Simulation of Atlantic Hurricane Tracks and Features: A Coupled Machine Learning Approach

RIKHI BOSE\textsuperscript{a}, ADAM L. PINTAR,\textsuperscript{b} and EMIL SIMIU\textsuperscript{a}

\textsuperscript{a} Materials and Structural Systems Division, Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland

\textsuperscript{b} Statistical Engineering Division, Information Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland

(Manuscript received 3 August 2022, in final form 31 January 2023, accepted 22 February 2023)

ABSTRACT: The objective of this paper is to employ machine learning (ML) and deep learning (DL) techniques to obtain, from input data (storm features) available in or derived from the HURDAT2 database, models capable of simulating important hurricane properties (e.g., landfall location and wind speed) consistent with historical records. In pursuit of this objective, a trajectory model providing the storm center in terms of longitude and latitude and intensity models providing the central pressure and maximum 1-min wind speed at 10-m elevation were created. The trajectory and intensity models are coupled and must be advanced together, 6 h at a time, as the features that serve as inputs to the models at any given step depend on predictions at the previous time steps. Once a synthetic storm database is generated, properties of interest, such as the frequencies of large wind speeds, may be extracted from any part of the simulation domain. The coupling of the trajectory and intensity models obviates the need for an intensity decay model inland of the coastline. Prediction results are compared with historical data, and the efficacy of the storm simulation models is evaluated at four sites: New Orleans, Louisiana; Miami, Florida; Cape Hatteras, North Carolina; and Boston, Massachusetts.

KEYWORDS: Statistical forecasting; Artificial intelligence; Decision trees; Deep learning; Machine learning; Neural networks

1. Introduction

Early estimates of landfalling storm wind speeds with specified return periods were based on (i) probabilistic models fitted to historical records of translation velocities, radii of maximum wind speeds, and central pressure deficits on storms arriving within a chosen distance of the target location; (ii) Monte Carlo simulations of those features; and (iii) physical models of the wind speeds of interest as functions of the simulated features (Russell 1971; Batts et al. 1980; Georgiou et al. 1983; Neumann 1987). The main weaknesses of these approaches were the estimation of tail probabilities from small datasets, the quality of the datasets, and the physical models available in the 1970s and 1980s.

Darling (1991) proposed a method to resolve this issue by introducing the relative storm intensity obtained by scaling the actual storm intensity by the potential intensity or the maximum possible storm intensity that mean seasonal climatic conditions would allow. This method is inadequate in regions where the potential intensity is small or zero. Vickery et al. (2000, 2009) simulated large numbers of synthetic storms (corresponding to periods of e.g., \( \sim20000 \) years) to generate wind maps at chosen mileposts in the U.S. coastline. Storm locations and intensities at 6-h intervals were predicted using linear models; the storm trajectory model depended on storm latitude, longitude, translation speed, and direction at two previous 6-h instants, while the (relative) intensity model was a function of the storm intensity and the sea surface temperature at two and three previous time instants, respectively. The constant coefficients of the models were obtained in chosen grids discretizing the 2D latitude–longitude domain. Synthetic storm descriptions included radial profiles of storms using the empirical profiles proposed by Holland (1980).

Emanuel et al. (2006) also generated large numbers of storms using two statistical models for storm trajectory and a deterministic model for storm intensity. One of the trajectory models utilized a Markov chain for each 6-h displacement, which accounted for the 6-h rate of change of direction of storm travel at the current location and time based on local climatological conditions. The second trajectory model also included the effect of the vertical wind shear, which also affects the intensity of a storm, and therefore, implicitly couples the trajectory and intensity models. Observed monthly statistics of climatological conditions were preserved while generating the synthetic tracks. Most importantly, instead of a statistical model, for the first time, a simple physical model for storm intensity coupled with an ocean model was used (Emanuel 2004). The intensity model intrinsically reduces a storm’s intensity at/after landfall.

The purpose of this paper is to present a machine and deep learning (ML–DL)-based methodology for the simulation of hurricane tracks and features. While such methods have been used for forecasting purposes, their application to the simulation of hurricane tracks and features is, to our knowledge, new. The success of ML–DL methods may be attributed to their superior ability to exploit large amounts of data to uncover complex relationships between feature and target variables. In the present work, input features are taken or derived from the HURDAT2 database (Landsea and Franklin 2013).
Table 1. Input features to the ML and DL models for trajectory, and corresponding model outputs at 6-h intervals. Here, $i$ denotes the $i$th time instant.

<table>
<thead>
<tr>
<th>Model (abbreviation)</th>
<th>Architecture</th>
<th>Database</th>
<th>Input features</th>
<th>Output variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trajectory–DL (T–DL–1)</td>
<td>LSTM–RNN</td>
<td>DB-1</td>
<td>$\lambda^{(i-1)}$, $\phi^{(i-1)}$, $v^{(i-1)}$, $\theta^{(i-1)}$, $w_m^{(i-1)}$, $p_k^{(i-1)}$</td>
<td>$\Delta \lambda^{(i)}, \Delta \phi^{(i)}$</td>
</tr>
<tr>
<td>Trajectory–DL (T–DL–2)</td>
<td>LSTM–RNN</td>
<td>DB-1</td>
<td>$\lambda^{(i-2)}$, $\phi^{(i-2)}$, $v^{(i-2)}$, $\theta^{(i-2)}$, $w_m^{(i-2)}$, $p_k^{(i-2)}$</td>
<td>$\Delta \lambda^{(i)}, \Delta \phi^{(i-1)}, \Delta \phi^{(i)}$</td>
</tr>
</tbody>
</table>

Developed before the recent reanalysis of Hurricane Camille (Kieper et al. 2016). We propose a combined ML–DL approach for the efficient simulation of storm trajectories and intensities over multiple 6-h time intervals. The approach uses separate models for storm trajectory and intensity. The trajectory model provides the storm center location. A joint model for predicting the central pressure and maximum 1-min wind speed at 10-m elevation constitutes the intensity model. Both the trajectory and storm intensity model must be advanced together, 6 h at a time, as the output features from one time step serve as input features for the next. The proposed approach has several advantages:

1) Unlike in many of the previous works, the trajectory and intensity models are coupled, which is physically more realistic.

2) Decay of a storm’s intensity on land is naturally acquired by ML models during training, which obviates the need to model decay by either statistical or analytical means, such as in Kaplan and DeMaria (1995), for example.

3) The accuracy of the DL trajectory models used herein is comparable to the accuracy of ensemble models used by the National Hurricane Center (NHC) for track forecasting, in spite of requiring significantly less input data (Bose et al. 2022).

4) As demonstrated herein, the ML–DL models are capable of emulating historical storm patterns at small (city) and large (Atlantic basin) spatial scales.

5) Once a synthetic storm database is generated, properties of interest, such as the probability that the wind speed exceeds 45 m s$^{-1}$ (~87.5 kt), may be directly extracted at any location in the simulation domain.

Simulating the evolution of dynamical systems is in some sense comparable to natural language processing. Recurrent neural networks (RNNs) were developed for this purpose. This work uses a class of RNN, the long short-term memory (LSTM) models with many-to-many prediction architecture developed for time marching a storm’s trajectory, and tested for hundreds of validation and test storms by Bose et al. (2021). Mean 6- and 12-h forecast errors were ~30 and 66 km, respectively. The 2/3 probability circle radii, a statistical measure used for characterizing errors incurred in trajectory forecasting, were found to be comparable to the state-of-the-art ensemble models currently in use for short-term storm trajectory forecasting. Random forest (RF) models are used herein to model storm intensities, on account of their requiring smaller volumes of data than neural networks, being relatively insensitive to outliers, having moderate variance and low bias, and having the advantage of interpretability (Hastie et al. 2009). RNN and RF models were utilized for their ability to automatically identify and exploit complex nonlinear relationships without explicit feature engineering, as would be the case with linear models, for example.

The storm simulation model may be used for meteorological purposes, for the estimation of extreme hurricane wind speeds with return periods specified by building codes, or for catastrophe risk modeling. For civil engineering structures, the return periods are typically 1000–3000 years. At a given location, the size of the available sample of hurricane wind speeds is typically too small to allow the reliable estimation of wind speeds with such long return periods. For this reason, HURDAT2 may be utilized to develop a simulation model, and then any desired number of storms may be simulated and used in conjunction with extreme value models to estimate wind speeds with the desired return periods.

The paper is structured as follows. Section 2 discusses the HURDAT2 database, the methodologies used for the calculation of 6-h storm displacement probabilities, the modeling of storm genesis, the types of ML–DL models used in this work, model architectures, implementation, training strategies, hyperparameter tuning, and the simulation methodology. Section 3 presents simulation results and compares global summaries of the simulated storms with corresponding summaries from HURDAT2. Section 4 demonstrates the efficacy of the synthetic storm simulation model in predicting long-term estimates of storm wind speeds at locations along the U.S. North Atlantic coastline. Section 5 presents conclusions based on our results.

2. Model development

a. Databases

In HURDAT2, the central pressure is tabulated for each storm since 1975. Data tabulated before the use of satellites in the 1970s were based on sparse observations. Different models are trained on different sets of input features depending on the purpose of the model (see Tables 1 and 2), so multiple distinct datasets were created from HURDAT2 by considering various sets of restrictions. The database for the LSTM trajectory models included all storms between 1851 and 2019 which satisfied the following conditions. Only storms with 6 or more time records are retained, which reduces the number of storms from 1893 to 1825. The database is further reduced to 1754 storms by retaining only the storms with $w_m$ (maximum 1-min wind speed at 10-m elevation) recorded for all time instants. The 71 storms for which latitude $\phi > 70^\circ$N, longitude $\lambda > 10^\circ$E, or translation speed $V > 25$ m s$^{-1}$ (48.6 kt) were
TABLE 2. Input features to the ML models for intensity, and corresponding model outputs at 6-h intervals. Here, \( t \) denotes the \( t \)th time instant.

<table>
<thead>
<tr>
<th>Model (abbreviation)</th>
<th>Architecture</th>
<th>Database</th>
<th>Input features</th>
<th>Output variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity–ML: ( p_c )</td>
<td>RF–ML</td>
<td>DB-3</td>
<td>( A^{(t-1)}, \phi^{(t-1)}, V^{(t-1)}, \theta^{(t-1)}, \psi^{(t-1)}, w_m^{(t-1)}, p_c^{(t-1)}, \xi^{(t-1)} )</td>
<td>( p_c^{(t)} )</td>
</tr>
<tr>
<td>(I–ML–1( p_c ))</td>
<td>Many-to-many-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensity–ML: ( p_c )</td>
<td>RF–ML</td>
<td>DB-3</td>
<td>( A^{(t-2)}, \phi^{(t-2)}, V^{(t-2)}, \theta^{(t-2)}, \psi^{(t-2)}, w_m^{(t-2)}, p_c^{(t-2)}, \xi^{(t-2)} )</td>
<td>( p_c^{(t)}, p_c^{(t+1)} )</td>
</tr>
<tr>
<td>(I–ML–2( p_c ))</td>
<td>Many-to-many-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensity–ML: ( w_m )</td>
<td>RF–ML</td>
<td>DB-3</td>
<td>( A^{(t-1)}, \phi^{(t-1)}, V^{(t-1)}, \theta^{(t-1)}, w_m^{(t-1)}, p_c^{(t-1)}, \xi^{(t-1)} )</td>
<td>( w_m^{(t)}, w_m^{(t+1)} )</td>
</tr>
<tr>
<td>(I–ML–1( w_m ))</td>
<td>Many-to-many-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensity–ML: ( w_m )</td>
<td>RF–ML</td>
<td>DB-3</td>
<td>( \hat{A}^{(t)}, \phi^{(t)}, \psi^{(t)}, p_c^{(t)}, w_m^{(t)} )</td>
<td>( w_m^{(t)}, w_m^{(t+1)} )</td>
</tr>
<tr>
<td>(I–ML–2( w_m ))</td>
<td>Many-to-many-2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

also excluded. The detailed reasons are provided in Bose et al. (2021, 2022). The resulting dataset of 1683 storms, called DB-1, is used to train the LSTM–RNN trajectory models. The RF models for intensity use central pressure \( p_c \) as an input feature, which is available at each time instant for a subset of DB-1, called DB-2, consisting of 583 storms between 1975 and 2019. To focus the wind speed predictions on the most extreme storms, a database called DB-3 that consists of 384 storms for which \( w_m \geq 25 \text{ m s}^{-1} \) (48.6 kt) at least once in their life span was created. This database is used to train the RF intensity models. In total, DB-1 contained 45,291 time records, while DB-3 contained 12,637 time records.

For the purpose of training and validation of the models, stormwise partitioning was favored over partitioning the individual time records. Storms were chosen randomly without replacement into the training, validation, and test storm sets for the trajectory models, and training and test storm sets for the intensity models. Thereafter, data tuples were created from the segregated storms for each of these purposes. The training set consists of 80% of the storms, and the rest were used for validation and testing of the trained models. For the intensity models, a two-way split was preferred to a three-way split due to limited database sizes.

b. Input features

The spherical coordinates (\( \lambda, \phi \)) are used unchanged from HURDAT2. The storm translation direction and speed are needed for storm intensity modeling as in Emanuel et al. (2006), Schwerdt et al. (1979). It is assumed that the storm translation is linear between two time instants 6 h apart. Translation direction (\( \theta \)) is obtained at the 6th time instant as

\[
\theta_t = \tan^{-1} \left( \frac{\phi_t - \phi_{t-1}}{\lambda_t - \lambda_{t-1}} \right).
\]

The distance (\( d \)) between storm locations at two time instants was calculated using the Haversine formula, and the 6-h-averaged translation speed (\( V \)) is

\[
V_t = \frac{d(\phi_{t-1}, \phi_t, \lambda_{t-1}, \lambda_t)}{\Delta t = 6 \text{ h}}.
\]

The maximum 1-min sustained wind speed (\( w_m \)) at 10-m elevation is used unchanged from HURDAT2. In HURDAT2, \( w_m \) is approximated to the nearest 5.14 m s\(^{-1}\) (10 kt) between 1851 and 1885 and to the nearest 2.57 m s\(^{-1}\) (5 kt) thereafter. The central pressure (\( p_c \)) in millibars is used as given in the HURDAT2 database.

For trajectory models, the input features should contain information associated with trends of past storm motion (Emanuel et al. 2006; Moradi Kordmahalleh et al. 2016). Emanuel et al. (2006) used a “transition matrix” to capture past trends and construct Markov chains. A similar approach is adopted here. The whole computational domain in (\( \lambda, \phi \)) coordinates (extents of which are defined later in section 2f) was decomposed into rectangular cells. The 6-h displacement probabilities associated with a cell (\( p_k \)) reflect the displacement tendency of a storm center to enter adjacent cells (6th cell) over any 6-h interval. A detailed description of the calculations of the 6-h displacement probabilities (\( p_k \)) is provided by Bose et al. (2021, 2022). Storm displacement probabilities were computed based on motions of DB-1 storms and used as input features of the DL trajectory models. For any given cell, displacement probabilities were calculated for the \( m = 13 \times 11 = 143 \) associated cells. Therefore, for the DL trajectory models, the number of input features at each time record was 148 (\( \phi \) and \( \lambda \) coordinates, storm translation speed \( V \) and direction \( \theta \), wind speed \( w_m \), and historical 6-h displacement probabilities \( p_{k,i} \) calculated at 143 cells associated with the cell containing \( \lambda \) and \( \phi \)). The ML intensity models’ input features include \( p_c \), but not the displacement probabilities for 6 features in total. The input features for the different trajectory and intensity models considered herein are summarized in Tables 1 and 2, respectively.

c. Trajectory models

The trajectory models use the RNN architecture with LSTM cells, invented by Hochreiter and Schmidhuber (1997), as the main recurrent unit. RNNs are utilized in DL to extract pattern and context from sequences, and they are used herein to advance the location of a storm in time. The LSTM layers attend to the vanishing gradient issue in sequence problems. In an LSTM unit, relevant information from the past may be retained over long sequences via a cell state that passes through all LSTM layers. Early use of LSTMs in weather forecasting is reported by Gómez et al. (2003). Except for the predicted quantities (see Table 1), the LSTM–RNN models used herein are similar in all aspects to the models used in Bose et al. (2021)
to forecast the evolution of North Atlantic hurricanes from genesis.

Bose et al. (2021) showed that the many-to-one-type recurrent prediction architectures are susceptible to compounded error accumulation for long-term forecasting beyond the minimum recurrence period (i.e., 6 h in the present case). The many-to-many prediction architectures were shown to reduce compounded error accumulation. Furthermore, the many-to-many-2 prediction architecture that takes in two time records as inputs and forecasts two time records in advance was shown to be the most accurate of all LSTM–RNN models for 6-h short-term forecasts. Therefore, the many-to-many-2 prediction architecture is preferred for the trajectory models used herein, in conjunction with the many-to-many-1 model, which advances a storm one time step from its initial condition/genesis. The bidirectional LSTM architecture is used (Bose et al. 2021). Schematics showing the architectures of the many-to-many-type models used herein are provided in Bose et al. (2022). In addition to the input and output layers, the bidirectional architecture consists of two bidirectional LSTM layers at either side of a repeat-vector layer for the many-to-many models. Each LSTM cell in each bidirectional layer had an input and output dimension of 64, so 128 neurons were used in each layer. The input features to the trajectory models and predicted quantities are presented in Table 1.

The LSTM–RNN models include $w_m$ as an input feature which couples the trajectory models and intensity models. Previous works, with few exceptions, such as one of the trajectory models used by Emanuel et al. (2006), did not attempt to couple the trajectory and intensity of a simulated storm. In reality, storm trajectories are dependent on the storm intensity and the ambient atmospheric conditions. Coupling of the trajectory and intensity models through the input features is a salient aspect of the present work. However, the coupling is only one way; the trajectory is updated first, and then the intensity. They are not updated simultaneously (see Fig. 2).

When simulating a storm, a set of input features at time steps $i - 2$ and $i - 1$ ($i - 1$ at genesis) are fed into a trajectory model to obtain 6-h increments in $A$ and $\phi$, $\Delta A$ and $\Delta \phi$, for the next two time instants. The storm locations at time steps $i$ and $i + 1$ are obtained by the following expressions, which involve the addition of error terms $e_{\Delta A,\Delta \phi}$ to the ML–DL model predictions:

$$\lambda^{(i)} = \lambda^{(i-1)} + \Delta \lambda^{(i-1)} + e_{\Delta \lambda}^{(i)}, \quad (3)$$

$$\phi^{(i)} = \phi^{(i-1)} + \Delta \phi^{(i-1)} + e_{\Delta \phi}^{(i)}, \quad (4)$$

$$\lambda^{(i+1)} = \lambda^{(i)} + \Delta \lambda^{(i)} + e_{\Delta \lambda}^{(i+2)}, \quad \text{and} \quad (5)$$

$$\phi^{(i+1)} = \phi^{(i)} + \Delta \phi^{(i)} + e_{\Delta \phi}^{(i+2)}. \quad (6)$$

In these expressions, the error terms $e_{\Delta A,\Delta \phi}$ are resampled from an error probability density function (PDF) obtained from the model predictions performed on test sequences not used for training. The longitude and latitude errors are resampled jointly to maintain correlations. The addition of the random error increases the likeness of simulated storms to real storms. The predicted trajectories are unphysically smooth otherwise. Similar error terms were included by other researchers for this purpose, such as Vickery et al. (2000, 2009) and Snaiki and Wu (2020).

d. Intensity models

Two sets of intensity models are created, one to predict $p_c$ and another to predict $w_m$. The input features to these models at different time steps, their abbreviated nomenclature, underlying model architecture, and the database used for model training/testing are listed in Table 2. The underlying architecture is an RF (Pedregosa et al. 2011) because there are fewer storms with $p_c$ and $w_m$ provided for each time instant. Additionally, the distributions of $p_c$ and $w_m$ have heavy tails that are important for estimating extreme wind speeds and better accommodated by RF models. The RF models are built upon decision trees (DTs) as implied by the name [see Quinlan (1986) for an example of a DT algorithm]. In an RF, the only tuned hyperparameter is the number of DTs, $n_{\text{estimators}}$, ranging from 5 to 1000. In the present work, all input features were considered for splitting of the predictor variables. In this manner, each DT in an estimator was expanded up to its maximum depth, that is, until the “leaves are pure.” The choice of $n_{\text{estimators}}$ is based on the test–train split. During training, the whole training database is used and consequently tested on the test dataset. The hyperparameter $n_{\text{estimators}}$ that provides the minimum mean squared prediction error for the test set is chosen.

The intensity models include the landfall status ($\xi$) as a binary input feature at all input time steps (see Table 2). Tropical storms dissipate on landfall because of the lack of latent and sensible heat from the ocean and because of increased surface roughness. Vickery et al. (2000, 2009) and Snaiki and Wu (2020) modeled the storm-relative intensity by equations whose parameter coefficients were functions of a storm’s location, the subbasin being traversed, intensity, heading, etc. Upon landfall, the relative intensity of a storm was obtained using a dissipation model (Vickery and Twisdale 1995). The decay models, while motivated by physical considerations, are essentially empirical. In a data-based approach, the signature characteristics of storm dissipation over land are inherent in the database, and therefore learned by the RF model. Distinguishing a storm’s landfall status helps the ML models learn the dissipative characteristics of a storm over land; a separate storm dissipation model is no longer necessary. The $\xi$ input feature is found to be an important predictor for both $p_c$ and $w_m$. Its inclusion facilitates the use of the same prediction model for storm intensity over both land and sea; a separate storm dissipation model is no longer necessary. For the implementation of the $\xi$ input feature, we have used the cartographic boundary shape files from the U.S. Census Bureau (2018) as the database.

Modeling of storm intensity based on a storm’s location is critical as the intensity of a storm is highly dependent on ambient conditions such as the sea surface temperature, salinity, etc. These conditions vary among subbasins/regions, and, therefore,
significant variability of storm intensity is expected between different regions. This is difficult to capture using a single model, but training a different model for each region would require training more models than is practical. Instead, three sets of intensity models are trained based on the magnitude of $w_m$. The lower limits of the training dataset are chosen with care. The first set of models is trained on the whole training database created from DB-3. The lower limit of $w_m$ for training the second set of intensity models, $w_m \geq 25$ m s$^{-1}$, corresponds to a tropical depression wind speed; the second lower limit for training the third set of intensity models, $w_m \geq 50$ m s$^{-1}$, corresponds to the lower limit of category-3 hurricane wind speeds. In this model training strategy, storm category status has been implicitly emphasized through application of the criteria in Table 3 on the input feature $w_m$. While simulating the tracks, the first set of models is applied if at all input time instants, $w_m$ remains below hurricane wind speeds, that is, $w_m < 33$ m s$^{-1}$ ($\sim 64$ kt); the second set of models is applied if a storm attains up to category-3 hurricane wind speeds, that is, at any of the input time instants $33$ m s$^{-1} \leq w_m < 58$ m s$^{-1}$; the final set of models is applied in case a storm attains category-4 hurricane wind speeds or above, that is, $w_m \geq 58$ m s$^{-1}$ ($\sim 112.74$ kt). The model names, databases used for training, and $w_m$ ranges are detailed in Table 3. In this flexible strategy, each segment of the distribution of $w_m$ is modeled separately by manipulation of the training database, DB-3; complexity associated with the use of zonal/regional intensity models may be avoided. Further, the correct spatial distribution of high wind speed records is retained. Most importantly, the high wind speeds in category-4 and category-5 hurricanes may be modeled accurately instead of the being “oversmoothed” by the much larger number of low-intensity storms.

For predictions of $p_c$ and $w_m$, an error term resampled from the PDFs of prediction errors calculated on test data sequences is added to the model predicted quantity using the set of equations

$$\gamma^{(i)} = \gamma_{\text{model}}^{(i)} + \varepsilon^{(+1)}$$

and

$$\gamma^{(+1)} = \gamma_{\text{model}}^{(+1)} + \varepsilon^{(+2)},$$

where $\gamma$ is a prediction of either $p_c$ or $w_m$, and $\gamma_{\text{model}}$ is the RF model predictions. The superscripts for the error terms $\varepsilon^{(+1)}$ and $\varepsilon^{(+2)}$ indicate the prediction step corresponding to the intensity models I–ML–1: $\gamma$ and I–ML–2: $\gamma$ from Table 3, respectively. The error PDFs are constructed using the Gaussian kernel density estimator (KDE) of the model prediction errors computed on test sequences. Joint resampling of the error $\varepsilon^{(+1,+2)}$ for the I–ML–2: $\gamma$ models ensures maintenance of correlations. Addition of these error terms is necessary to incorporate variability into model predictions. Otherwise, for a given set of input features, any of the intensity models would always output the same $\gamma_{\text{model}}$. In reality, based on the ambient conditions, a storm may evolve differently starting from the same initial conditions due to the chaotic nature of the governing physics of storm evolution.

e. Storm origination

The trajectory and intensity models described to this point require a complete set of input features to make their first prediction. The initial input features, with the exception of $p_c$, are simulated from a joint probability distribution constructed using the initial conditions from all storms with more than six time records and an initial speed, $V > 0$. The joint distribution was constructed by first fitting a Johnson $S_U$ distribution (see, e.g., Johnson et al. 1994, chapter 12) to the histogram of each input feature. The Johnson $S_U$ distribution is a four-parameter family of probability distributions over the whole real line. In contrast to the family of Gaussian distributions, which are all symmetric, it accommodates a wide variety of skewed distributional shapes. A Gaussian copula (Nelsen 2007) is used to combine the univariate marginal probability distributions into a joint probability distribution $F_y$. The covariance matrix for the copula was the observed correlation matrix of the input features. The initial input features for a simulated storm are sampled from $F_y$, but a rejection step (see, e.g., 10.3 of Gelman et al. 2013) is performed to ensure that the sampled proportion of storms from each of the five subbasins (viz., the tropical Atlantic, the Caribbean, the Gulf of Mexico, the East Coast, and the subtropical Atlantic) matches the observed proportion.

An initial value for $p_c$ must also be generated. Of the 1719 storms used to construct $F_y$, 603 have $p_c$ recorded. For those storms, multiple linear regression (Neter et al. 1996) is used to predict $p_c$ based on the values of the other input features. The root-mean-square prediction error of the multiple linear regression is less than 1%.

The prescription to generate an initial value for a simulated storm is to sample from $F_y$ first and then perform a rejection step to ensure agreement with subbasin proportions. Last, use the sampled input features to predict $p_c$, yielding a complete initial condition for a storm. To simulate one full year of storms, the number of storms to simulate in that year, which is random, must be selected. This is done using a negative binomial distribution (Johnson et al. 2005) fitted to counts of storms per year in HURDAT2 since 1975. Figure 1 shows a comparison of initial conditions from HURDAT2 (plus signs) to simulated conditions (stars). Generally, the simulated initial conditions match those found in HURDAT2 well. In all cases, a convex hull around the simulated initial conditions would envelop the corresponding
convex hull for the initial conditions from HURDAT2. The simulation approach is thus successfully generating all important initial combinations, in agreement with HURDAT2. We note that some spurious storms are also generated, such as storms with initial wind speed \( w_m > 30 \text{ m s}^{-1} \) (58.3 kt), \( p_c < 990 \text{ mb} \) (1 mb = 1 hPa), etc. They are attributable to the non-limit of the Johnson SU distribution. However, such spurious storms are very few in number compared with the total number of storms simulated herein.

f. Simulation methodology

A flowchart of the simulation procedure to generate a single storm using the trajectory and intensity models is presented in Fig. 2. The synthetic storm genesis model provides an initial condition (time step \( i = 0 \) in Fig. 2) for a storm. The input features at \( i = 0 \) are fed into the trajectory model T–DL–1, which predicts the 6-h increments in \( \lambda \) and \( \phi \) at \( i = 0 \), that is, \( \Delta \lambda^{(0)} \) and \( \Delta \phi^{(0)} \). The increments provide the storm location at \( i = 1 \) using Eqs. (3) and (4). Next, the I–ML–1 model and Eq. (7) are used to obtain \( p_c^{(1)} \), a necessary input into the I–ML–1 model. The I–ML–1 model and Eq. (7) may then be used to predict \( w_m^{(1)} \). This completes the execution of the prediction time step \( i = 1 \).

After completion of each prediction time step, several checks are performed, denoted as \( Q \) in Fig. 2. The conditions that must be satisfied at each time step \( i \) for the procedure to continue are:

\[ \Delta \lambda^{(i)} > 10^\circ, \Delta \phi^{(i)} > 5^\circ, p_c^{(i)} > 975 \text{ mb}, w_m^{(i)} > 30 \text{ knots} \]

The algorithm is repeatedly applied (colored box in Fig. 2) utilizing the two-time-step DL–2/ML–2 models until the storm is terminated.
advance to the next prediction step are listed below (the storm is otherwise terminated):

- \( w_m^{(0)} \geq w_m^{\text{critical}} = 8 \text{ m s}^{-1} \),
- \( w_m^{(0)} > V_l^{(0)} \),
- 102.5°W \( \geq \lambda \geq 10^\circ\text{W} \),
- 30°N \( \geq \phi \geq 0^\circ\text{N} \),
- \( i + 1 < i_{\text{max}} = 120 \).

The first criterion ensures termination of a storm with low \( w_m \), that is, the critical \( w_m = 8 \text{ m s}^{-1} \) (15.55 kt). The other constraint on \( w_m \) is based on the physical consideration that for a Rankine vortex model to be valid, \( w_m \) must be greater than the translation speed of the storm. The next two criteria ensure that the calculations are performed in the computational domain bounding box defined as \( \phi \in [0^\circ, 50^\circ\text{N}] \) and \( \lambda \in [10^\circ, 102.5^\circ\text{W}] \).

The last criterion limits the maximum possible life span of a storm. A storm may have a maximum of \( i_{\text{max}} = 120 \) time records, that is, the maximum life span of a storm is \( t_{\text{max}} = 720 \text{ h} \).

3. Results: Efficacy of the synthetic storm simulation model

Synthetic storms are generated in sets of 100-yr periods. In total, five synthetic storm databases are simulated and compared with the storms in HURDAT2 since 1920 (1302 storms). Since the intensity models were trained on storms only since 1975, there exist many storms in this comparison set that were not part of the training data for the intensity models. The initial/genesis conditions for the simulated storms are obtained as described in section 2, which is based on storms occurring after 1975. Due to enhanced storm detection and tracking capabilities, the average number of storms listed per year in the HURDAT2 database has increased since the beginning of the satellite era in the 1970s (Vecchi and Knutson 2008, 2011; Vecchi et al. 2021); additionally, beginning in 2000, a new bias has emerged in the Atlantic basin with an increase in the number of very short-lived (2 day) tropical storms (Landsea et al. 2010). These short-lived storms are due to modern observational capabilities and are not attributable to anthropogenic climate change (Villarini et al. 2011). Therefore, the average number of storms generated by the synthetic storm generator for a 100-yr period is generally larger than 1302. On average, 1619 storms were simulated in each simulated database. After simulation, on average, each database consisted of 49046 time records compared with 34589 observed time records in HURDAT2 over the 100-yr period within the computational domain considered for the simulations.

a. Efficacy of the trajectory models

The joint distribution of the 6-h increments in \( \phi \) (i.e., \( \Delta\phi \)) and \( \lambda \) (i.e., \( \Delta\lambda \)) for the simulated storms over a 100-yr period...
(one of the five simulated storm databases) is compared with its HURDAT2 counterpart in Fig. 3. As the simulated storms were generated in the computational domain bounded by \( \phi \in [0^\circ, 50^\circ N] \) and \( \lambda \in [10^\circ, 102.5^\circ W] \), the HURDAT2 records since 1920 with storm positions within these limits are used for comparison. Figure 3 suggests that 6-h storm movements in HURDAT2 are predominantly in the northeast direction (for the simulated database considered herein, 47.7% of the records correspond to a 6-h motion in the northeast direction compared with \( \sim 37.5\% \) of the records in HURDAT2). The simulated storms generally agree with that trend, and the storm motions in other quadrants, that is, 6-h storm motions toward the southwest, northwest, or southeast are also captured well by the simulation. The most extreme storm movements, that is, \( \Delta \phi > 2 \) and \( \Delta \lambda > 4.5 \), are not well captured by the trajectory models.

Figure 3 only provides a qualitative view. Both the left and right tails of \( \Delta \phi \) and \( \Delta \lambda \) from HURDAT2 were underpredicted by the models. The frequency in the midrange for both quantities was overpredicted. This is investigated later.

Plotting the combinations \( \{\Delta \phi, \Delta \lambda\} \times \{\phi, \lambda\} \), (Fig. 4) again shows that the extreme tails of the distributions of \( \Delta \phi \) and \( \Delta \lambda \) from HURDAT2 are not captured by the trajectory models. This again is due to smoothing by the LSTM–RNN models. However, the relevant ranges of these increments are adequately represented for the simulated storms.

The frequency distributions of the life span of storms in hours for the five simulated storm databases are compared with their HURDAT2 counterparts in Fig. 5 and account for the larger number of storms in the simulated storm databases by plotting density on the vertical axis instead of raw frequency. The life span of a simulated storm is dependent on the efficacy of the intensity models. The frequency distributions are in good agreement only in the long-life-span range above 400 h. HURDAT2 has a larger proportion of short-life-span storms (life span between 48 and 144 h) than the simulated...
storm databases, whereas the simulated databases have a larger proportion of medium-life-span-range storms (between 168 and 400 h) and very-short-life-span storms (life span below 48 h). The discrepancies are possibly due to the use of DB-3 for storm intensity modeling. Overemphasis on storms that attain larger values of $w_m$ may result in storms of longer duration. Also, underemphasis on low values of $w_m$ means that storms with low $w_m$ die out sooner than HURDAT2 storms. This results in the large overshoot in the very-short-life-span range.

To account for the larger number of storms in the five simulated storm databases, the PDFs of $\Delta \lambda$ and $\Delta \phi$ (instead of raw frequencies) are shown for these five databases in Fig. 6 and compared with the historical storms in the HURDAT2 database since 1920. The distribution shape for simulated storms matches the HURDAT2 storms pretty well. In particular, the peaks occur at the same place and the asymmetry in the distribution of $\Delta \lambda$ is captured by the simulated storms. For both $\Delta \phi$ and $\Delta \lambda$, there is a spike at zero for HURDAT2 storms that is not reflected by the simulated storms.

The trajectories of storms generated in each of the five subbasins from one of the five simulated storm databases are compared with those of the HURDAT2 database in Fig. 7.
FIG. 7. Comparison of storm trajectories based on their inception subbasin: (left) simulated storms for an arbitrary 100-yr period; (right) the HURDAT2 storms since 1920. Colors are arbitrary and only intended to make individual trajectories more distinguishable.
Excellent visual similarity is apparent between the trajectories of storms generated in all subbasins, although the trajectories of the simulated storms are smoother than the HURDAT2 storms even after adding the error terms in Eqs. (3)–(6). A few synthetic storms generated in the Caribbean and the Gulf of Mexico subbasins and in the Pacific Ocean travel westward, and some storms generated in the tropical Atlantic (close to the equator) travel southward. These anomalies are expected because the initial/genesis coordinates of these storms fall out of the range of the coordinates listed in HURDAT2 and therefore are not available for model training. Neural networks are known to provide erroneous results for out-of-bound inputs. The movements of these storms show up as the anomalous time records in Fig. 4 at the left end of φ.

Landfall status is an important derived property of the trajectory models that is pertinent to accurate estimation of wind speeds on the U.S. mainland. In the HURDAT2 database, 363 storms made landfall in the United States out of the 1302 storms listed since 1920 (i.e., ~27.9% of storms). For the five simulated storm databases presented herein, each for a 100-yr period with an average of 1619 storms per database, the average number of storms making landfall in the United States was 480.2 (i.e., ~29.6% of storms). The good agreement between the proportion of simulated and actual storms that reach the mainland further validates the accuracy of the trajectory models and shows that they are fit for the purpose of estimating hurricane-induced extreme coastal wind speeds.

b. Efficacy of the intensity models

The trajectory and the intensity models are coupled via their input features; therefore, their efficacy cannot be assessed separately. In this section, we demonstrate that the predictions from the intensity models, that is, $p_c$, and $w_m$, faithfully emulate the characteristics in HURDAT2. To account for the larger number of storms in the simulated storm databases, PDFs are preferred over frequency distributions for statistical comparison purposes.

In the top frame of Fig. 8, the central pressure ($p_c$) predicted by the intensity models is plotted against latitude ($\phi$) for one of the simulated storm databases and is compared with the HURDAT2 storms since 1920. The simulated $p_c(\phi)$ values are in good agreement with the HURDAT2 storms. The lowest values of $p_c$ in HURDAT2 are obtained in an intermediate range of $\phi \in [12^\circ, 30^\circ]$. The synthetic storms capture this property. Additionally, the synthetic storms accurately capture the spread in the $p_c$ at all $\phi$. The time records corresponding to the anomalous southward-traveling storms are clearly discernible in the top panel of Fig. 8 at low $\phi < 10^\circ$, for which the intensity models assign values in the median range of $p_c$ from the HURDAT2 database.

The pressure deficit $\Delta p = p_c - p_m$, where $p_m$ is the ambient pressure ($\equiv 1013$ mb), is plotted against $w_m$ for one of the five simulated storm databases with the storms in the HURDAT2 database since 1920 in the lower panel of Fig. 8. These quantities are clearly negatively correlated, and the simulated storms emulate the characteristics from HURDAT2 well. The spread in $\Delta p$ at a given $w_m$, and the spread in $w_m$ at a given

Fig. 8. (top) Distribution of central pressure $p_c$ plotted against the latitude $\phi$ from a simulated storm database for a 100-yr period compared with that from the HURDAT2 database since 1920. (bottom) Joint scatterplots of pressure deficit ($\Delta p = p_c - p_m$) and the max wind speed $w_m$ from a simulated storm database for a 100-yr period compared with that from the HURDAT2 database since 1920.

$\Delta p$, are accurately captured by the simulated storms. Moreover, the models are able to predict high values of $w_m$ commensurate with HURDAT2. Therefore, the sequential application of separate models for $p_c$ and $w_m$ accounts for the intensity of the simulated storms effectively.

The PDFs of $w_m$ for the five sets of synthetic storms, each for a period of 100 years, and for the HURDAT2 storms since 1920 are compared in the top frame of Fig. 9. The middle frame of Fig. 9 compares the PDFs of $w_m$ for the time records with landfall status in the continental United States. Finally, in the bottom frame of Fig. 9, the PDFs of $w_m$ at the very first instance of storm landfall in the continental United States for the five simulated databases are compared with those of the HURDAT2 storms since 1920. The proportion of wind speeds greater than $25$ m s$^{-1}$ (48.6 kt) is slightly overpredicted by the five sets of simulated storms in the top frame of Fig. 9. The overprediction decreases as wind speed increases, and may be due to the use of the DB-3 for storm intensity modeling to emphasize the high- wind speed storms in the training database, which are more pertinent for extreme wind estimation purposes. The wind speed distributions for the simulated
storms are in good agreement with HURDAT2 despite the discrepancies in the distributions of storm life span. The proportion of time records with predicted $w_m \geq 75 \text{ m s}^{-1} \left(\sim 145.78 \text{ kt}\right)$ in the top frame and $w_m \geq 50 \text{ m s}^{-1} \left(\sim 97.2 \text{ kt}\right)$ in the middle frame of Fig. 9 is slightly lower than observed in HURDAT2. However, the underprediction of the proportion of very high wind speeds over land is quite small and is probably due to a combination of smoothing and the large number of low–wind speed records over land in DB-3. The comparisons of the distributions in the bottom frame show the effect of the use of $\xi$, the binary landfall status feature. Excellent agreement between the distributions is evident in the bottom frame, indicating the effectiveness of the input feature. Again, the slight underprediction in the high $w_m > 50 \text{ m s}^{-1} \left(\sim 97.2 \text{ kt}\right)$ range is to be expected due to the smoothing associated with the use of the RF models. However, overall, the simulation model is able to capture the high $w_m$ range at the very first instance of storm landfall effectively.

![Graphs showing distributions of max wind speed](image-url)
4. Local wind speed estimates

In this section, our simulation results are compared with HURDAT2 storms since 1920 in terms of the cumulative exceedance probability of the maximum (max) wind speed ($w_m$) recorded at four major locations in the U.S. mainland affected by destructive hurricanes. The trajectories of the storms making landfall near/around those sites are also examined. The locations are Miami, Florida; New Orleans, Louisiana; Cape Hatteras, North Carolina; and Boston, Massachusetts. Miami and New Orleans are chosen because they are severely affected by storm surge. Additionally, they are affected by storms with different trajectory trends. Storms originating in the tropical Atlantic and the Caribbean regularly come near Miami, but storms from these subbasins often make landfall in New Orleans. Also, many storms making landfall at New Orleans originate in the Gulf of Mexico. Therefore, wind speed estimation at New Orleans is especially challenging. Cape Hatteras is one of the most vulnerable regions for hurricanes. Boston is representative of the New England region, where extratropical storms often make landfall. Additionally, storms affecting Boston are typically fast moving compared with tropical storms.

Only storms that pass within 100 km of a chosen site’s location in spherical coordinates attaining $w_m \geq 40$ kt (1 kt = 0.51 m s$^{-1}$) are considered. A threshold of 40 kt was chosen because it was also used by Emanuel et al. (2006). In figures below (see Figs. 11, 14, 17, and 20), trajectories are plotted for one of the five simulated databases (left frame) and compared with the trajectories of storms from HURDAT2 since 1920 (right frame). The cumulative exceedance probability of $w_m$ based on simulated storms and the storms in HURDAT2 since 1920 in knots at Miami, New Orleans, Cape Hatteras, and Boston, respectively, is shown below (see Figs. 10, 13, 16, and 19). Also in these figures (Figs. 10, 13, 16, and 19), the records from the HURDAT2 database are shown in maroon bars, and the frequencies from the five databases are all shown in green bars. In other figures below (see Figs. 12, 15, 18, and 21), the storm that induces the maximum $w_m$ at the chosen locations in each database is analyzed by plotting the storm intensity characteristics, $w_m$, $V$, and $\Delta p$, over the storm’s life span in the left frames and their trajectories in the right frames. The five simulated storms are shown in color and the storm from HURDAT2 is shown in black in the characteristic plots, and in the storm trajectory plots, the symbols are colored by the value of $w_m$.

There are 29 storms in HURDAT2 since 1920 which at some point in their lifetime pass within 100 km of downtown Miami with $w_m \geq 40$ kt, as shown in the right frame of Fig. 11. Considering the small number of time records available for comparison, the cumulative exceedance frequency of $w_m$ for the simulated storms plotted in Fig. 10 is in good agreement with the cumulative exceedance frequency of $w_m$ for the HURDAT2 storms. The exceedance frequencies for the simulated storms are overpredicted at low values of $w_m$ because of the larger number of simulated storms in each of the five synthetic storm databases. However, in the medium–high $w_m$ range, the exceedance frequencies for the five simulated storm databases are very similar to the HURDAT2 storms. Also, the trajectories of the simulated storms are in good agreement with the HURDAT2 storms in Fig. 11. Storm headings on arrival to and departure from Miami are also in good agreement. The trajectories of the simulated storms are smoother than the HURDAT2 storm trajectories due to smoothing by the LSTM–RNN models. The high wind speeds induced by the storms generated in the tropical Atlantic and the Caribbean subbasins are replicated in the simulated storm databases at approximately the correct frequencies.

Storm intensity characteristics of the most powerful storm recorded within 100 km of Miami in each of the five simulated databases, and from HURDAT2 since 1920, are compared in...
Fig. 12. (left) Max wind speed $w_m$, translation speed $V$, and pressure deficit $\Delta p$ plotted over life span of the storms that induce maximum $w_m$ at Miami in each of the five simulated databases and the HURDAT2 storms (black symbols). The red squares indicate the time of landfall. (right) Trajectories of storms plotted in the left panels. The colors in the trajectories indicate the wind speeds $w_m$ (kt).

The 6-hourly evolutions of $w_m$, $V$, and $\Delta p$ for the five simulated storms (colored symbols) are similar to the evolution of the historical storm (black symbols) in the left frame of Fig. 12. Although the trajectory and intensity models emulate the historical storm evolution properties, the individual storms are distinctly different from each other. The storm life spans also vary significantly. The highest $w_m$ induced at Miami by the simulated storms is 150 kt, compared with 145 kt (due to Hurricane Andrew in 1992) in HURDAT2 since 1920. High $w_m$ is generally induced by storms with $V \geq 25$ kt. The correlation between $w_m$ and $\Delta p$ for the simulated storms is similar to the historical storms. The trajectories of the storms colored by $w_m$ in the right frame of Fig. 12 show that the historical storm trajectory trends are indeed replicated by the DL trajectory models. Interestingly, as is the case for the HURDAT2 storm, the most destructive simulated storms recorded at Miami are generated in the tropical Atlantic subbasin. However, the trajectory of the HURDAT2 storm is qualitatively different from the simulated storms. The spatial distribution of high wind speeds from the simulated storms is in good agreement with the historical storms.

The cumulative exceedance frequencies of $w_m$ for the storms passing within 100 km of New Orleans are shown in Fig. 13. At New Orleans, the exceedance frequencies of $w_m$ for the simulated storms are more similar to those from HURDAT2 in the low-$w_m$ range as compared with Miami. On the other hand, in the high–wind speed range, $w_m \geq 95$ kt, the exceedance frequencies of $w_m$ for the simulated storms tend to be smaller than for the HURDAT2 storms. The prediction of the very high wind speeds from HURDAT2 is more complicated in the Gulf region because storms heading northward make landfall and quickly dissipate, causing a steep drop in $w_m$ for storm trajectories in the training dataset. The smoothing inherent in RF models predicts a more gradual drop in $w_m$, and therefore, typically lower wind speeds are predicted by the RF intensity models along the Gulf Coast even before landfall. There are also fewer storm records with $w_m \geq 40$ kt for New Orleans as compared with Miami (Fig. 14). Only 24 storms in HURDAT2 since 1920 satisfy this criterion. However, as was observed at Miami, the trajectory trends are in remarkably good agreement with HURDAT2, even for storms originating in the Gulf of Mexico. Additionally, the simulated storm trajectories after landfall are in good agreement with the HURDAT2 storms traversing over land.

The evolution of storm intensity properties for the most destructive storm recorded within 100 km of New Orleans is compared with the historically strongest storm recorded since 1920 in the left frame of Fig. 15. The life span of the historical storm is shorter than all but one of the simulated storms. However, the highest $w_m$ recorded at New Orleans by the historical storm ($w_m = 150$ kt due to Hurricane Camille in 1969) is larger than the highest $w_m$ recorded by any of the five
simulated storms. The highest wind speed, 116 kt, induced by the simulated storms is close to the second-highest wind speed recorded in HURDAT2 since 1920, 115 kt. This may seem like a large discrepancy between the most intense simulated storm and the most intense historical storm; however, Hurricane Camille occurred in 1969, and the RF models for $p_c$ and $w_m$ are trained on HURDAT2 storms since 1975, the year in which measurements of $p_c$ become available. Neither the training nor testing datasets for the RF intensity models include Hurricane Camille. In fact, the largest hurricane wind speed within 100 km of New Orleans since 1975 was due to Hurricane Katrina in 2005 and was 110 kt, which is in good agreement with the most powerful simulated storms.

The storm trajectories shown in the right frame of Fig. 15 show that, unlike Miami, the most destructive simulated storms at New Orleans arise from any of the tropical Atlantic, Caribbean, or Gulf of Mexico subbasins. The spatial distribution of high-$w_m$ records is also captured by the simulated storms. The dissipative characteristics of the intensity models over land are indicated in the right frame of Fig. 15 by the reduction in $w_m$ as the simulated storms make landfall.

The cumulative exceedance frequencies of $w_m$ at Cape Hatteras (Fig. 16) are representative of the efficacy of the simulation model in generating storms that travel along the East Coast of the U.S. mainland. Again, the simulated storm databases show higher cumulative exceedance frequencies of $w_m$ than the HURDAT2 storms since 1920 in the low–medium ranges of $w_m$ owing to the larger number of storms in the simulated databases (also see Fig. 17). However, the models successfully predict high wind speeds at approximately the right frequencies, which is the most important consideration for the purpose of this work. The storm trajectories of the simulated storms affecting Cape Hatteras are also in good agreement with the HURDAT2 storms (Fig. 17), although the simulated storm trajectories are smoother. There are 44 storms in HURDAT2 since 1920 compared with 58 storms in one of the simulated

![Graphs showing wind speed and pressure deficit over time](image1)

**Fig. 15.** (left) Max wind speed $w_m$, translation speed $V$, and pressure deficit $\Delta p$ plotted over life span of the storms that induce maximum $w_m$ at New Orleans in each of the five simulated databases and the HURDAT2 storms (black symbols). The red squares indicate the time of landfall. (right) Trajectories of storms plotted in the left panels. The colors in the trajectories indicate the wind speeds $w_m$ (kt).
databases which at some point in their lifetime pass within 100 km of Cape Hatteras with $w_m > 40$ kt in Fig. 17. The evolution of intensity properties at 6-h time steps for the most powerful simulated storms (colored symbols) at Cape Hatteras and the most destructive historical storm since 1920 (black symbols) is shown in Fig. 18. Significant variation in life span of the simulated storms is noticeable. The highest $w_m$ recorded at Cape Hatteras for the simulated storms is in good agreement with the historically highest wind speed (115 kt due to Hurricane Helene in 1958). Again, high $w_m$ occurs when $V$ is low, and the simulation models incorporate adequate variability. The storm trajectories in the right frame of Fig. 18 show that the pertinent trajectory trends are emulated by the trajectory models, and the intensity models are able to provide the correct spatial distribution of the highest wind speeds.

Storms impacting New England often have higher translation speeds ($V$). Moreover, storms undergoing extratropical transition may restrengthen before impacting the northeastern coastline of the United States. The choice to study Boston is motivated by these reasons. The cumulative exceedance frequencies of $w_m$ for the five simulated databases are compared with the HURDAT2 storms since 1920 in Fig. 19. Additionally, the left frame of Fig. 20 shows the trajectory of storms passing by within 100 km of Boston from one of the five simulated databases; the right frame shows such storms from the HURDAT2 database. Although the cumulative exceedance frequencies for the simulated databases are overestimates for $w_m < 60$ kt, excellent agreement with the HURDAT2 storms is evident at higher $w_m$. The highest $w_m$ induced at Boston by the simulated storms is 115 kt, compared with 95 kt due to Hurricane Edna (1954). The trajectories of the synthetic storms are also clearly in good agreement with the HURDAT2 storms in Fig. 20. There are 18 simulated storms affecting Boston in a 100-yr period compared with 9 storms recorded in HURDAT2 since 1920.

Six-hour storm intensity evolution properties for the most intense storm affecting Boston from each of the five simulated databases (colored symbols) are shown and compared with the most intense storm since 1920 (black symbols) in Fig. 21. Significant variations in $w_m$ and $D_p$ are noticeable for the simulated storms. The highest $w_m$ recorded at Boston for the simulated storms is in good agreement with the historically highest wind speed. Again, high $w_m$ is predicted for slow-moving storms (low $V$). Adequate variability is demonstrated by the simulation models. The storm trajectories in the right frame of Fig. 21 show that the pertinent trajectory trends are emulated by the trajectory models. The intensity models also provide the correct spatial distribution of the highest wind speeds.

The results indicate that the simulations are repeatable and the variability between simulations is realistic. The estimation of extreme wind speeds over the East Coast is easier compared with the southern Gulf Coast. This is due to a storm’s bearing being typically acute to the coastline along the East Coast at landfall and near normal along the southern Gulf Coast.

5. Discussion and conclusions

This paper demonstrates the applicability of DL and ML techniques for estimating hurricane wind speeds. The recent success of ML–DL methods in several aspects of atmospheric science and weather forecasting (Reichstein et al. 2019; Schultz et al. 2021) provided the impetus for this research. Storm trajectory models developed by Bose et al. (2021) using LSTM–RNN model architectures and 6-h storm displacement probabilities as input features are also leveraged here. These models incur errors similar to those inherent in ensemble models used by the NHC for trajectory forecasts up to 12 h. A set of intensity models was developed using RFs to complete
the synthetic storm generation. These intensity models predict a storm’s central pressure $p_c$ and maximum wind speed $w_m$. Previous works typically modeled storm intensities zonally, as local ambient conditions affect the evolution of a storm’s intensity. In this work, alternatively, three sets of intensity models based on storm intensity ($w_m$) were used. These intensity models have the advantage of embedding storm dissipation characteristics over land through the use of landfall status as an input feature. By coupling the input variables to both the trajectory and intensity models, a coupled storm simulation model is obtained (see Fig. 2). The storm simulation model advances a storm’s location and intensity 6 h at a time, until one of the storm termination criteria is satisfied.

The efficacy of the simulation of the trajectory and the intensity components is tested by simulating five storm databases, each for a period of 100 years, which is then compared with storms from HURDAT2 since 1920. The synthetic storm genesis model selects a number of storms for one year based on a negative binomial distribution fitted to the number of storms each year in HURDAT2 since 1975. The average number of storms in the five simulated databases is 1619, compared with 1302 historical storms since 1920 in HURDAT2. The trajectory models faithfully represent the important statistical properties of the historical storms, such as the 6-h increments in latitude and longitude (Figs. 3, 4, and 6). By using historical storm motion trends as input features, these models capture the dominant storm motion trends, but due to the inherent smoothing associated with LSTM–RNN models, the largest 6-h increments $|\Delta \phi|$ and $|\Delta \lambda|$ (Fig. 3) are underrepresented. However, historical storms that induce high $w_m$ typically have low translation speed $V$ and consequently low $|\Delta \phi|$ and $|\Delta \lambda|$. The values for $\Delta \phi$ and $\Delta \lambda$ plotted against the local coordinates $\phi$ or $\lambda$ in Fig. 4 show good qualitative agreement between simulated and historical trajectories. The use of a reduced database (DB-3) emphasizing the most powerful storms for training the intensity models results in synthetic storms sustaining longer for storms in the medium–high $w_m$ range, and relatively low wind speed storms dying out sooner than historical storms (see Fig. 5). The global trajectories of simulated storms originating in different subbasins are in good qualitative agreement with their historical counterparts (Fig. 7).

Simulated values of $p_c$ and $w_m$ are also demonstrated to be in qualitatively good agreement with the HURDAT2 storms when considering the Atlantic basin as a whole. Figure 8 shows that the distributions of central pressure $p_c(\phi)$ and pressure deficit $\Delta p_c(w_m)$ for the simulated storms faithfully represent the historical storms. The near-linear relation between $p_c$ and $w_m$ is appropriately depicted by the simulated storms. Spreads in the distributions of $p_c$ and $w_m$ are correctly captured by the intensity models. The PDFs of $w_m$ in Fig. 9 for all five simulated databases are in excellent agreement with the historical storms.
Storms achieving $w_m \geq 40$ kt have been analyzed further for four chosen cities, Miami, New Orleans, Cape Hatteras, and Boston, which are all greatly affected by storm surges. The cumulative exceedance frequencies of $w_m$ for these cities (Figs. 10, 13, 16, and 19) show good agreement between the five simulated databases and the historical storms for high wind speeds $w_m > 70$ kt. The largest value of $w_m$ the simulated databases is either about the same (Miami) or larger than (Cape Hatteras and Boston) the largest value of $w_m$ in HURDAT2 for these cities. An exception is New Orleans, for which the largest simulated value of $w_m$ is about the same as for the historical storm with the second-largest wind speed. This discrepancy is due to the fact that the training database for the RF models (starting in 1975) did not include any storms within 100 km of New Orleans producing a maximum wind speed as large as the strongest historical storm since 1920. The cumulative frequencies are overpredicted for $w_m$ below this limit, which is expected on account of the higher number of storms in the simulated databases. The storm trajectories of the storms passing within 100 km of these areas are larger over a 100-yr period due to a larger number of storms in the simulated databases. The trajectory trends of the simulated storms are in good agreement with the historical storms, although the trajectories of the simulated storms are noticeably smoother. The most violent storm from each of the six databases (five simulated and HURDAT2) within 100 km of these areas is extracted, and the intensity properties, $w_m$, $V$, and $D_p$, are plotted in Figs. 12, 15, 18, and 21. These figures show that the 6-hourly rate changes in the simulated storms’ intensity and trajectory are similar to those of the actual storm but still evolve distinctly, highlighting the stochastic nature of the model predictions. The trajectory of the most intense simulated storm in each of the five simulated databases, colored by

![DBS: 18 storms sampled at Boston](image1)

![HURDAT-2: 9 storms sampled at Boston](image2)

**Fig. 20.** Trajectories of storms passing within 100 km of Boston for (left) the simulated storms over an arbitrary 100-yr period compared with (right) those from the HURDAT2 database since 1920.

![DBS: 18 storms sampled at Boston](image3)

![HURDAT-2: 9 storms sampled at Boston](image4)

**Fig. 21.** (left) Max wind speed $w_m$, translation speed $V$, and pressure deficit $D_p$ plotted over life span of the storms that induce maximum $w_m$ at Boston in each of the five simulated databases and the HURDAT2 storms (black symbols). The red squares indicate the time of landfall. (right) Trajectories of storms plotted in the left panels. The colors in the trajectories indicate the wind speeds $w_m$ (kt).


