ᎧTropical Cyclone Intensity Evolution Modeled as a Dependent Hidden Markov Process𝒫

RENZHI JING AND NING LIN

Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey

ABSTRACT

A hidden Markov model is developed to simulate tropical cyclone intensity evolution dependent on the surrounding large-scale environment. The model considers three unobserved (hidden) discrete states of storm intensity change and associates each state with a probability distribution of intensity change. The storm's transit from one state to another is described as a Markov chain. Both the intensity change and state transit components of the model are dependent on environmental variables including potential intensity, vertical wind shear, relative humidity, and ocean feedback. This Markov Environment-Dependent Hurricane Intensity Model (MeHiM) is used to simulate the evolution of storm intensity along the storm track over the ocean, and a simple decay model is added to estimate the intensity change when the storm moves over land. Data for the North Atlantic (NA) basin from 1979 to 2014 (555 storms) are used for model development and evaluation. Probability distributions of 6- and 24-h intensity change, lifetime maximum intensity, and landfall intensity based on model simulations and observations compare well. Although the MeHiM is still limited in fully describing rapid intensification, it shows a significant improvement over previous statistical models (e.g., linear, nonlinear, and finite mixture models).

1. Introduction

Tropical cyclones (TCs) are one of the most destructive natural phenomena on Earth and are the leading cause of great social and economic losses (Geiger et al. 2016; Pielke and Landsea 1998). Accurately assessing TC-related risk is of great importance. However, while TC intensity plays a large role in the damage potential of a storm, TC intensity modeling is still limited (Rappaport et al. 2009; Manganello et al. 2012). This is mainly because the intensity evolution of TCs involves complicated dynamic and thermodynamic processes that have not been fully resolved (Bender et al. 2010; Camargo and Wing 2016). Apart from internal variability (e.g., eyewall replacement cycles; Willoughby et al. 1982), the storm

environment plays a key role in controlling the storm's intensity evolution. The prestorm sea surface temperature (SST) has been considered as one of the most important environmental variables (Knutson et al. 2010), since TCs draw their energy from the underlying seawater. Emanuel et al. (2004) used an ocean coupled model to show that other environmental factors, including maximum potential intensity (MPI) and vertical wind shear, also have great influences on storm intensity. Furthermore, the interaction between TC and the ocean can significantly influence TCs' intensity. The entraining of cold, deep water induced by storm's surface winds may cause strong cooling of SST and thus negatively affect a storm's intensification (Schade and Emanuel 1999; Lloyd and Vecchi 2011).

Due to a combination of inner-core, oceanic, and large-scale environmental processes, many TCs undergo rapid intensification (Kaplan and DeMaria 2003) during their life cycles. Rapid intensification (RI) is the significant strengthening of a TC in a short time [specifically, intensity increases by 30 kt (1 kt ≈ 0.51 m s⁻¹) or more over 24 h], and almost all historical category 4 and 5 hurricanes are RI storms (Kaplan and DeMaria 2003). Accurately predicting RI is not only important in

---

real-time forecasting (Toepfer et al. 2010), but also plays an important role in reproducing TC intensity climatology, as the RI storms control the tail (including possibly a secondary peak) of the statistical distribution of a TC's maximum lifetime intensity (LMI) (Lee et al. 2016a). However, the prediction of RI is still a great challenge, and few operational models are skillful in predicting RI (Elsberry et al. 2007). To fill the gap, Kaplan and DeMaria (2003) defined the rapid intensification index (RII) as a probabilistic guidance. A revised RII is further developed to estimate the probability of RI occurrence for various RI thresholds and for different basins (Kaplan et al. 2010). Although the probabilistic RII forecast is generally skillful compared to the baseline climatology, the false alarm rate of RII is still high, especially for the North Atlantic basin (Kaplan et al. 2010).

Despite the challenges, various statistical methods have been applied to model TC intensity change as a response to environmental forcing. One widely used model is the Statistical Hurricane Intensity Prediction Scheme (SHIPS; DeMaria and Kaplan 1994; DeMaria and Kaplan 1999; DeMaria et al. 2005), which estimates TC intensity change as a linear regression function of key environmental variables (e.g., MPI, vertical wind shear, upper-level winds) and TC characteristics (e.g., current intensity, previous intensity change, and storm's translation speed). Lee et al. (2015) narrowed down the essential predictors in SHIPS to a smaller pool and developed a simpler linear regression model. Lee et al. (2016b) improved this model by adding a stochastic component to better capture the statistical distributions of TC intensification rate and LMI.

More recently, Lin et al. (2017) applied advanced statistical techniques to further explore the dependence of TC intensity change on the environmental variables, including the newly developed ventilation index (VI; Tang and Emanuel 2012). They found that applying the three component variables of VI (i.e., MPI, vertical wind shear, and entropy deficit) is statistically better than applying VI as a single variable, and the simpler relative humidity represents the entropy deficit relatively well. After identifying the most important environmental variables, Lin et al. (2017) developed linear, nonlinear, and mixture regression relationships between TC intensity change and these variables. In particular, they found that the statistical R^2 of their linear and nonlinear models is similar to that of SHIPS, but the statistical R^2 of their mixture model (involving multiple subgroup models) is significantly higher, indicating that the dependence of TC intensity change on the environment is nonhomogeneous. More interestingly, their mixture regression analysis identified three regimes/states of intensity change: static, normal, and extreme, reflecting a storm's slow, moderate,

and rapid intensity change, respectively, during its life cycle. However, applying the mixture model to predict TC intensity evolution requires adding a classification model to identify which regime the storm belongs to at each time step, which can induce large uncertainties/errors, especially for extremes related to RI (Lin et al. 2017).

Motivated by the findings of Lin et al. (2017), here we model TC intensity evolution as a dependent hidden Markov process. This new model is called the Markov Environment-Dependent Hurricane Intensity Model (MeHiM). Like the mixture model, the MeHiM considers three unobserved (hidden) states of intensity change and associates a probability distribution of intensity change with each state. Unlike the (statistical) mixture model, which assumes that the state identity for a time step is conditionally independent of that for another time step, the (stochastic) MeHiM considers the storm's transit from one state to another as a Markov chain.

Both the intensity change and state transition components of the MeHiM are dependent on identified environmental variables. To keep the model relatively simple, we apply only four environmental variables (i.e., MPI, vertical wind shear, relative humidity, and an ocean feedback parameter) and two storm variables (i.e., current intensity and previous intensity change). Few of the developed ocean feedback parameters (e.g., Schade and Emanuel 1999; Vincent et al. 2012; Balaguru et al. 2015), which indicate ocean–storm interaction, have been evaluated in statistical modeling of TC intensity. Here we adopt the ocean feedback parameter of Schade and Emanuel (1999), which has also been used in a simplified dynamical model (Emanuel 2017), and we test its impact on the overall performance of the MeHiM.

The MeHiM is developed to predict the intensity change of storms moving over the ocean. To simulate the intensity evolution over storm's entire life cycle, we add a simple decay model to predict a storm's intensity change over land. Here we develop the model for the North Atlantic basin, though the same methodology can be applied to other basins or on the global scale. To evaluate the developed model, we apply it to simulate intensity evolution of historical storms using observed tracks and environment taken from reanalysis datasets. The performance of the model is measured by its capability to reproduce the observed TC intensity metrics, including spatial distribution of intensity over the basin, probabilistic distributions of intensification rate, LMI, and landfall intensity, as well as temporal evolution of intensity for sample cases. We focus particularly on the model's capability to capture the extremes related to RI.

The MeHiM improves on the mixture model of Lin et al. (2017), because 1) it accounts for the (possibly large) temporal correlation of the states (e.g., slow,

moderate, rapid) of intensity change over storm's life cycle and 2) it integrates the mixture model of intensity change and the identification model of intensity change state within a Markov framework, avoiding statistical inconsistency between the two components. It is interesting to ask whether these improvements in methodology will lead to significant improvements in the model performance, especially in capturing RI related extremes. To answer this question, we build the mixture model (with an added identification model) using the same probability distributions (i.e., Gaussian for intensity change and logistic for state identification), environmental and storm variables, and observational datasets as in the MeHiM. In addition, we build the (Gaussian) linear regression model (similar to SHIPS) using the same variables and datasets to provide a baseline. The linear and mixture models are run in parallel with the MeHiM, and the simulation results are compared among the models and with the observations.

This paper is organized as follows: Following this introduction, the development of the models (i.e., linear regression model, mixture regression model, MeHiM, and land decay model) is described in section 2. Model simulation results are shown and evaluated in section 3. Section 4 discusses the effects of adding the ocean feedback parameter and correctly identifying the extreme state of intensity change in the MeHiM. Section 5 summarizes the main findings of the study.

2. Model development

Three statistical TC intensity models are developed to predict 6-hourly storm intensity change (DV): the linear regression model, mixture model (with an added state identification model), and MeHiM. We build these models on the same six variables: intensity change in the previous step (DV$_p$), the current intensity (V), MPI, vertical wind shear (SHR), relative humidity (RH), and an ocean feedback predictor (OCN). Apart from these intensity models, which are used to simulate a storm's intensity change over the ocean, we also add a simple land model to simulate the decay of the storm intensity over time when the storm moves on land.

All models are developed based on historical records from 1979 to 2014 in the North Atlantic basin. The TC dataset is taken from the IBTrACS WMO archive (Knapp et al. 2010). It includes 6-hourly latitude and longitude positions and 10-min maximum sustained wind speeds at 10 m above the sea surface. The location data are used to calculate the translational velocity of the storm. The storm maximum wind intensity V is estimated by subtracting the surface background wind, estimated as a fraction (0.55 at 20° cyclonically) of the

storm translation velocity (Lin and Chavas 2012), from the observed maximum wind. The 6-h intensity change DV (and DV$_p$) is then obtained from the time series of the estimated storm maximum wind intensity.

The atmospheric variables are derived from the ERA-Interim reanalysis with a resolution of 0.75° × 0.75°, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF; Dee et al. 2011). The maximum potential intensity is theoretically derived following Emanuel (1995) and Bister and Emanuel (1998, 2002). Similar to that in previous studies (e.g., DeMaria and Kaplan 1994; Lin et al. 2017), the deep-layer vertical wind shear is defined as the difference between the 850- and 200-hPa level winds, averaged over a 200–800-km annulus around the storm center. High-level relative humidity is computed as the averaged relative humidity in the layer between 300 and 500 hPa within a 500–800-km annulus around the storm center.

In addition to the atmospheric predictors, we also incorporate OCN, developed by Schade and Emanuel (1999) based on numerical modeling, to represent the ocean's negative impact on storm intensification. OCN is defined as

$$\text{OCN} = 1 - 0.87e^{-z}, \tag{1a}$$

where

$$z \equiv 0.01\Gamma^{-0.4}h_m u_T \frac{\text{MPI}}{V}, \tag{1b}$$

where u_T is the storm translation speed, h_m is the ocean mixed layer depth, and Γ is the thermal stratification below the ocean mixed layer. OCN is a number between 0 (strong weakening influence on the storm) and 1 (weak influence); it is closer to 0 when the storm is stronger, moving more slowly, with a shallower mixed layer depth, and/or with stronger thermal stratification. OCN is more sophisticated than other ocean variables such as the ocean temperature averaged over the mixing layer and the ocean heat content, which indicate only the amount of heat stored in the ocean and are incapable of representing the interaction between the storm and ocean. The ocean heat content is used in SHIPS, but the data are available only for a limited time period and over a limited part of the Atlantic basin (Mainelli et al. 2005); the ocean temperature averaged over the mixing layer was found to be an insignificant predictor of TC intensification (Lee et al. 2015). To estimate the variables in OCN, we take the ocean's salinity and potential temperature from the Ocean Reanalysis System 4 (ORAS4; Balmaseda et al. 2013). The mixing layer depth h_m is computed following de Boyer Montégut et al. (2004) with a temperature difference criterion of

0.5°C. The variable Γ is estimated as the temperature gradient within 100 m below the mixed layer depth.

The following four subsections describe the development of the three over-the-ocean intensity models and the land decay model with statistical fitting results. In all statistical analyses, variables are standardized so that all variables are on the same scale. To maximize the utilization of the data, we train the models considering all 555 North Atlantic TCs from 1979 to 2014. However, since a TC behaves differently when over land, we filtered out these observations (used for training the land decay model) by dividing the storm's intensity evolution into multiple temporal sequences (as temporal sequences of storm intensity are needed for training MeHiM). That is to say, the time series may begin with either the genesis of the storm or the first observation after the storm makes landfall and moves back to the ocean again. The time series ends when the storm dies or hits land. Sometimes the temporal sequences can be very short when a storm hits land twice within a short period of time. These short sequences are considered as the "noise" of the training data and thus are not used. After removing sequences that are shorter than 12 observations (i.e., 3 days), 9809 observations coming from 332 storms (347 time series) remain for training each of the three intensity models. The models are evaluated in terms of their capability of capturing the statistical distributions of various intensity and intensification metrics and simulating realistic sample cases of intensity evolution, for all the 555 storms from 1979 to 2014, as discussed in following sections. We also evaluate the models using nonoverlapping training (271 storms) and testing (284 storms) datasets, as shown in the online supplemental material.

a. OLS

For a linear regression model with ordinary least squares (OLS), the probability distribution of the predictand Y is Gaussian, as shown in Eq. (2), where Y is the 6-h intensity change (DV), and $\mathbf{X}$ is the vector of predictors (i.e., DV_p, V, MPI, SHR, RH, and OCN):

$$Y \sim \mathcal{N}(\alpha + \mathbf{X}\boldsymbol{\beta}, \sigma). \qquad (2)$$

The model coefficients α and $\boldsymbol{\beta}$ are estimated by minimizing the squared residual error (i.e., OLS). The estimated model coefficients as well as the standard deviation σ are shown in Table 1. The signs of the coefficients in front of all variables are as expected (based on previous studies; e.g., DeMaria and Kaplan 1999; Schade and Emanuel 1999; Lin et al. 2017), that is, DV is positively (negatively) correlated to DV_p (V), and it is expected to increase when MPI and RH are larger, SHR

TABLE 1. Coefficients of the OLS model.

Intercept	DV_p	V	MPI	SHR	RH	OCN	σ
0.000	0.472	−0.123	0.083	−0.081	0.046	0.044	0.844

is smaller, and/or ocean cooling is weaker (OCN larger). The storm's last-step intensification DV_p is shown to be the strongest predictor, followed by the current intensity V. Among the environmental variables, MPI and SHR are more important than RH and OCN. The intercept α is nearly zero, which means that the expected value of 6-h intensity change DV is nearly 0 when all five predictors are at their means (i.e., standardized values being zero).

b. FMR

The finite mixture regression (FMR) extends the OLS by incorporating a heterogeneous structure into the model. In this model, each observation is considered to be a random sample generated by one of its k components (groups/clusters), while the identity of the generating component is not observed. In a Gaussian setting, each of the k components is a linear regression, and the probability distribution of the predictand Y is mixed Gaussian, that is,

$$Y \sim \sum_{r=1}^{k} w_r \mathcal{N}(\alpha_r + \mathbf{X}\boldsymbol{\beta}_r, \sigma_r), \qquad (3)$$

where w_r is the component mixing proportion/weight, α_r and $\boldsymbol{\beta}_r$ are regression coefficients, and σ_r is the standard deviation, for each component r. These model parameters can be estimated from data using the maximum likelihood estimation method with the generalized expectation-maximization (EM) algorithm (McLachlan and Krishnan 2007). The EM algorithm also estimates the probability that each observation was generated from each of the k groups, providing an optimal group identification for each observation.

Based on the same dataset as for the OLS, the estimated model parameters for FMR with three groups ($k = 3$) are shown in Table 2. The obtained group weights and features are similar to those from the FMR analysis of Lin et al. (2017). Specifically, over 60% of the observations are optimized by the EM algorithm into group 2, nearly 30% into group 1, and only less than 10% into group 3. Groups 1, 2, and 3 again feature "static," "normal," and "extreme" intensity change (DV), respectively, in the sense that, first, the expectation of intensity change as a response to the environment is weak for group 1, relatively strong for group 2, and the strongest for group 3 (as indicated by the magnitude of the model coefficients in front of the environmental

TABLE 2. Coefficients of the FMR response model.

	Weights	Intercept	DV_p	V	MPI	SHR	RH	OCN	σ_r
Group 1	0.298	−0.107	0.010	0.002	0.035	−0.004	−0.009	0.011	0.094
Group 2	0.610	0.067	0.687	−0.123	0.104	−0.085	0.057	0.054	1.119
Group 3	0.092	0.086	0.275	−0.424	0.093	−0.263	0.110	0.123	2.706

variables); and, second, the model standard deviation σ_r (representing the width of the subgroup Gaussian distribution) increases significantly from group 1, to group 2, and to group 3. As a result, the sample mean of DV (average over all observations optimally assigned to the group) vary around zero in absolute value (−0.07 for group 1, 0.84 for group 2, and −0.13 for group 3) but the sample standard deviation of DV increases substantially from group 1 (0.48), to group 2 (5.75), and to group 3 (13.90), indicating that, regardless of the environmental conditions, small intensity change observations are assigned into group 1, extreme intensification (and deintensification) into group 3, and the rest into group 2. Such grouping features will be further discussed with the Markov model analysis. Unexpected signs of the model coefficients for V and RH are obtained for group 1, possibly due to statistical compensation.

The FMR is not a predictive model, and conducting simulations with FMR requires assigning group weights or membership for the new observation. Similar to Lin et al. (2017), we add a multinomial logistic regression [Eq. (4)] as a classification model to determine the group membership probabilities π_r for the new observation **X** (i.e., DV_p, V, MPI, SHR, RH, and OCN),

$$\pi_r(\mathbf{X}) = \frac{\exp(\theta_r + \mathbf{X}\gamma_r)}{\sum\limits_{j=1}^{k} \exp(\theta_j + \mathbf{X}\gamma_j)}. \quad (4)$$

The multinomial logistic regression model is fitted based on the FMR optimized group memberships, and the obtained model coefficients (θ_r, γ_r) are shown in Table 3. To meet the constraint $\pi_1 + \pi_2 + \pi_3 = 1$, we have arbitrarily picked group 3 as the baseline group (all coefficients are assigned to be zero). Thus, the linear combination of covariates for group 1 (first row) describes the log odds of belonging to group 1 versus

belonging to group 3, and, similarly, the linear combination of covariates for group 2 (second row) describes the log odds of belonging to group 2 versus belonging to group 3. As an example, a one-unit, or one standard deviation, increase in the variable DV_p is associated with a decrease in the log odds of belonging to group 1 versus belonging to group 3 in the amount of 0.360 (coefficient of DV_p). At the same time, a one-unit increase in the variable DV_p is also associated with a decrease in the log odds of belonging to group 2 versus belonging to group 3 by 0.295. Thus, with other covariates fixed, an increase in DV_p will cause the new observation more likely to be assigned to group 3.

To obtain a more general view of the obtained classification model, we estimate the group membership probabilities for three specific scenarios. When all variables of the new observation are at their medians (over all samples in the dataset), the probabilities that it belongs to groups 1, 2, and 3 are 0.41, 0.57, and 0.02, respectively. When all variables of the new observation are at the 95th percentile favorable for larger DV (according to the expected coefficient signs shown in Table 1), the probabilities that it belongs to groups 1, 2, and 3 are 0.37, 0.61, and 0.02, respectively. When all variables of the new observation are at the 95th percentile unfavorable for larger DV, the probabilities that it belongs to groups 1, 2, and 3 are 0.25, 0.72, and 0.03, respectively. The probabilities obtained in the three scenarios are relatively close, indicating that the model may be limited in capturing the change in the membership probability as a response to the change of the predictor variables. Moreover, the probability of belonging to the extreme group (group 3) is always significantly smaller than the group weight from the optimization (~9%; Table 2). This limitation in capturing the extreme state by the FMR classification model was also noted by Lin et al. (2017). This limitation is reduced in the Markov model.

TABLE 3. Coefficients of the FMR classification model.

	Intercept	DV_p	V	MPI	SHR	RH	OCN
Group 1	2.851	−0.360	−1.043	0.037	−0.178	−0.285	−0.216
Group 2	3.354	−0.295	−0.728	−0.173	−0.220	−0.100	−0.121
Group 3	0	0	0	0	0	0	0

c. MeHiM

The MeHiM is based on the idea of the hidden Markov model (HMM; Rabiner 1989; Cappé et al. 2006), an effective tool for modeling stochastic processes. In an HMM, the visible observations are distributed according to the underlying hidden states, where the states are discrete and are assumed to form a first-order Markov process. With a selected number k of states, an HMM can be fully determined by the following three sets of parameters:

1) Initial state probability π_i: the probability of starting from state $i (i = 1, 2, \ldots, k)$.
2) Transition probability π_{ij}: the probability of moving from state i to $j (j = 1, 2, \ldots, k)$.
3) Response distribution f_{Y_i}: the distribution of the response variable Y in state i.

The HMM can be extended to the dependent HMM with the above model parameters dependent on covariates (Zucchini et al. 2016). Then, the likelihood of a response observation sequence $y_{1:T} = y_1 y_2 \cdots y_T$, given the covariates sequence $\mathbf{x}_{1:T} = \mathbf{x}_1 \mathbf{x}_2 \cdots \mathbf{x}_T$, is

$$
\begin{aligned}
& L(y_{1:T} | \mathbf{x}_{1:T}) \\
& = \sum_{s_1=1}^{k} \pi_{s_1}(\mathbf{x}_1) f_{Y_{s_1}}(y_1 | \mathbf{x}_1) \prod_{t=2}^{T} \pi_{s_{t-1} s_t}(\mathbf{x}_t) f_{Y_{s_t}}(y_t | \mathbf{x}_t),
\end{aligned}
\tag{5}
$$

where s_t denotes the state at time t, and all three model components (i.e., initial state probability, transition probability, and response probability) depend on the covariates. For the entire dataset, the total likelihood is the product of the individual likelihood functions for all sequences.

The basic idea of a dependent HMM is applied in the MeHiM to model the sequence of TC intensity change (DV) over the storm's lifetime dependent on the selected storm and environmental covariates (i.e., DV_p, V, MPI, SHR, RH, and OCN). Motivated by the FMR analysis, we assume the storm is in one of three hidden states ($k = 3$) at each time step.

Given the state, the response distribution is assumed to be Gaussian, that is,

$$
Y_i \sim \mathcal{N}(\alpha_i + \mathbf{X}\boldsymbol{\beta}_i, \sigma_i),
\tag{6}
$$

dependent on the six predictor variables as in the OLS and FMR models.

Similar to the FMR analysis, we apply the multinomial logistic regression to model the transition probability that the storm moves from the present state to the next state, that is,
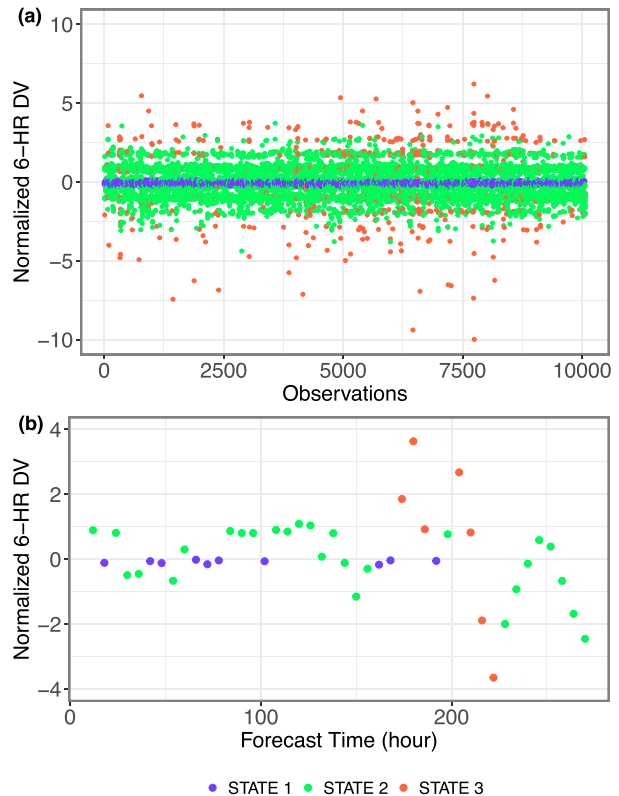

FIG. 1. MeHiM optimal states for (a) all data and (b) a sample storm. All observations of DV are plotted with colors denoting the optimal states given by MeHiM fitting.

$$
\pi_{ij} = \frac{\exp(\theta_{ij} + \mathbf{X}\boldsymbol{\gamma}_{ij})}{\sum_{j=1}^{k} \exp(\theta_{ij} + \mathbf{X}\boldsymbol{\gamma}_{ij})}.
\tag{7}
$$

The transition model (3×3 probability matrix) contains three multinomial logistic regressions for each of the three states, dependent on all 6 predictor variables. The initial state probabilities π_i are similarly determined by a multinomial logistic regression.

The parameter fitting process for MeHiM involves the backward–forward algorithm for likelihood calculation and the Baum–Welch method, an iterative procedure for optimization similar to the EM method. In addition to model coefficients, the optimal state sequence can be estimated through the Viterbi algorithm (Zucchini et al. 2016). A more detailed description of the algorithms can be found in the statistical R package depmixS4 (Visser and Speekenbrink 2010).

Figure 1 shows observed DV colored by estimated optimal states. The first state has a sample mean close to zero (-0.05) with a small sample standard deviation (0.50); the second state features a moderate positive mean (0.43) and variation (5.31); the third state has a

TABLE 4. Coefficients of the MeHiM response model.

	Intercept	DV_p	V	MPI	SHR	RH	OCN	σ
State 1	−0.106	0.010	0.004	0.012	−0.003	−0.008	0.014	0.103
State 2	0.038	0.612	−0.183	0.097	−0.070	0.082	0.028	0.742
State 3	0.280	0.398	−0.424	0.521	−0.563	−0.091	0.336	1.579

large positive mean (3.04) but also a large standard deviation (10.85) with DV taking large negative and positive values. The three regimes of intensity change also appear clearly as three phases during the lifetime of individual storms. As listed in Table 4, the model standard deviations of the three states vary in similar patterns as for the three subgroups in FMR (Table 2). Similar explanations can be applied here so that the three underlying states represent storm's ''static'' (state 1), ''moderate'' (state 2), and ''extreme'' (state 3) intensification. (A sensitivity test shows that, if $k = 2$, the ''static'' and ''moderate'' states are combined and, if $k = 4$, the ''static'' group is separated into a group with nearly no intensity change and a group with slight intensification. In both cases, the ''extreme'' group includes the largest intensification and deintensification cases, similar to the case of $k = 3$.)

Other estimated parameters of the response component of the MeHiM are also shown in Table 4. Similar to FMR results (Table 2), the estimated coefficients are mostly with expected signs, except for V in state 1 and RH in states 1 and 3, though these coefficients are all relatively small in absolute value. Actually, all coefficients in state 1 are quite small compared with those in states 2 and 3, and thus DV in state 1 is not sensitive to the change in the predictors. These parameters of the response model will be further examined together with the parameters of the transition model.

The estimated coefficients of the transition component of the MeHiM are shown in Table 5. Three multinomial logistic models are developed for each of three original states, and transitioning to state 3 is picked as

the baseline (all coefficients set as zero). The signs and the relative magnitude of the regression coefficients jointly determine the change of the transition probabilities as a response to the change of the covariates. Consider the coefficients for storms that are currently in state 1 as an example. The linear combination of covariates describes the log odds of the storm staying in state 1 versus transitioning to state 3 (Table 5, first row) and that of the storm transitioning to state 2 versus to state 3 (Table 5, second row). For instance, a one-unit, or one standard deviation, increase in the variable DV_p is associated with an increase in the log odds of staying in state 1 versus transitioning to state 3 in the amount of 0.329. At the same time, an increase in the variable DV_p is also associated with an increase in the log odds of transitioning to state 2 versus to state 3 in the amount of 0.501. Thus, when all other covariates are fixed, an increase in DV_p will cause storms in state 1 less likely to transit to state 3. Similarly, a one-unit increase in MPI will lead to a decrease in both the odds of staying in state 1 versus transitioning to state 3 and the odds of transitioning to state 2 versus to state 3. Thus storms are more likely to transit to state 3. In contrast, a one-unit increase in RH will lead to an increase in both of these odds and thus a decrease in probability of transitioning to state 3. In this case, the probabilities of both transitioning to state 2 and staying in state 1 will grow, with the probability of transitioning to state 2 growing more rapidly than that of staying in state 1.

The effect on DV of the change in each covariate depends on both the transition probabilities and response distributions; this is particularly true for the MeHiM, where the two model components are coupled.

TABLE 5. Coefficients of the MeHiM transition model.

		Intercept	DV_p	V	MPI	SHR	RH	OCN
From state 1	To state 1	4.696	0.329	−1.095	−2.907	0.388	0.079	−0.105
	To state 2	4.564	0.501	−1.200	−3.176	0.334	0.242	−0.215
	To state 3	0	0	0	0	0	0	0
From state 2	To state 1	1.840	−0.887	−2.824	−1.206	−0.172	0.099	−0.506
	To state 2	3.513	−0.907	−1.549	−1.430	−0.298	0.385	−0.176
	To state 3	0	0	0	0	0	0	0
From state 3	To state 1	−1.327	−0.274	0.134	0.340	0.079	−0.520	−0.599
	To state 2	−0.315	−0.247	−0.317	−0.022	0.391	−0.101	−1.116
	To state 3	0	0	0	0	0	0	0

Thus, one should combine Tables 4 and 5 to examine the environment's control on TC intensity change. We first consider the two storm-related variables DV_p and V. As the coefficients in front of DV_p in the response model are all positive (Table 4), increases in DV_p always lead to increases in DV, as expected. Specifically, for storms that are originally in states 2 and 3 (Table 5), an increase in DV_p will result in an increase in the probability of staying in or transitioning to state 3, where the storm can significantly grow, as the response model coefficient in front of DV_p is relatively large (0.398). For storms that are originally from state 1 (storms that keep their intensity nearly as constant, with the response model coefficient in front of DV_p as low as 0.009), an increase in DV_p leads to a greater chance in transitioning to state 2 (Table 5), where storms will start to develop more quickly (with the response model coefficient in front of DV_p as high as 0.612). Thus, the effect of DV_p (storm inertia) is always significant, even though the response model coefficient in front of DV_p appears to be low for state 1. The response model coefficient in front of V is negative for states 2 (-0.183) and 3 (-0.424), as expected, but it is marginally positive for state 1 (0.004). However, with larger V, storms that are in states 1 and 2 are more likely to transit to state 3 (Table 5). Only storms that are originally in state 3 are more likely to transit to state 1, where V does not have an expected negative effect on the storm's intensification (possibly due to statistical compensation).

The model coefficients for the four environmental variables can be examined in a similar way. As the MPI increases, storms that are originally in states 1 or 2 are more likely to enter state 3 (Table 5) and hence will intensify more rapidly due to the larger, positive coefficient for state 3 (0.521 vs 0.097 and 0.012) in the response model (Table 4). However, for storms that are originally in state 3, a larger MPI may lead to the storm's transitioning to state 1, where the storm will achieve a smaller intensification. This mechanism may prevent storms from continuously intensifying at a large rate. In contrast, as SHR increases, storms that are originally in state 2 are more likely to enter state 3, and hence will deintensify more rapidly due to the larger, negative coefficient for state 3 (-0.563 versus -0.070 and -0.003) in the response model. Storms that are originally in state 1 are more likely to stay in state 1. Storms that are originally in state 3 are more likely to transit to states 1 or 2, where the storm will achieve a relatively smaller deintensification. This mechanism may prevent storms from continuously deintensifying at a large rate. The response model coefficient in front to RH is positive (0.082) as expected for state 2, but it is moderately negative for states 1 (-0.008) and 3 (-0.091). However,

as RH increases, storms in states 1 and 2 will be more likely to transit to state 2 (than to state 1; Table 5). Only storms in state 3 will be more likely to be in state 3, where the unexpected, negative response in DV is likely due to statistical compensation to the large responses to other variables. The response model coefficients in front of OCN are all positive, with the coefficient for state 3 (0.336) larger than states 2 (0.028) and 1 (0.014). As OCN increases (ocean weakening effect decreases), storms at all states are more likely to enter state 3 (Table 5) and intensify more rapidly. The model seems to capture the expected large effect of ocean cooling on storm intensification, which will be discussed further in section 4.

Finally, we consider the overall effects of joint changes in the covariates. We estimate the transition probabilities for the three specific scenarios considered in the FMR classification analysis—all variables are at the 95th percentile of their samples that are unfavorable for larger DV, all variables are at their medians, and all variables are at the 95th percentile that are favorable for larger DV—as shown in Fig. 2. We find that storms in state 1 are around 50% likely to stay in state 1 and around 50% likely to transit to state 2, almost regardless of the environment. In other words, storms in state 1 are likely to stay in state 1 (with small DV response to the environment) or move to state 2 (with moderate DV response to the environment). This mechanism makes weak storms develop slowly. On the contrary, storms in states 2 and 3 respond to the environment in different directions. Moving from unfavorable conditions to favorable conditions, storms in state 2 are more and more likely to transit to state 1 (with probability changes from 3% to 23% to 30%), and less and less likely to stay in state 2 (96%–75%–68%). Storms that are originally in state 3 are significantly more and more likely to stay in state 3 (4%–60%–85%), as the environment changes from unfavorable to favorable conditions. Thus, although the probability of entering state 3 from states 1 or 2 is always low (0%~2%), once the storm enters state 3, it is likely to stay in state 3 and continuously achieve rapid intensification if the environment is favorable. This "lock-in" mechanism supports the development of intensity extremes, which is a significant advantage of the MeHiM over FMR. On the other hand, when the environment is unfavorable, the storms are not likely to enter state 3 or stay locked in state 3, and thus the model does not support much the development of rapid deintensification, similar to FMR.

d. Land decay model

TCs weaken quickly over land mainly due to the lack of fuel sources and the entrainment of dry air.
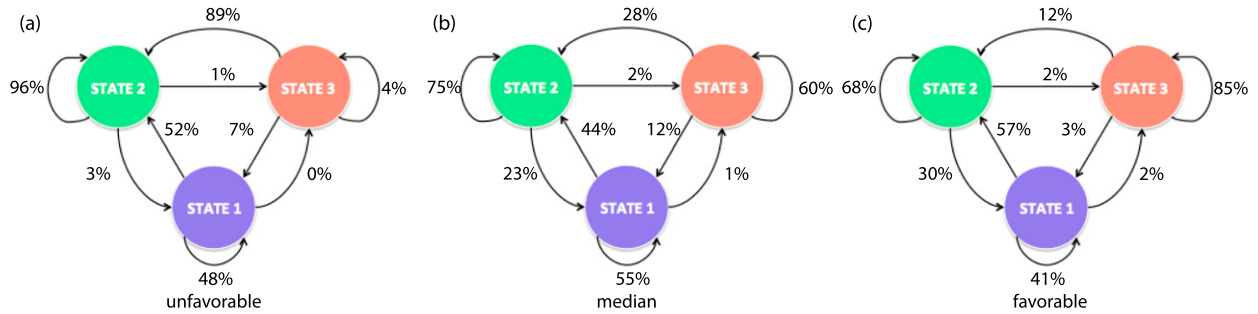
FIG. 2. Examples of MeHiM transition probabilities when all covariates are (a) at 95th percentile of their samples that are unfavorable for larger DV, (b) at their medians, and (c) at 95th percentile of their samples that are favorable for larger DV. Variables in (a) are $DV_p = -9.14$ kt $(6\,h)^{-1}$, $V = 94.4$ kt, MPI = 24.48 kt, SHR = 41.2 kt, RH = 37.73%, and OCN = 0.398. Variables in (b) are $DV_p = 0.10$ kt $(6\,h)^{-1}$, $V = 34.4$ kt, MPI = 141.0 kt, SHR = 13.9 kt, RH = 50.0%, and OCN = 0.933. Variables in (c) are $DV_p = 9.61$ kt $(6\,h)^{-1}$, $V = 15.3$ kt, MPI = 161.6 kt, SHR = 1.3 kt, RH = 69.5%, and OCN = 0.999.

Estimating the decay of storms after making landfall or when moving across small islands is also a key component of TC intensity modeling. Georgiou (1985) modeled the increase of central pressure proportional to the distance from the landfall point. DeMaria and Kaplan (1995) modeled the decrease in the maximum wind speed as an exponential decay function of time after making landfall. DeMaria et al. (2006) proposed a similar model and applied it particularly for storms moving over narrow landmasses. The decay models of DeMaria and Kaplan (1995) and DeMaria et al. (2006) have been applied to adjust operational forecasting, and the verification on SHIPS has shown that the inclusion of such decay models can improve intensity forecasts in various basins (DeMaria et al. 2006).

Here, we develop a land model based on the same dataset (storms of 1979–2014) applied for the overocean intensity models. The model structure is similar to that in DeMaria and Kaplan (1995), except that we set the reduction factor to be 1 for simplicity, neglecting the assumed sudden intensity decrease when storms approach land. In this model, the storm's intensity V exponentially decays as a function of time t, as shown in Eq. (8):

$$V(t) = V_b + (V_0 - V_b)e^{-\alpha t}, \qquad (8)$$

where V_0 is the storm's original intensity prior to landfall, which decays with a constant temporal rate α, to a constant background intensity V_b that a storm may maintain over land.

To estimate α and V_b in this model, we collect all TC intensity sequences when the storm's center is over land. While most of these segments are quite short, with an average length of about 5 observations, we exclude segments that are shorter than 2 observations (12 h). Also, as the model does not apply for storms weaker

than V_b, we assume the storms will not decay if the landfall intensity is smaller than V_b. Accordingly, we exclude segments with V_0 smaller than 20 kt with trial and error. The model parameters α and V_b are then fitted based on 229 segments, and the optimal decay rate and background intensity are estimated to be 0.049 kt $(6\,h)^{-1}$ and 18.82 kt, respectively, by nonlinear least squares regression. Figure 3 shows model predicted intensity decay over land compared with the observation for storms with various initial intensities. The model can capture the intensity decay well for storms with relatively small landfall intensities ($V_0 < 70$), but it may underestimate the decay for more intense storms (Fig. 3).

3. Model evaluation

To evaluate the developed models (based on 332 historical storms during 1979–2014), we perform Monte Carlo simulations of the intensity evolution over storms' lifetime for all 555 historical storms during 1979–2014 and compare model simulations with observations. Here the removed surface background wind is added back to the estimated storm intensity in order to compare with the observed maximum wind. We examine the models' capability to reproduce the observed TC intensity metrics, including spatial distribution of intensity over the basin, probabilistic distributions of intensification rate, LMI, and landfall intensity, as well as temporal evolution of intensity for sample cases. We also evaluate the models using nonoverlapping training (271 storms) and testing (284 storms) datasets; the results are similar and shown in the supplemental material.

a. Monte Carlo (MC) simulation

For all simulations, the storm is initialized by the first two steps of observed intensity to obtain initial V and
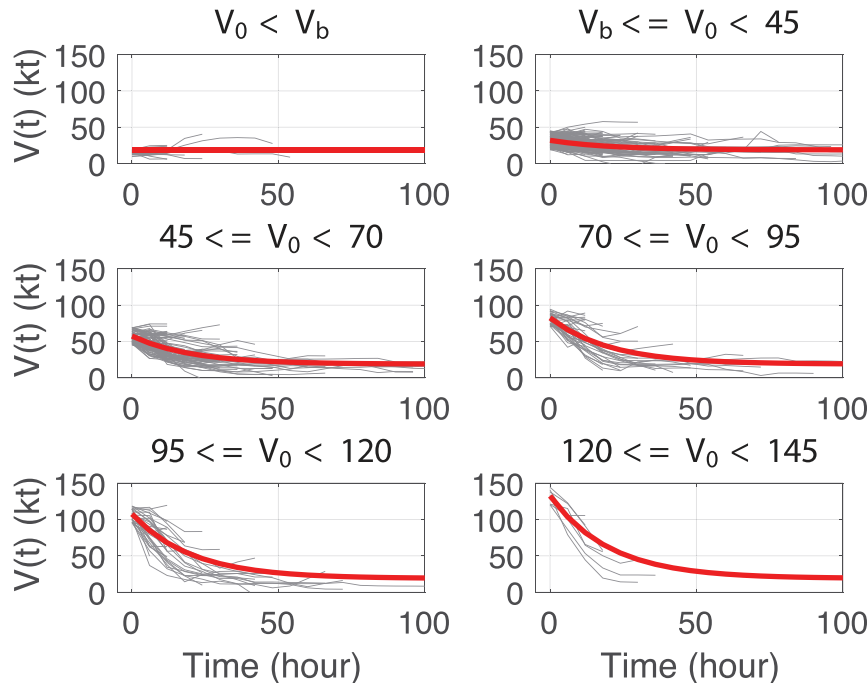
FIG. 3. Comparison of land decay model (red curves) and observed intensity evolutions (gray curves) for different initial intensity ranges. The modeled decay is calculated with an initial intensity equal to the midpoint value of the intensity range.

DV_p. Then, for the OLS, the intensity change for each step is obtained by sampling from a Gaussian distribution [Eq. (2) and Table 1]. For the FMR, the classification model is first applied to compute the group membership probabilities [Eq. (4) and Table 3], based on which group membership is determined with the discrete acceptance–rejection method. The intensity change is then obtained by sampling the Gaussian distribution for the selected group (Table 2).

Simulation with the MeHiM has one more step in the initialization. Given the environmental parameters (i.e., MPI, SHR, and RH) at the storm's occurrence, we use the initial state model with the acceptance–rejection method to determine the state for the first two steps. Then, we continue the simulation by repeatedly choosing the next state to visit according to the transition probabilities [Eq. (7) and Table 5] with the discrete acceptance–rejection selection and updating storm intensity with the intensity change sampled from the state's corresponding response distribution [Eq. (6) and Table 4].

During all simulations, the land model is applied after the storm's center hits land. The intensity change when moving from ocean to land is computed as an inverse-distance-weighted average of the predictions from the land model and the over-the-ocean intensity model. Storms with landfall intensity V_0 smaller than the background intensity V_b are assumed to keep their landfall intensity over land, while storms with V_0 greater than V_b are simulated to decay according to the filling model [Eq. (8)]. The over-the-ocean model is resumed when the storm moves back to the ocean from land. For the MeHiM, when the storm moves back to the ocean, its initial state is assumed to be the same as the state before it last made landfall. The simulation stops at the end of the historical record of the track or when the storm' intensity becomes lower than 10 kt. Each historical storm is simulated 100 times with each of the three intensity models, and the comparisons with observations are based on all realizations.

b. Spatial distributions

Figure 4 displays observed and MeHiM-modeled TC intensity along historical tracks. The basin is divided as a $2° \times 2°$ grid and, as an example, the 75th and 50th percentiles of the intensity values in each grid from the observations are compared with those from simulations (median over the 100 realizations). The simulations capture relatively well the spatial variation of TC intensity, although they appear smoother and seem to underestimate the most extreme intensities in lower latitudes (occurring over the Caribbean Sea and the Gulf of Mexico) and overestimate the intensities in midlatitudes especially on land (due to the limitation of
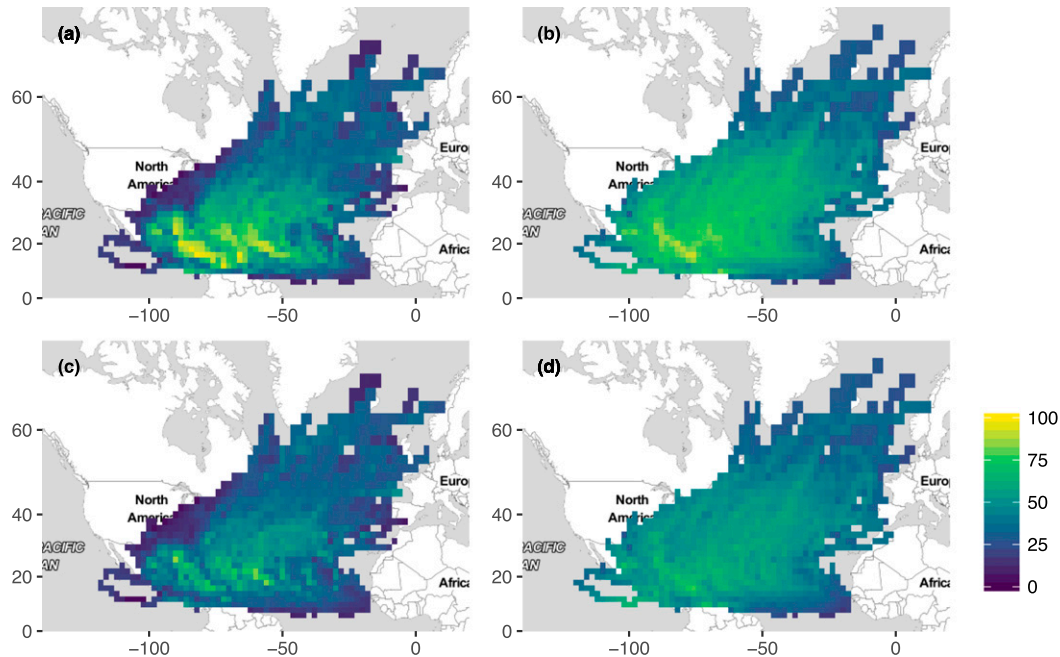
FIG. 4. Comparison of observed and MeHiM-modeled TC intensity (kt; color) in each 2° × 2° grid. The intensity metrics plotted are (a) 75th percentile of observations, (b) median (over 100 realizations) of 75th percentile of MeHiM simulations, (c) 50th percentile of observations, and (d) median of 50th percentile of MeHiM simulations.

the land decay model). Nevertheless, the MeHiM improves over the OLS and FMR in capturing the extremes (not shown here; see other comparisons below).

c. Probability distributions

In this subsection, we investigate the models' capability in capturing the probability distributions of various key intensity and intensity change metrics. First, a comparison of the histograms of observed (IBTrACS) and simulated 6-h intensity change rate (6-h DV) is shown in Fig. 5. Here observations and simulations over land are removed to focus on the comparison between the three over-the-ocean intensity models. All three models capture the general features of the observed distribution. However, simulations from the OLS have 6-h DV fall within a relatively narrow range from −29 to 23 kt. The FMR improves over the OLS to a range from −40 to 33 kt. The MeHiM generates a larger range of 6-h DV, from −54 to 41 kt, covering nearly the full range of observations except the outlier. IBTrACS has a range from −30 to 55 kt for 6-h DV, and the outlier at the right tail of IBTrACS (Fig. 5b) comes from one observation of 6-h DV of 55 kt during the life cycle of Hurricane Wilma (2005), one of the most intense storms ever recorded in the Atlantic basin. MeHiM did not capture this outlier of extreme intensification, and it generated larger extremes of deintensification compared to observations (Fig. 5b).

We also compare the histograms of 24-h intensity change (24-h DV) between observations and simulations (Fig. 6). The distribution from MeHiM simulations is wider than that from the OLS and FMR and compares better with IBTrACS. The simulated range of 24 h DV is from −78 to 73 kt for the OLS, from −104 to 85 kt for FMR, and from −101to 103 kt for MeHiM, compared to a range from −110 to 95 kt for IBTrACS. We find that although both the FMR and MeHiM are able to simulate extreme 24-h DV, simulations from FMR have lower probability densities than IBTrACS in most bins that are larger than 35 kt. MeHiM can better capture the large intensity changes. Particularly, although the MeHiM does not capture the outlier extreme 6-h intensification (Fig. 5b), it produces higher probability densities than IBTrACS for most bins of 24-h DV greater than 30 kt (i.e., RI; Fig. 6b). On the other hand, although the MeHiM overestimate the extreme 6-h deintensification (Fig. 5b), it underestimates the rates of most extreme 24-h deintensifications (Fig. 6b). The large 24-h intensifications are likely generated by the "lock-in" mechanism in the MeHiM, which supports continuous large intensification once the storm enters the extreme state and when the environment is favorable. The MeHiM lacks such a mechanism to support continuous large deintensification, as discussed earlier.

Next, we compare the probability distribution of LMI in the simulations and observations (Fig. 7). Among the
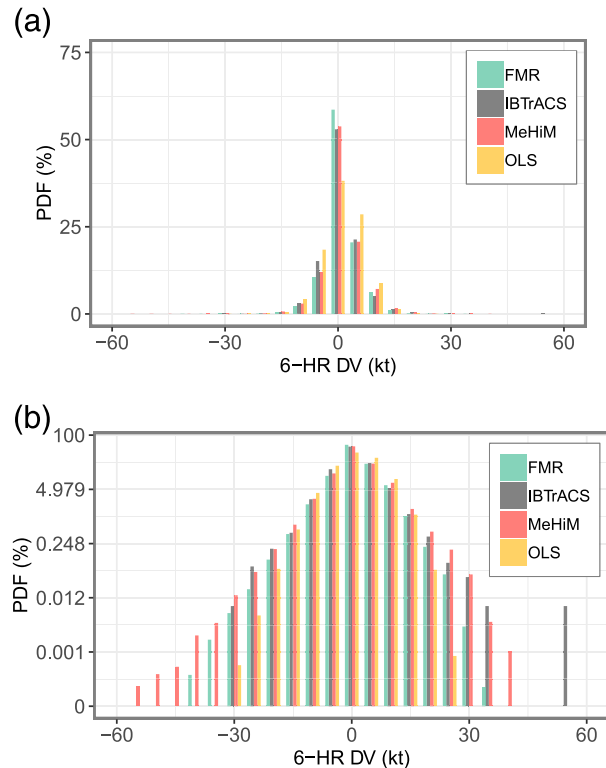
(a)

(b)

FIG. 5. Comparison of histograms of observed (gray) and modeled (OLS in yellow, FMR in green, and MeHiM in red) 6-h intensity change in (a) linear scale and (b) logarithmic scale. Overland points are removed. Data are binned in 10-kt intervals. Simulation results include 100 realizations.

three models, the MeHiM compares with IBTrACS the best. It captures the shape and peak of the LMI distribution quite well, although it is smoother and misses the "fluctuations and bumps" in the tail, and as a result it underestimates the extreme LMIs to some extent. The OLS and FMR estimated LMI distributions have significantly thinner tails. Also, FMR-estimated distribution has a substantially higher peak, and OLS-estimated simulation has a significant right shift in the peak, compared to IBTrACS and the MeHiM. Thus, the OLS and FMR are likely to underestimate the LMI for extreme storms and overestimate the LMI for relatively weak storms.

To further investigate this matter, we examine the LMI distribution for non-RI and RI storms separately. The percentages of simulated RI storms by the OLS, FMR, and MeHiM are 24.7%, 20.1%, and 26.1%, respectively, while in IBTrACS this fraction is 30.5%. Thus, the MeHiM improves over the OLS and FMR in simulating the right fraction of RI storms. Furthermore, as shown in Fig. 8, the two LMI distributions of the OLS simulations are relatively close, so that the peak in the LMI distribution of non-RI storms is shifted to the right
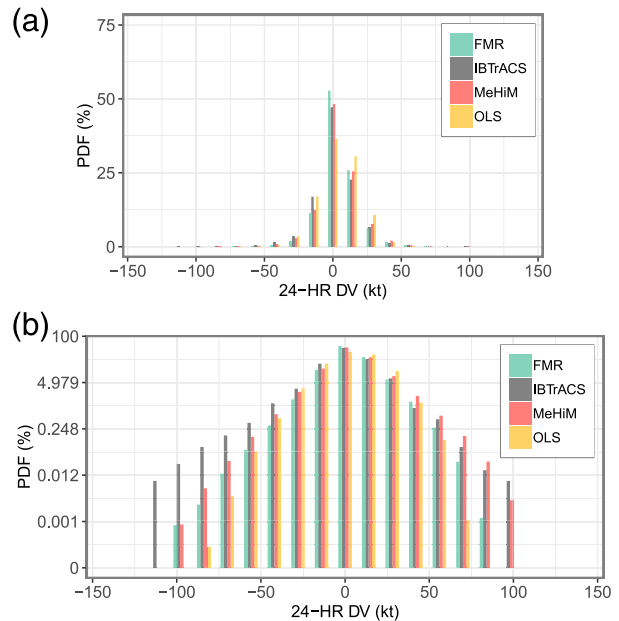
(a)

(b)

FIG. 6. As in Fig. 5, but for 24-h intensity change. Data are binned in 15-kt intervals.

while the peak in the LMI distribution of RI storms is shifted to the left, compared to IBTrACS. Therefore, the OLS tends to overestimate the LMI of non-RI storms and underestimate the LMI of RI storms. FMR has a similar right shift in the LMI distribution for non-RI storms and it overestimates the peak. Also, similar to the OLS, the LMI distribution for RI storms in FMR is left shifted and featured with a much thinner tail compared to IBTrACS. Thus, FMR also tends to underestimate the LMI for RI storms. Significantly improving over the OLS and FMR, the MeHiM captures the LMI distribution of non-RI storms very well. The MeHiM also captures the shape of the LMI distribution for RI
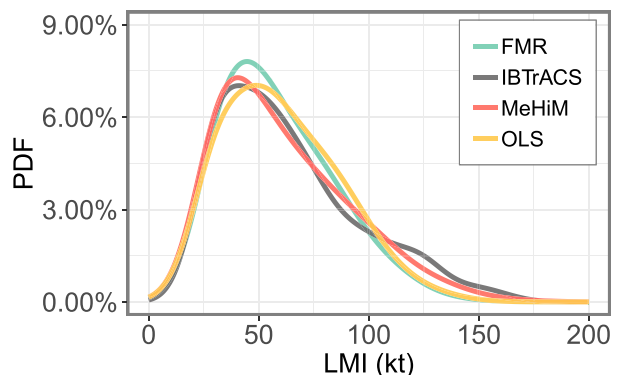

FIG. 7. Comparison of PDF of lifetime maximum intensity between observations (gray) and model simulations (OLS in yellow, FMR in green, and MeHiM in red). Overland points are removed. Simulation results include 100 realizations.
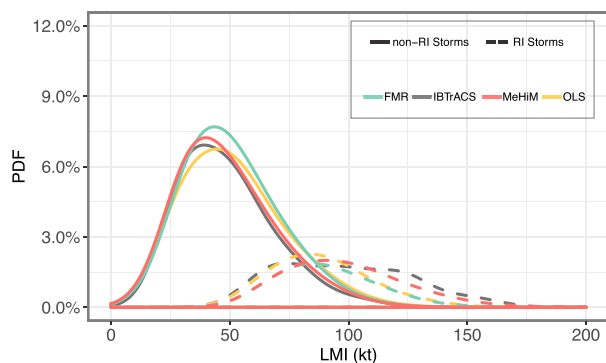
FIG. 8. As in Fig. 7, but for RI storms (intensity increases by 30 kt or more over 24 h; dashed curves) and non-RI storms (solid curves).

storms much better, although the most extreme LMIs for RI storms are still underestimated.

Comparison between Figs. 7 and 8 indicates that the tail of the LMI distribution is largely contributed by RI storms, as noted by Lee et al. (2016a), and thus the statistical intensity models underestimate the tail of the LMI distribution mainly because they underestimate the LMI of extreme RI storms. Given that 80% of major TCs (LMI > 96 kt) underwent RI historically, it is important to capture the RI distribution in order to capture the LMI distribution. Being able to capture the distribution of RI extremes, however, does not imply being able to capture the tail of the LMI distribution. Improving over the OLS and FMR, the MeHiM captures well the tail of the 24-h DV distribution representing RI (Fig. 6b), but it still underestimates the tail of the LMI distribution (Figs. 7 and 8). This underestimation of extreme LMIs also explains the underestimation of spatial extremes in Fig. 4.

Next, we compare the simulated and observed probability distribution of the landfall intensity for the North Atlantic coastline and its divisions of U.S. Northeast, Southeast, and Gulf of Mexico regions (Fig. 9). All three models can capture the general shape of the landfall probability distribution, but the MeHiM better captures the tail of the distribution. Also, MeHiM has smaller uncertainty bonds, compared to OLS, and the bonds better contain the IBTrACS estimates, compared to FMR. It is noted that for the northeastern U.S. coastline, the distribution from IBTrACS is quite rough and the percentile band of the model estimates is relatively large, due to data limitations for this region.

d. Sample cases

Apart from capturing various statistical distributions, we compare the performances of the OLS, FMR, and MeHiM in simulating the intensity evolution of individual storms. Figures 10a, 10d, and 10g show as an example the intensity

evolution of a typical TC, Hurricane Richard in 2010, simulated by the OLS, FMR, and MeHiM, compared to the IBTrACS observation. All three models can capture the main features of the storm's development at the early stage, although the large dissipation at the end is not well captured due to the deficiency of the simple land decay model. The ensemble mean (over the 100 realizations) from each model compares relatively well with the observation. The MeHiM performs significantly better as the LMI of the ensemble mean shifts toward the observed LMI, compared to the OLS and FMR.

The intensity simulations for an extreme RI storm, Hurricane Wilma (2005), are shown in Figs. 10b, 10e, and 10h. Hurricane Wilma experienced the most intense rapid intensification (95 kt in 24 h) over the period of 1979–2014 in the North Atlantic basin and reached an LMI of 160 kt. The LMI of the ensemble mean of the models is significantly lower than the observation, but the MeHiM performs better. This is somewhat expected given the models' underestimation of the tail of the LMI distribution (see Figs. 7 and 8). However, even for such an extreme case, we find that the ensemble of the MeHiM is able to cover the observed LMI—the maximum (over the 100 realizations) of the simulated LMI is 175 kt from the MeHiM, while it is 143 kt from OLS and 150 kt from FMR. In addition, more than 40% of the realizations from the MeHiM can reach a LMI over 100 kt, which is much more than that from OLS and FMR (~15%). Indeed, an upward shift over the early development stage in the MeHiM realizations (indicated by the shadings) indicates a larger simulated intensification rate in the MeHiM compared to the OLS and FMR. Also, MeHiM captures well the second intensity peak and the following rapid deintensification.

In Figs. 10c, 10f, and 10i, an unnamed tropical depression in 2003 is shown as an example of weak storms with LMI less than 45 kt. In this case, both FMR and the MeHiM are capable of simulating the intensity evolution, with the ensemble mean matching the observation. OLS, on the other hand, significantly overestimates storm's intensity, with mean as well as the 20th–80th percentile envelope shift upward away from the observation. This overestimation by OLS is typical for weak storms, as discussed before (see Figs. 7 and 8).

From the above results, we find that both FMR and the MeHiM are able to simulate relatively weak storms, while OLS tends to overestimate weak storms. The MeHiM has significant advantage over both FMR and OLS when simulating RI storms or storms with large intensification. It is also noted that the models tend to overestimate the intensity after making landfall for intense storms, which indicates that the simplified land model may not be sufficient to simulate a storm's extreme decay over land. These conclusions hold when the
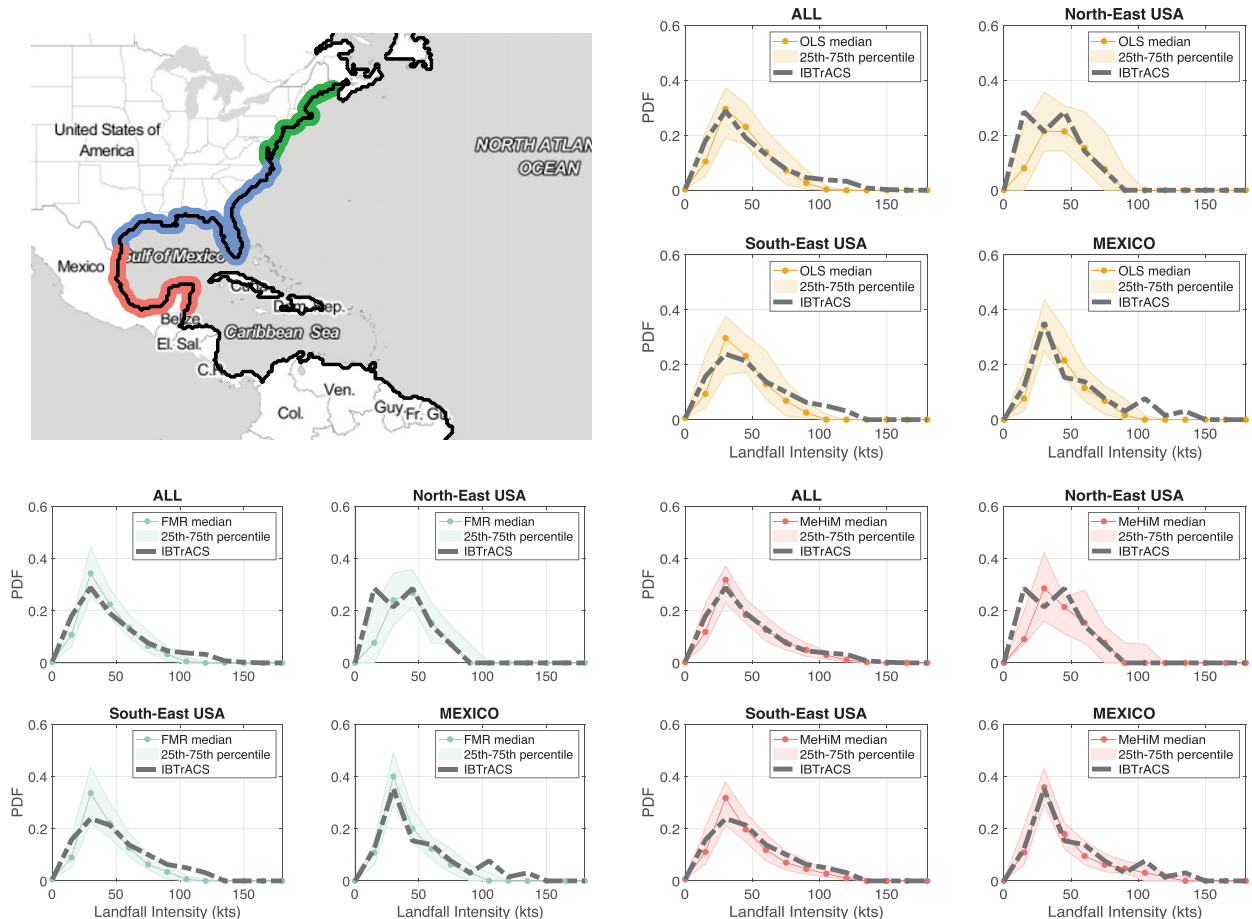
FIG. 9. Comparison of observed (thick dashed curves) and modeled (thin solid lines and shadings; yellow for OLS, blue for MFR, and red for MeHiM) probability distributions of TC landfall intensity for (a) North Atlantic coastline (green, blue, and red segments shown on the map), (b) northeastern United States from Maine to Virginia (green segment), (c) southeastern United States from North Carolina to Florida plus Gulf Coast of the United States (blue segment), and (d) Gulf Coast of Mexico (red segment). The 25th–75th percentile uncertainty bonds of the simulations (shadings) are estimated from 100 realizations.

models are trained and tested using nonoverlapping datasets. The results for the models trained on 271 TCs during 1979–2014 and tested on the remaining 284 TCs during 1979–2014 are shown in the supplemental material. Obtained distributions of 6-h intensity change (Fig. S1 in the online supplemental material), 24-h intensity change (Fig. S2), LMI (Fig. S3), and LMI conditioned on the RI status (Fig. S4) are similar to those shown in Figs. 5–8, respectively. Testing on sample cases that are out of the training dataset is also further discussed in the supplemental material (Fig. S5).

4. Discussion

a. Effect of the ocean feedback variable

The three (overocean) intensity models are developed with four environmental covariates including three

atmospheric and one ocean feedback variable. As OCN is not used in previous statistical models, here we test its significance and examine how OCN represents TC–ocean interaction. We use two storms as examples: Tropical Storm Larry (2003) and Hurricane Katia (2011). The upper two panels in Fig. 11 show simulated intensity evolution for these two storms from the original MeHiM (as discussed above), and the lower two panels show simulations from the MeHiM without OCN (rebuilt with OCN removed). The MeHiM estimation compares well with the observation, while the MeHiM without OCN overestimate the intensity for Tropical Storm Larry. In this case, OCN played a significant role in preventing the simulated storm from intensifying too fast, given that Larry was a slow-moving storm. It is more interesting to note, however, that including OCN may also better simulate extreme intensification. For Hurricane Katia, an RI storm in 2011, simulations from the MeHiM with OCN are
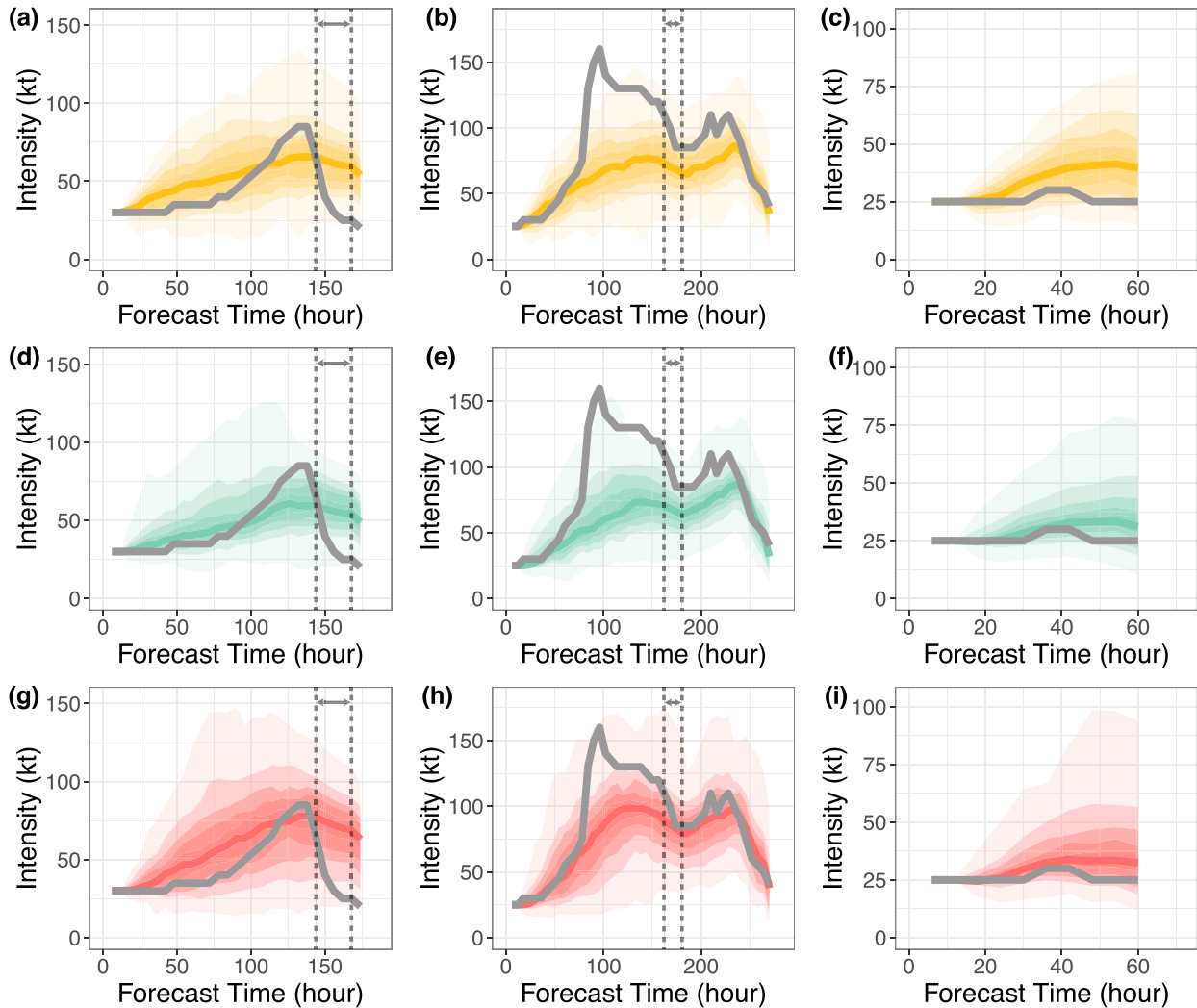
FIG. 10. Comparisons of observed (gray curve) and simulated intensity evolution for (a),(d),(g) Hurricane Richard (2010), (b),(e),(h) Hurricane Wilma (2005), and (c),(f),(i) an unnamed tropical depression in 2003. The simulated intensity evolutions from OLS (yellow) are shown in (a)–(c); simulations from FMR (blue) are shown in (d)–(f); and simulations from MeHiM (red) are shown in (g)–(i). The colored shading boundaries represent the deciles of 100 realizations while the solid colored curve is the mean over the 100 realizations. The vertical dashed lines highlight the period of observations that are over land, during which the land model is applied. Overland observations that are less than two steps (12 h) are not highlighted.

better in terms of capturing the rapid RI phase, compared to the simulations from the MeHiM without OCN. In this case, the relatively weak effect of the ocean feedback during RI is captured by OCN. Thus, adding OCN does not always make storms weaker in a statistical model, and it may actually help to simulate storm's rapid intensification.

b. Effect of extreme state identification

The MeHiM performs well for simulating normal and relatively weak storms but may underestimate extremes (as shown in Figs. 7, 8, and 10). This underestimation of a storm's LMI is often due to the underestimation of rapid

or extreme intensification, which in turn is likely due to the limitation in capturing the underlying extreme state of intensity change [i.e., underestimating the probability of entering the extreme state, as discussed in Lin et al. (2017) for the FMR analysis]. Here we investigate this limitation in the MeHiM by experimenting on four extreme RI storms with maximum intensification rate ranging from 30 to 85 kt within 24 h. For each of these storms, we run the MeHiM freely until the storm experiences rapid intensification, when the occurrence of RI is indicated by IBTrACS. At the time of RI, the underlying state of intensity change is manually fixed to the extreme state (regardless of the transition probability in
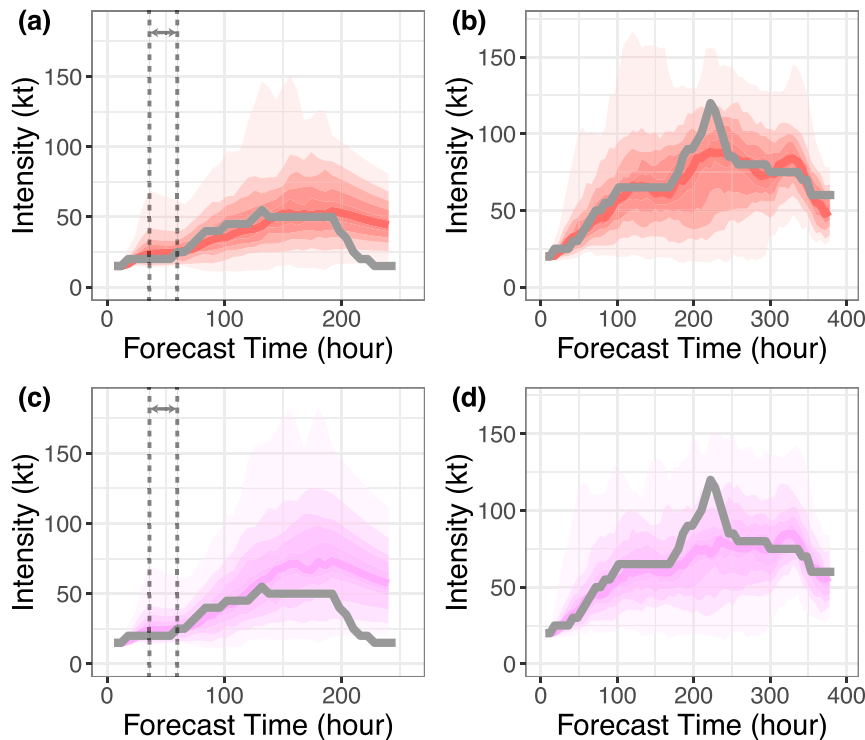
FIG. 11. Comparisons of MeHiM simulated intensity evolution with (red) and without (purple) ocean parameter. The observation is shown by the gray curve. Selected storms are (a),(c) Tropical Storm Larry (2003) and (b),(d) Hurricane Katia (2011). The vertical dashed lines highlight the period of observations that are over land, during which the land model is applied.

the model) during the next four steps (24 h, consistent with the definition of RI). Thereafter, the model is resumed to run freely. The results are shown, in comparison with the original MeHiM runs, in Fig. 12, where the black arrow indicates the occurrence of RI in each case. We find that the simulations improve significantly with the correction of the underlying states at the occurrence of RI. This is especially the case for Hurricane Andrew (1992). The MeHiM performs well in simulating the first part of the storm's intensity evolution, but it fails to capture the RI and thus greatly underestimates the peak intensity. With state correction, the simulations are shifted greatly toward observation and thus well capture the observed LMI. (The rapid decay after landfall is again not well captured for these extreme cases due to the limitations of the land model. Further studies may include combining MeHiM with a more advanced land model, e.g., with decay rate modeled as a function of initial intensity prior to landfall.) We also evaluate the MeHiM's performance and limitation in simulating three recent destructive hurricanes in the 2017 Atlantic hurricane season, that is, Hurricanes Harvey, Irma, and Maria, which are all RI storms and out of the training

dataset, as shown in the supplemental material. Similar to the cases in Fig. 12, the LMI of these extreme RI storms are significantly underestimated by the MeHiM, but the simulations from the MeHiM with state correction improve greatly.

This experiment confirms that although the MeHiM improves over FMR in capturing the extreme state, it may still underestimate the likelihood of extreme states. Future research may focus on improving the state transition probability modeling of the MeHiM, that is, applying more sophisticated classification models than the simple logistic regression. This analysis also indicates that the MeHiM may have a great potential for real-time operational forecasting, since, unlike a homogenous model such as the OLS (and similarly SHIPS), the state of intensity change in the MeHiM can be manipulated through incorporating/assimilating real-time information on the RI state, for example, the probabilistic guidance RII (Kaplan et al. 2010).

5. Summary

In this study, we have developed a Markov environment-dependent hurricane intensity model, or the MeHiM,
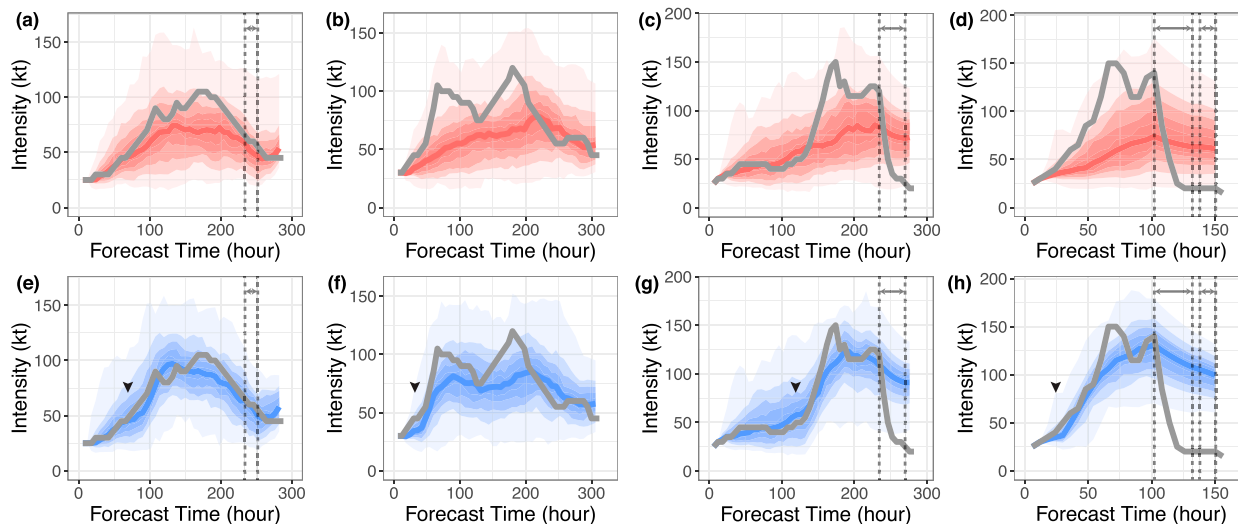
FIG. 12. Comparison of MeHiM simulated intensity evolution without (red) and with (blue) RI state correction. The observation is shown by the gray curve. Selected four RI storms are (a),(e) Irene (1981); (b),(f) Isaac (2000); (c),(g) Andrew (1992); and (d),(h) Felix (2007), from left to right. The black arrow indicates the onset of RI for each storm. The vertical dashed lines highlight the period of observations that are over land, during which the land model is applied.

based on the North Atlantic TC records from IBTrACS WMO archive and environmental parameters derived from ERA-Interim reanalysis data over the period of 1979–2014. The model considers three hidden discrete states, and each state is associated with a probability distribution of intensity change. The movement of storm intensity change from one state to another is considered as a Markov chain described by a transition probability matrix. Both the intensity change and state transition components of the model are dependent on environmental variables (including potential intensity, vertical wind shear, high-level relative humidity, and ocean feedback) and storm variables (including current intensity and previous intensity change). All the variables, including the ocean feedback variable, which is applied in statistical modeling for the first time, are found to be significant. Similar to the FMR model discussed in Lin et al. (2017), the MeHiM's three states of intensity change turn out to represent the storm's slow, moderate, and rapid intensity change, respectively. The MeHiM improves over FMR as it accounts for the temporal correlation of the states of intensity change and integrates the intensity change and state identification components in a Markov framework. As a result, the MeHiM significantly improves over FMR in better capturing the extreme state and simulating rapid and continuous intensification.

We evaluate the MeHiM by comparing simulated historical storms with observations, where the influence of land is included through a simple land decay model. We find that the MeHiM shows a great improvement

over previous statistical models (such as the OLS and FMR, built as baseline models in this study) in simulating TC intensity climatology. The probability distributions of TC intensity change rates (6- and 24-h DV) and LMI simulated by the MeHiM are closer to the observations. With the capability in simulating category 4 and 5 storms, the MeHiM can better capture the tail of the LMI distribution, although it still slightly underestimates the tail. The MeHiM is limited in simulating the most extreme LMIs, even though it can well capture the tail of the 24-h DV distribution representing RI, indicating that a mechanism of continuous intensification beyond the temporal window of RI (i.e., 24 h) exists but is not well captured by the model. The MeHiM also improves over previous models in reproducing the observed distribution of landfall intensity for regions with relatively adequate data. The model can also be applied to provide an advisory to areas with limited historical data.

The MeHiM improves over the OLS (which is similar to SHIPS) and FMR in probabilistically forecasting the temporal evolution of intensity for individual storms, although the MeHiM still may not capture a storm's RI, especially for storms that strengthen extremely in a short time period. Nevertheless, we find that, when combined with an extreme state correction for RI, the forecasting can be greatly improved for most historical RI storms. This result indicates the potential to significantly improve the MeHiM through improving its state transition probability model to better capture the extreme state. This result also indicates a potential application of the

MeHiM combined with a probabilistic RI guidance in real-time operational forecasting.

Future studies may also include a comparison between the MeHiM and dynamic models, such as the Coupled Hurricane Intensity Prediction System (CHIPS; (Emanuel et al. 2004) and its simplified algorithm (Emanuel 2017). The ultimate goal of our study is to integrate the intensity model with a genesis model and a track model into a complete hurricane climatology modeling system, similar to Emanuel et al. (2008) and Lee et al. (2018).

REFERENCES

Balaguru, K., G. R. Foltz, L. R. Leung, E. D. Asaro, K. A. Emanuel, H. Liu, and S. E. Zedler, 2015: Dynamic potential intensity: An improved representation of the ocean's impact on tropical cyclones. *Geophys. Res. Lett.*, **42**, 6739–6746, https://doi.org/10.1002/2015GL064822.

Balmaseda, M. A., K. Mogensen, and A. T. Weaver, 2013: Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quart. J. Roy. Meteor. Soc.*, **139**, 1132–1161, https://doi.org/10.1002/qj.2063.

Bender, M. A., I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, 2010: The operational GFDL coupled hurricane–ocean prediction system and a summary of its performance. *Mon. Wea. Rev.*, **135**, 3965–3989, https://doi.org/10.1175/2007MWR2032.1.

Bister, M., and K. A. Emanuel, 1998: Dissipative heating and hurricane intensity. *Meteor. Atmos. Phys.*, **65**, 233–240, https://doi.org/10.1007/BF01030791.

——, and ——, 2002: Low frequency variability of tropical cyclone potential intensity. 1. Interannual to interdecadal variability. *J. Geophys. Res.*, **107**, 4801, https://doi.org/10.1029/2001JD000776.

Camargo, S. J., and A. A. Wing, 2016: Tropical cyclones in climate models. *Wiley Interdiscip. Rev.: Climate Change*, **7**, 211–237, https://doi.org/10.1002/wcc.373.

Cappé, O., E. Moulines, and T. Ryden, 2006: *Inference in Hidden Markov Models.* Springer Science & Business Media, 653 pp.

de Boyer Montégut, C., G. Madec, A. S. Fischer, A. Lazar, and D. Iudicone, 2004: Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *J. Geophys. Res.*, **109**, 1521, https://doi.org/10.1029/2004JC002378.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system.

Quart. J. Roy. Meteor. Soc., **137**, 553–597, https://doi.org/10.1002/qj.828.

DeMaria, M., and J. Kaplan, 1994: A Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic Basin. *Wea. Forecasting*, **9**, 209–220, https://doi.org/10.1175/1520-0434(1994)009<0209:ASHIPS>2.0.CO;2.

——, and ——, 1995: A simple empirical model for predicting the decay of tropical cyclone winds after landfall. *J. Appl. Meteor.*, **34**, 2499–2512, https://doi.org/10.1175/1520-0450(1995)034<2499:ASEMFP>2.0.CO;2.

——, and ——, 1999: An updated Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **14**, 326–337, https://doi.org/10.1175/1520-0434(1999)014<0326:AUSHIP>2.0.CO;2.

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

——, J. A. Knaff, and J. Kaplan, 2006: On the decay of tropical cyclone winds crossing narrow landmasses. *J. Appl. Meteor. Climatol.*, **45**, 491–499, https://doi.org/10.1175/JAM2351.1.

Elsberry, R. L., T. D. B. Lambert, and M. A. Boothe, 2007: Accuracy of Atlantic and eastern North Pacific tropical cyclone intensity forecast guidance. *Wea. Forecasting*, **22**, 747–762, https://doi.org/10.1175/WAF1015.1.

Emanuel, K. A., 1995: Sensitivity of tropical cyclones to surface exchange coefficients and a revised steady-state model incorporating eye dynamics. *J. Atmos. Sci.*, **52**, 3969–3976, https://doi.org/10.1175/1520-0469(1995)052<3969:SOTCTS>2.0.CO;2.

——, 2017: A fast intensity simulator for tropical cyclone risk analysis. *Nat. Hazards*, **88**, 779–796, https://doi.org/10.1007/S11069-017-2890-7.

——, C. DesAutels, C. Holloway, and R. Korty, 2004: Environmental control of tropical cyclone intensity. *J. Atmos. Sci.*, **61**, 843–858, https://doi.org/10.1175/1520-0469(2004)061<0843:ECOTCI>2.0.CO;2.

——, R. Sundararajan, and J. Williams, 2008: Hurricanes and global warming: Results from downscaling IPCC AR4 simulations. *Bull. Amer. Meteor. Soc.*, **89**, 347–368, https://doi.org/10.1175/BAMS-89-3-347.

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

Georgiou, P. N., 1985: *Design Wind Speeds in Tropical Cyclone-Prone Regions.* University of Western Ontario, 590 pp.

Kaplan, J., and M. DeMaria, 2003: Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Wea. Forecasting*, **18**, 1093–1108, https://doi.org/10.1175/1520-0434(2003)018<1093:LCORIT>2.0.CO;2.

——, ——, and J. A. Knaff, 2010: A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **25**, 220–241, https://doi.org/10.1175/2009WAF2222280.1.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS). *Bull. Amer. Meteor. Soc.*, **91**, 363–376, https://doi.org/10.1175/2009BAMS2755.1.

Knutson, T. R., and Coauthors, 2010: Tropical cyclones and climate change. *Nat. Geosci.*, **3**, 157–163, https://doi.org/10.1038/ngeo779.

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

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

——, ——, ——, and ——, 2016b: Autoregressive modeling for tropical cyclone intensity climatology. *J. Climate*, **29**, 7815–7830, https://doi.org/10.1175/JCLI-D-15-0909.1.

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

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

——, R. Jing, Y. Wang, E. Yonekura, J. Fan, and L. Xue, 2017: A statistical investigation of the dependence of tropical cyclone intensity change on the surrounding environment. *Mon. Wea. Rev.*, **145**, 2813–2831, https://doi.org/10.1175/MWR-D-16-0368.1.

Lloyd, I. D., and G. A. Vecchi, 2011: Observational evidence for oceanic controls on hurricane intensity. *J. Climate*, **24**, 1138–1153, https://doi.org/10.1175/2010JCLI3763.1.

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

Manganello, J. V., and Coauthors, 2012: Tropical cyclone climatology in a 10-km global atmospheric GCM: Toward weather-resolving climate modeling. *J. Climate*, **25**, 3867–3893, https://doi.org/10.1175/JCLI-D-11-00346.1.

McLachlan, G., and T. Krishnan, 2007: *The EM Algorithm and Extensions*. John Wiley & Sons, 400 pp.

Pielke, R. A., Jr., and C. W. Landsea, 1998: Normalized hurricane damages in the United States: 1925–95. *Wea. Forecasting*, **13**, 621–631, https://doi.org/10.1175/1520-0434(1998)013<0621:NHDITU>2.0.CO;2.

Rabiner, L. R., 1989: A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, **77**, 257–286, https://doi.org/10.1109/5.18626.

Rappaport, E. N., and Coauthors, 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting*, **24**, 395–419, https://doi.org/10.1175/2008WAF2222128.1.

Schade, L. R., and K. A. Emanuel, 1999: The ocean's effect on the intensity of tropical cyclones: Results from a simple coupled atmosphere–ocean model. *J. Atmos. Sci.*, **56**, 642–651, https://doi.org/10.1175/1520-0469(1999)056<0642:TOSEOT>2.0.CO;2.

Tang, B., and K. Emanuel, 2012: A ventilation index for tropical cyclones. *Bull. Amer. Meteor. Soc.*, **93**, 1901–1912, https://doi.org/10.1175/BAMS-D-11-00165.1.

Toepfer, F., R. Gall, F. Marks, and E. Rappaport, 2010: Hurricane Forecast Improvement Program: Five-year strategic plan. NOAA, 60 pp., http://www.hfip.org/documents/hfip_strategic_plan_yrs1-5_2010.pdf.

Vincent, E. M., M. Lengaigne, J. Vialard, G. Madec, N. C. Jourdain, and S. Masson, 2012: Assessing the oceanic control on the amplitude of sea surface cooling induced by tropical cyclones. *J. Geophys. Res.*, **117**, C05023, https://doi.org/10.1029/2011JC007705.

Visser, I., and M. Speekenbrink, 2010: depmixS4: An R package for hidden Markov models. *J. Stat. Software*, **36**, 1–21, https://doi.org/10.18637/jss.v036.i07.

Willoughby, H.E., J.A. Clos, and M.G. Shoreibah, 1982: Concentric eye walls, secondary wind maxima, and the evolution of the hurricane vortex. *J. Atmos. Sci.*, **39**, 395–411, https://doi.org/10.1175/1520-0469(1982)039<0395:CEWSWM>2.0.CO;2.

Zucchini, W., I. L. MacDonald, and R. Langrock, 2016: *Hidden Markov Models for Time Series: An Introduction Using R*. CRC Press, 360 pp.