume wind field predictions are not required for multiple time steps along a TC track.

A “leave-one-out” procedure is adopted (cross validation) where the three models are repeatedly trained using 29 of the historical events simulated (the training set) and then deployed to predict the left-out storm (the testing set whose data are unseen during model building). Although the PCA results barely change when leaving one storm out of the analysis, for consistency the decomposition is repeated at each step using data from the training set only. The wind targets from the testing set are then projected onto the PCs to obtain $W_1(t)$, $W_2(t)$, and $W_3(t)$ for all time steps in the testing set.

As was the case in the previous section the quantile regression forest algorithm is used for all three models and all model parameters are set to their default. For each target from the left-out storm the models can
predict any quantile in the conditional distribution of the weights. Figure 7 shows the time series of hourly weight values for one of the left-out storms (red lines) along with model predictions for the 0.05, 0.25, 0.5, 0.75, and 0.95 quantiles. Each light gray shaded bar is the prediction for a single target and it extends from the 0.05 to the 0.95 quantile (i.e., it shows the 90% confidence interval). In other words, according to the model and under the conditions specified by all input features, 90% of the observations should fall in that range.

In a similar vein, the darker shaded bars show the 50% confidence interval while the black lines indicate the median model predictions. A first observation is that the range of these confidence intervals varies across the targets (i.e., some targets have a higher level of uncertainty than others). Figure 7 also indicates that the models perform well by accounting for most of the fluctuations in the observed weight values—the shift in the location of the median prediction and confidence intervals is generally in the right direction.

By repeating the process for all left-out storms we can assess the quality of the model predictions for the full dataset. Figure 8 shows all median predictions (black dots) sorted in increasing order along with the observed weights (red crosses) and the 90% and 50% confidence intervals. For the $W_1(t)$ model, 92.3% of the observed weights fall within the 90% interval (CI90; Fig. 8) while 55.3% are contained in the 50% interval (CI50). For $W_2(t)$ and $W_3(t)$ the 90% interval covers 89.1% and 91.3% of the observed values while 50.3% and 54.2%, respectively, fall within the 50% interval. These results are gratifying and confirm the ability of the models to predict the range of PC weights with confidence intervals consistent with the distribution of observed outcomes.

With the three models capturing the variability in observed weights, the framework described in section 2d provides a way to account for uncertainty when producing a wind field forecast at a given time $t$. By repeating steps 1–4 in section 2d for a large number of iterations, a probabilistic view of the spatial distribution of risk can be compiled. Figure 9 shows examples (using the same two targets in Fig. 2) of the approach when simulating the likelihood of reaching the category 1 wind speed threshold ($33 \text{ m s}^{-1}$), and experiencing wind speeds of at least 90% of $V_m$, with 1000 iterations. For each iteration, the four steps are executed and each modeled field is multiplied by $V_m$. The cells reaching the category 1 and 0.9 $V_m$ thresholds are recorded. The probability of reaching these thresholds can then be approximated by dividing the total counts for each cell by the number of iterations (1000).

A clear difference exists in the probability maps produced for the two targets (Fig. 9). Although $V_m$ for target 310 (47.5 m s$^{-1}$) is higher than for target 11 (40.5 m s$^{-1}$), the probability of reaching the category 1 threshold is higher in the $Y < 0$ section of the domain for target 11 (cf. Figs. 9a and 9c) (i.e., the likelihood of having strong winds on both sides of the storm is larger).

A similar conclusion can be drawn from Figs. 9b and 9d, which show the likelihood of exceeding the 0.9 $V_m$ thresholds. For target 310, this area is limited to the neighborhood around the location of $V_m$ while the prediction for target 11 shows that the chance of observing wind speeds of at least 0.9 $V_m$ is around 35% for some cells on the opposite side of the $V_m$ location.

These examples illustrate the value of including uncertainty in a wind field forecast: although target 310 is more threatening in terms of maximum wind magnitude, the target 11 wind field has the potential to cover a much broader area with wind speeds that are still significant relative to what would be the forecast peak value ($V_m$).

In both examples, a very steep gradient in probability occurs inside the eye of the TC, especially along the $X = 0$ line. Analysis of the PC fields (Fig. 4) shows that none of the three PCs captures much variation for $X = 0$ and $0 < Y < R_m$. To better represent model uncertainty in this area additional PCs will need to be added to the model. Note, however, that for the purpose of risk
assessment, the weaker winds that exist in the eye of the TC would not generally be a priority.

3) MODELS FOR PREDICTION BEYOND TIME t

Having built models for $W_1(t)$, $W_2(t)$, and $W_3(t)$ at time $t$ the focus now shifts toward prediction for multiple time steps. As outlined in section 2d, here the predictions made for the three PC weights at a given time step become input features to the models for the following time step. Three new quantile regression forest models are trained using the 16 input features in Fig. 1, the model chain approach described in section 2d, and including $W_1(t-1)$, $W_2(t-1)$, and $W_3(t-1)$.

Targets 310 and 11 are again used as examples and Fig. 10 provides a 24-h forecast using our 3 new models. The forecast for each target is produced by first sampling three quantiles that are then used to determine $W_1(t)$, $W_2(t)$, and $W_3(t)$ using the models from section 3c(2). These three predictions are then fed into the new models along with the values of the input features at time step.

![Time series of (top to bottom) hourly $W_1$-$W_3$ values for one of the left-out TC case (red × symbols and line) along with the median out of sample model predictions (black), as well as the 50% (dark gray) and 90% (light gray) confidence intervals.](image-url)
and with all updated values for the predictors at time \( t + 1 \). Here again three quantiles from the modeled distributions can be sampled to produce the \( W_1(t + 1) \), \( W_2(t + 1) \), and \( W_3(t + 1) \) estimates. This wind field forecast at time \( t + 1 \) is by construction dependent on the wind field produced for time \( t \) since the predictions were made conditional on the \( W_1(t) \), \( W_2(t) \), and \( W_3(t) \) predictions. The process can then be repeated forward in time for any number of steps where predictor values are available. The two time series forecasts in Fig. 10 represent only one possible simulation path for each target. Although one could search for the combinations of quantiles that best replicate the WRF Model output for each target, this is not the purpose of the exercise and the simulation paths shown were selected to only broadly mimic the WRF-simulated patterns. Beyond the analysis of model behavior for these individual targets, what the comparison of the two cases really shows is the ability of our approach to cover a wide range of realistic wind field shapes under a single modeling framework. In particular the model is able to cope with cases like target 11 where strong winds occur on both sides of the TC.

In the context of a risk assessment framework the types of quantile regression forest models presented here would greatly increase the range of wind field patterns simulated and therefore provide a much broader view of the risk associated with TC winds.

4. Discussion and future steps

In an attempt to better characterize uncertainty in tropical cyclone wind field modeling the potential of machine learning methods was explored as a substitute for traditional parametric models. In particular, quantile regression modeling approaches were employed as these have the virtue of allowing simulation of the full conditional distribution of a response variable. This approach was implemented to simulate the conditional distributions of the weights associated with the first three PCs derived from a set of high-resolution TC wind field simulations. The list of input features (predictors)
includes commonly used TC track and intensity parameters along with indicators of the location of maximum winds. Autocorrelation and dependency between the three PC weights was also accounted for by including the predicted weights, when relevant, in the list of predictors.

A cross-validation exercise confirmed the ability of this modeling framework to reproduce the range of PC weight values contained in the training set with good agreement between the predicted model wind speed confidence intervals and the distribution of observed outcomes. The model, which can simulate a wide variety of wind field patterns, also allows easy generation of spatial wind speed probability maps through repeated sampling of the modeled conditional distributions.

Incorporating a wind field model of the type presented here in a probabilistic TC forecasting system such as that described by De Maria et al. (2009, 2013) would ensure the uncertainty in the spatial distribution of wind speeds is better accounted for. Similarly, when deployed as part of a risk assessment framework (e.g., Vickery et al. 2009), our approach is able to generate a much wider and more realistic range of TC wind structures than is possible from current parametric formulations. In both applications our method would lead to a more complete representation of the risk profile.

Given enough data the influence of more complex atmospheric and oceanic features could also be incorporated in the models. A strong candidate for inclusion is the environment shear given the link observed between it and surface wind field asymmetries.

Finally, the method can easily be applied to other perils and further research could include the modeling of TC-induced precipitation. Using a similar modeling

---

**FIG. 9.** Probability maps for the risk of reaching the category 1 wind threshold (33 m s$^{-1}$) under the conditions leading to (a) target 310 and (c) target 11. (b),(d) As in (a),(c), but for the probability of reaching winds of at least 0.9$V_m$. 

AUGUST 2017 LORIDAN ET AL. 3217
chain to the one introduced here, the dependency of the rainfall field on the wind field pattern can be represented by including the PC weights for the wind field as predictors.

Acknowledgments. The authors thank UCAR/NCAR/MMM for provision of the WRF Model, NCEP for provision of the CFSR reanalysis, and JMA/IBTrACS for the track data. This research was funded by Risk Frontiers and undertaken with the assistance of resources from the National Computational Infrastructure (NCI), which is supported by the Australian government.

APPENDIX A

High-Resolution WRF Simulations

The initial and boundary conditions for all WRF simulations are taken from the Climate Forecast System Reanalysis product (CFSR, 0.5° resolution; Saha et al. 2010) and three nested domains are implemented with horizontal resolutions of 36, 12, and 4 km, using 36 vertical levels. The physics options selected are as in Loridan et al. (2015), including the Yonsei University planetary boundary layer scheme (Hong et al. 2006), Kain–Fritsch cumulus parameterization (Kain and Fritsch 1990), Rapid Radiative Transfer Model long-wave radiation model (Mlawer et al. 1997), Dudhia shortwave radiation scheme (Dudhia 1989), and WRF single-moment 5-class microphysics scheme (Hong et al. 2004).

The objective of the simulations is to model physically realistic, high-resolution TC winds under the same atmospheric conditions as was observed during the event but not necessarily to perfectly replicate the historical track and magnitude of each case selected. However, as the 30 cases (Table 1) are selected to create a diverse dataset, there is a need to ensure each TC does not depart too much from its historical occurrence. For this purpose analysis nudging is applied to the parent domain while the track central pressure values reported by the Japan Meteorological Agency are nudged on both nested domains to keep the TC center in the neighborhood of the observed track trajectory. As the pressure nudging would not be feasible with vortex following nests, fixed domains are used and their definition varies from case to case to match track trajectories.

The cases are typically simulated by the WRF Model for 6–7 days with the first 48 h omitted from the analysis as this represents the model spinup period. Hourly outputs from the finer domain (4-km resolution) are stored for further processing. The 10-m horizontal wind components, surface pressure field, and land mask are extracted from the WRF Model hourly outputs stored from the finer 4-km domain. For a given hourly wind output the center of the storm is first identified from a search of the grid point with the minimum surface pressure in a radius of 3° around the location of the maximum 10-m winds. An additional step is then used to identify a second storm center as the location of the minimum 10-m wind speed within a radius of 0.5° around the first center. For cases where these searches return two different locations the location of the minimum...
wind (second storm center) is deemed to be the center as the focus here is on the wind vortex structure.

Note several filtering criteria enter the process described above; to remove any impact of land surface roughness, the land mask is used to discard targets with winds over land. Furthermore, and by definition of the standard grid (section 3a), only the targets for which the $6R_{\text{max}} \times 6R_{\text{max}}$ grid can fit in the WRF Model domain are kept.

APPENDIX B

Implementation of the Kepert (2001) Model

A wide range of parametric surface wind field models exist and as an example we here introduce the linear analytical boundary layer vortex model derived by Kepert (2001). This formulation is able to represent the full three-dimensional wind field and accounts for the impact of surface roughness on the boundary layer flow via surface drag and vertical diffusivity, making it one of the most complete parametric formulations developed to date.

Starting from the horizontal momentum equations for a steady-state vortex, with constant vertical turbulent diffusivity ($K$), and in storm-centered cylindrical coordinates moving with the TC, Kepert (2001) derives analytical solutions for the radial ($u_r$) and azimuthal ($u_\theta$) storm-relative wind components at a given height $z$. The key assumptions include the following: a linearization of the equations discarding second-order terms, a semislip condition at the surface that involves a drag coefficient ($C_d$), the latitude of the storm center, the storm translational speed ($T$), radius of maximum winds ($R_m$), the drag coefficient $C_d$, and the turbulent diffusivity $K$. The two parameters $A$ and $B$ are linked via the radius of maximum winds: $A = R_m^6$.

Since the modeled wind field is representative of the top of the surface layer it needs to be scaled to a 10-m height for comparison. Khare et al. (2009) tested several combinations of the parameters $C_d$ and $K$, and then applied a further empirical scaling to adjust the winds to a 10-m height—1-min average estimate. Based on their analysis the following settings are selected to illustrate the behavior of parametric wind field formulations: $z = 0$, $K = 70 \text{ m}^2\text{s}^{-1}$, $C_d = 0.001$, and $\alpha = 0.95$ (scaling parameter). Although parameter tuning is not the purpose of the current exercise, we note that the performance of this model with regards to the Fig. 3 case could likely be improved with a different setting.

REFERENCES


