JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2021MS002796

Characterizing the Impacts of Turbulence Closures on Real Hurricane Forecasts: A Comprehensive Joint Assessment of Grid Resolution, Horizontal Turbulence Models, and Horizontal Mixing Length

Oussama Romdhani[1], Jun A. Zhang[2,3], and Mostafa Momen[1]

[1]Department of Civil and Environmental Engineering, University of Houston, Houston, TX, USA, [2]NOAA/AOML/Hurricane Research Division, Miami, FL, USA, [3]Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, FL, USA

Key Points:

- Multiple weather simulations are conducted to study the combined effects of grid resolution and turbulence closures on five past hurricanes
- Although refining the grid sizes enhances hurricane wind intensity forecasts, all considered turbulence models underpredict observed data
- Reducing the default horizontal mixing length in the considered turbulence models remarkably improves hurricane's intensity forecasts

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

M. Momen,
mmomen@uh.edu

Citation:

Romdhani, O., Zhang, J. A., & Momen, M. (2022). Characterizing the impacts of turbulence closures on real hurricane forecasts: A comprehensive joint assessment of grid resolution, horizontal turbulence models, and horizontal mixing length. *Journal of Advances in Modeling Earth Systems*, *14*, e2021MS002796. https://doi.org/10.1029/2021MS002796

Received 30 AUG 2021
Accepted 29 JUL 2022

Author Contributions:

Conceptualization: Mostafa Momen
Data curation: Oussama Romdhani
Formal analysis: Oussama Romdhani, Jun A. Zhang, Mostafa Momen
Funding acquisition: Mostafa Momen
Investigation: Oussama Romdhani, Jun A. Zhang, Mostafa Momen

Abstract Hurricanes are highly complex geophysical flows that have caused billions of dollars in damage in recent years. Despite the significance of these extreme weather events, the turbulence mechanisms that derive the dynamics of hurricane flow systems are poorly understood and ineffectively parameterized in numerical weather prediction (NWP) models. The objective of this study is to bridge these knowledge gaps by assessing the accuracy and deficiencies of existing horizontal turbulence models in NWPs for hurricane forecasts. In particular, the Weather and Research Forecasting (WRF) Model is employed to conduct 135 simulations of five real hurricanes by varying the grid resolution, turbulence models, and horizontal mixing length values. Decreasing the default horizontal mixing length values both in low and high resolution WRF simulations significantly improves the wind intensity forecasts. This result indicates that the existing horizontal diffusion parameterizations are overly dissipative for hurricane flows, and thus, generate a weaker vortex compared to observations. These deficiencies are shown to stem from the horizontal mixing-length parameterization in WRF that is prescribed as a function of grid size without considering the physics of the flows (e.g., rotation). The paper provides notable insights into the role of turbulent fluxes in simulated hurricane evolutions that can be useful to advance the turbulence parameterizations of NWP models for hurricane forecasts.

Plain Language Summary With climate change and global warming, more intense hurricanes are projected to occur, potentially causing significant hazards to humans and infrastructure. It is thus imperative to improve hurricane forecasts in order to mitigate their economic ramifications. Despite the significance of turbulence in the evolution of hurricanes, our knowledge on the impact of turbulence models on real hurricane simulations is limited. This study bridges this gap by conducting multiple real hurricane simulations using different turbulence models. The results show that increasing the grid resolution generally improves the wind intensity forecasts in hurricanes. However, all considered turbulence closures underpredict the observed maximum surface wind speed in the investigated cases. This underperformance is partly associated with the improper horizontal mixing parameterizations in weather models that do not consider hurricane dynamics. We conducted new hurricane simulations by modifying the default turbulence models. The results indicated a remarkable improvement in hurricane intensity forecasts. The study shows that such physical parameterizations in weather models need to consider the unique structure of hurricane flows in order to enhance their forecasts.

1. Introduction

Since the early 1970s, hurricanes have made up eight of the ten costliest natural disasters that struck the United States (Pita et al., 2015), causing more than $500 billion in total damages (Guha-Sapir et al., 2011). Climate change and global warming can potentially increase the frequency (Emanuel, 2017; IPCC, 2007), and intensity of future tropical cyclones [TCs; (Bender et al., 2010; Emanuel, 2005; Mei & Xie, 2016)], leading to serious hazards for the environment and society (Cheikh & Momen, 2020; Marsooli et al., 2019; Weinkle et al., 2018). Over the last two decades, the United States has seen a significant number of major hurricanes landfalling on its shores.

To mitigate the adverse impacts of this natural threat, numerous scientific and technological innovations have been developed by the research community to advance our understanding and modeling of hurricanes. Accurate predictions of these extreme weather events enable the communities to prepare and issue evacuation orders far

Methodology: Oussama Romdhani, Mostafa Momen
Project Administration: Mostafa Momen
Resources: Mostafa Momen
Software: Oussama Romdhani
Supervision: Mostafa Momen
Validation: Oussama Romdhani
Visualization: Oussama Romdhani
Writing – original draft: Oussama Romdhani, Mostafa Momen
Writing – review & editing: Oussama Romdhani, Jun A. Zhang, Mostafa Momen

in advance of the hurricane's landfall. Numerical weather prediction (NWP) models are the prominent advanced available tools for forecasting the hurricane track and intensity. These models, when employed properly with efficient evacuation procedures, resulted in a reduction in hurricane mortality rates, especially in developed countries (Arguez & Elsner, 2001; Peduzzi et al., 2012).

Weather Research and Forecasting (WRF) Model is the state-of-the-art NWP model that is used in the US for operational weather forecasts (Al-Yahyai et al., 2010). It has been extensively employed for both research and practical applications (Nasrollahi et al., 2012), and has shown great potential in capturing the physics of hurricanes (Abarca & Corbosiero, 2011). The Advanced Research WRF (ARW) is one of the two dynamical solvers of WRF that was primarily developed by the National Center for Atmospheric Research (NCAR). The ARW code has been employed to study hurricane dynamics, for example, by forecasting five landfalling Atlantic hurricanes during 2005 (Davis et al., 2008), finding secondary eyewall formations (Abarca & Corbosiero, 2011), and assimilating observational data for improving TC tracks (Cavallo et al., 2013).

Despite the significant improvements of the TC forecasts in recent years, they still face many challenges for accurately predicting hurricane track and intensity, particularly for rapidly intensifying hurricanes (Emanuel, 2017). One of the main reasons for these NWP shortcomings is the inaccurate physical parameterizations of hurricane flows. A mature hurricane is ideally an asymmetric high-Reynolds number vortex in hydrostatic and rotational balance that consists of a wide range of fluid-dynamical processes such as boundary layer (BL) turbulence, rotating convection, and air-sea interactions (Emanuel, 1991; Montgomery & Smith, 2017). The atmospheric boundary layer (ABL), which plays a major role in hurricane dynamics, is defined as the lowest portion of the atmosphere extending up to about 2 km above the ground in which most of the momentum and heat fluxes are carried by turbulent motions (Garratt, 1994; Lemone et al., 2019; Stull, 1988). Turbulence in ABLs is affected by many complex factors such as unsteadiness (Momen & Bou-Zeid, 2016, 2017), baroclinicity (Floors et al., 2015; Momen, 2022; Momen et al., 2018), surface heterogeneity (Anderson & Chamecki, 2014; Bou-Zeid et al., 2020; Calaf et al., 2010), and stratification (Mahrt, 1999; Momen & Bou-Zeid, 2017; Ramamurthy et al., 2007; van de Wiel et al., 2012).

In hurricane flow systems, turbulence has a pivotal role in the evolution and intensification of hurricanes (Zhang, 2010). Recent studies showed a strong correlation between the intensification of a hurricane and the turbulent transport in the BL and the eyewall region (Emanuel, 1995; Persing & Montgomery, 2003; Rotunno & Bryan, 2012). Turbulent transport processes in hurricane eyewalls are essential for initiating the rapid intensification of TCs (Zhu et al., 2021). The vertical turbulent mixing is shown to be significant in maintaining the intensity and structure of hurricanes (Braun & Tao, 2000; Foster, 2009; Kepert, 2012; Zhang et al., 2015, 2017).

The NWP models typically use meshes with horizontal spacings of several kilometers (2–30 km) due to the computational limitations for simulating large domains. These grid resolutions are so coarse that cannot resolve the three-dimensional (3D) turbulence and energy-containing eddies (∼10–100 m) in the ABL (Moeng et al., 2007). Therefore, turbulent fluxes are parameterized to calculate the sub-grid scale stresses in such models. Despite the importance of vertical turbulent fluxes in hurricane simulations (Braun & Tao, 2000; Foster, 2009; Kepert, 2012; Zhang et al., 2015, 2017), the present study focuses on investigating different horizontal diffusion models since they have received relatively less attention. Horizontal diffusion significantly impacts the simulated hurricane intensity (Rotunno & Bryan, 2012), maximum wind speed, TC vortex size, inflow layer depth (Zhang & Marks, 2015), and storm structure (Zhu et al., 2014). Horizontal diffusion has been shown to be the most controlling factor for simulating maximum TC intensity (Bryan & Rotunno, 2009; Rotunno & Bryan, 2012). Furthermore, a shift of power spectrum toward higher wave numbers in hurricane systems compared to the typical ABL flows is observed both in measurements (Zhang, 2010) and high-resolution large-eddy simulations (Momen et al., 2021; Worsnop et al., 2017), indicating the stark difference of horizontal turbulence structures in hurricanes compared to conventional ABLs.

In NWPs, the horizontal turbulent fluxes are typically parameterized using the *K*-theory (aka gradient transport theory; Stull, 1988) and the deformation of the large-scale flow (Mirocha et al., 2010). *K*-theory is a first-order closure approximation for the turbulent fluxes that are parameterized as a function of the eddy diffusivity (K) times the local gradient of the mean quantity (Stull, 1988). The eddy diffusivity in WRF is modeled using a mixing-length formulation. Note that the mixing length in Prandtl's mixing-length theory, which forms the basis of these models, depends on the nature of the flow and is space dependent. Hence, the mixing length in hurricanes

can be significantly different from typical ABL flows (Momen et al., 2021), and as such it has to be specified for such complex flows that have not been established before (Pope, 2000). Thus, the paper aims to comprehensively assess and evaluate the horizontal mixing length scales and different horizontal turbulence closures in hurricane flow systems.

Despite the evidence that underscores the significance of turbulence characteristics in hurricane simulations, our knowledge on the impacts of horizontal turbulence closures on the WRF model's real hurricane forecasting performance is limited. Only a few studies have thus far examined the effects of horizontal eddy diffusivity on hurricane simulations (e.g., Zhang et al., 2018; Zhang & Marks, 2015). Most of these studies primarily assessed a single diffusion parameterization scheme by using the non-hydrostatic mesoscale core of WRF. Nonetheless, the interacting effects of grid resolution, different horizonal turbulence models, and diffusion length scales on the underlying dynamics (e.g., Reynolds stresses) and accuracy of real hurricane simulations are not yet comprehensively established. In this paper, we aim to bridge these knowledge gaps by systematically varying the horizontal mixing length parameters, grid resolution, and horizontal turbulence models in the WRF-ARW system. In particular, our research questions are:

1. What are the effects of different horizontal turbulence closure models under different grid resolutions on the accuracy of real hurricane forecasts?
2. What are the impacts of changing the horizontal mixing lengths on the simulated hurricane intensity, wind profile, sea-level pressure, and Reynolds stresses?
3. How does varying the horizontal mixing length in the existing horizontal turbulence closures impact the accuracy of real hurricane forecasts?

To address these questions, we first compare the performance of different horizontal turbulence models in WRF-ARW at different grid resolutions for five real hurricane cases. Then, we delineate the impacts of the horizontal mixing lengths on the hurricane flow characteristics. Finally, the performance of the horizontal turbulence models with the newly modified mixing lengths in hurricane forecasts are compared with the default values, and potential pathways toward improving these models are proposed.

The current paper has the following structure. Section 2 describes the selected hurricanes, the numerical framework, and the details of the turbulence models. Then, the section provides an overview of the conducted WRF-ARW simulations to probe the full parameter space of the problem. Section 3 presents the results of our simulations. The first part of the section shows the results of the default turbulence models at different grid resolutions and evaluates their track and wind speed forecasts compared to the measured data. The next part of the section analyzes the impacts of varying the horizontal mixing length on the simulated hurricane intensity, wind profile, sea-level pressure, and Reynolds stresses. The last part of the section comprehensively evaluates the performance of the WRF hurricane forecasts with the newly modified turbulence mixing lengths compared to the default values. Finally, Section 4 provides a summary of the main findings of the paper.

2. Methods

2.1. Hurricane Cases and Numerical Domains

To investigate the impacts of horizontal turbulence and mixing length parametrizations on the NWP model's hurricane forecasts, five TCs of categories 4 or 5 were selected. These hurricanes were chosen based on their maximum intensity, economic footprint, and spatial distribution over the Atlantic Ocean. The selected hurricanes occurred between 2005 and 2018 leading to total estimated damage of more than $295 billion [(Blake et al., 2021); see Table 1]. Katrina, a category 5 storm has been the costliest and one of the most intense Atlantic hurricanes to make landfall in the US thus far (Houze et al., 2006). Gustav, a tropical depression formed near the Caribbean, quickly developed into a category 4 hurricane that caused considerable damages and casualties along its track across the Dominican Republic, Haiti, Jamaica, Cuba, and southeast Louisiana (Beven & Kimberlain, 2009). Florence, originated from a convectively tropical wave, moved off the west coast and made landfall near Wrightsville Beach, North Carolina on 14 September 2018. The two other selected hurricanes are Irma that made 6 landfalls in 2017 before striking Florida as a category 5 hurricane (Ginzburg et al., 2018) and Maria, which was the third costliest hurricane in US history as of 2021 (Pasch et al., 2017). More details about these hurricanes are provided in Table 1.

Table 1
List of Investigated Hurricanes and Their Simulation Periods in Weather and Research Forecasting Model

Hurricanes	Year	Category	Formation to dissipation dates	Simulation running hours	Approximate domain size	Estimated damages	Observed max speed
Katrina	2005	5	August 23–August 31	30 h, August 28, 12 am–August 29, 6 am	2100 km × 2940 km	$125 billion	77 m/s
Gustav	2008	4	August 25–September 7	30 h, August 31, 6 am–Sept. 1, 12 pm	2700 km × 2400 km	$6 billion	70 m/s
Irma	2017	5	August 30–September 14	48 h, September 2, 12 am–September 4, 12 am	3000 km × 2000 km	$50 billion	80 m/s
Maria	2017	5	September 16–October 2	48 h, September 22, 12 am–September 24, 12 am	2100 km × 2700 km	$90 billion	77 m/s
Florence	2018	4	August 31–September 17	48 h, September 11, 12 am–September 13, 12 am	2100 km × 2700 km	$24 billion	67 m/s

The main focus of the present paper is to assess the ARW's performance over the Ocean as landfalling processes and surface heterogeneities add extra complicating factors to the turbulence characteristic of the problem, for example, by forming internal boundary layers. Hence, the simulated period and the domain size were governed by the hurricanes' track and their associated intensities over those trajectories. The hurricanes were simulated for up to 2 days in which they achieved their peak speed over the Atlantic Ocean. Both Katrina and Gustav were simulated for 30 hr during which they regained strength across the Gulf of Mexico Basin. For Irma, Maria, and Florence the simulations were conducted for 2 days respectively before hitting the Dominican Republic, after the departure from Cuba, and before making landfall in North Carolina. Figure 1 depicts the hurricanes' simulated period, the best observed full track, and the WRF domain size for each case.

2.2. Numerical Methodology: The Eddy Viscosity Models

The ARW version 4.1 is adopted in this study to simulate the considered hurricane cases. The ARW has been extensively used and validated for hurricane simulations (Alvey et al., 2020; Cavallo et al., 2013; C. Davis

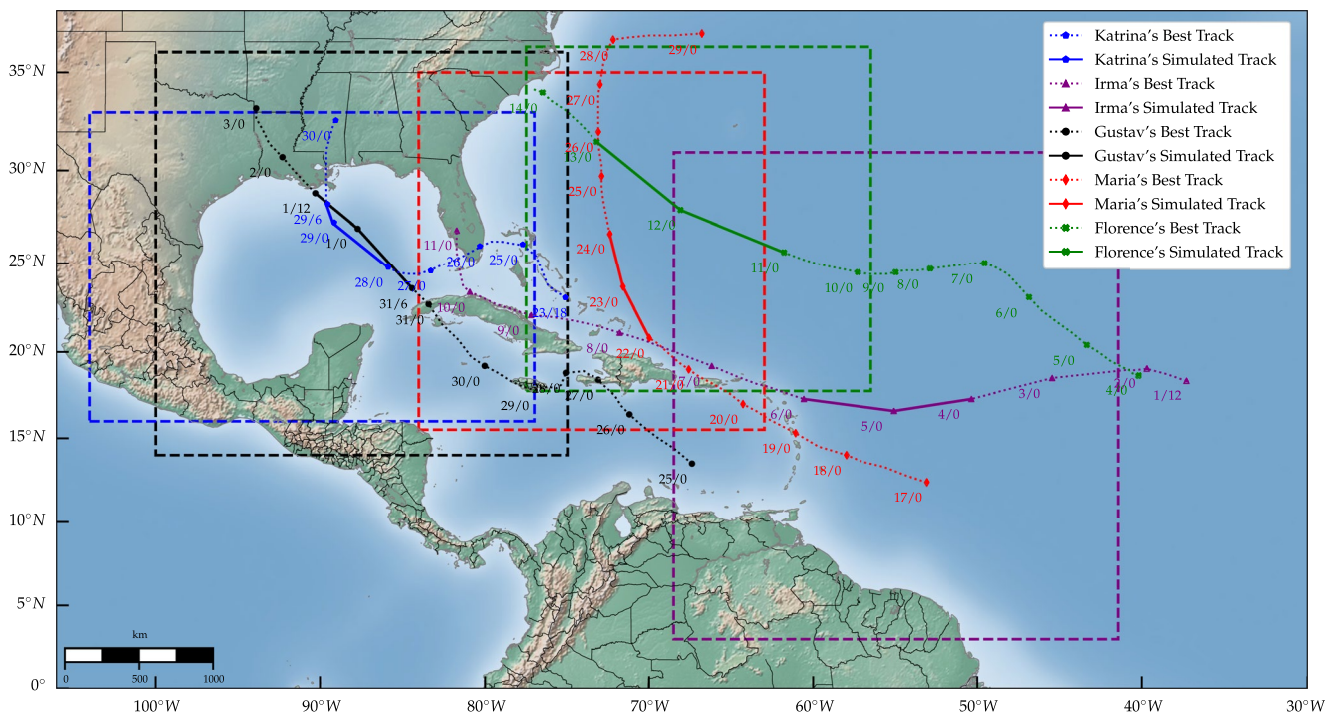

Figure 1. Best observed track of the five chosen hurricanes. The solid lines depict the hurricanes trajectories during the simulation period, and the dashed boxes display the size of the simulated domain for each hurricane. The markers show the hurricane's eye location.

et al., 2008; Chen et al., 2006; Fierro et al., 2009; Islam et al., 2015; Pattanayak et al., 2006; Xue et al., 2013). The ARW code integrates a set of fully compressible, non-hydrostatic Euler equations at each time step. The dynamics solver of the ARW is built upon an Arakawa C-grid staggering horizontally and terrain-following hydrostatic pressure coordinates vertically (Skamarock et al., 2019). Further details about the governing equations of the ARW and the employed planetary boundary layer (PBL) schemes for modeling the vertical turbulent fluxes can be found in Supporting Information S1. Please refer to Supporting Information S1 for more details about the selected microphysics and radiation schemes.

The ARW system provides four options to determine the horizontal eddy viscosity (K_h): (a) the constant option in which the user externally defines constant values for the eddy viscosity, (b) the two-dimensional (2D) first-order Smagorinsky (Smag2D) option, (c) the three-dimensional (3D) Smagorinsky turbulence closure that calculates both K_h and the vertical diffusion, and (d) a prognostic approach predicting turbulent kinetic energy (TKE) that determines both K_h and the vertical eddy viscosity. If the PBL scheme is turned on in the TKE mode, the vertical diffusion is calculated by the PBL scheme and only the horizontal diffusion is computed by the prognostic TKE equations.

In the current study, we employ Smag2D and TKE options to assess the impacts of the horizontal diffusion parameterizations on simulating real hurricanes. The utilized models are coupled with a diffusion parametrization in the physical space. These two turbulence closures are also evaluated against a no horizontal turbulence model option—that is, explicit spatial numerical filters simulations—to which we refer as NoHorizTurb hereafter. This assessment allows us to indicate the degree to which the existing horizontal turbulence models in WRF (Smag2D and TKE) improve or worsen the real hurricane simulations compared to a scenario when there is no horizontal turbulence model (NoHorizTurb).

2.2.1. Turbulent Kinetic Energy Model

The TKE model is a 3D 1.5-order turbulence closure that is built upon a prognostic equation for the evolution of the TKE *"e"* (Stull, 1988). The prognostic TKE closure evaluates the mixing terms in the physical space. This physical space formulation computes the geometric height coordinates at each Runge-Kutta time step using the prognostic geopotential ($\phi = gz$) in the ARW solver

$$z_x = g^{-1}\frac{\partial \phi}{\partial x}, \quad z_y = g^{-1}\frac{\partial \phi}{\partial y}. \tag{1}$$

These metric terms are defined on the vertical levels, and horizontally, they coincide with the horizontal velocity points (u, v). The forcing terms in the continuous momentum equations for evaluating diffusion in the physical space can be written as follows,

$$\frac{\partial U_i}{\partial t} = \cdots - m_i^*\left[\frac{\partial \tau_{1i}}{\partial x} + \frac{\partial \tau_{2i}}{\partial y} - z_x\frac{\partial \tau_{1i}}{\partial z} - z_y\frac{\partial \tau_{2i}}{\partial z}\right] - \frac{\partial \tau_{3i}}{\partial z}, \tag{2}$$

where $m_i^* = (m_x, m_y, m_y)$ are map scaling factors (see Supporting Information S1) and τ_{ij} is the stress tensor defined as

$$\begin{aligned} \tau_{ij} &= -\mu_d K_h D_{ij}, \quad i, j \neq 3, \\ \tau_{i3} &= -\mu_d K_v D_{i3}, \quad i = 1, 2, 3 \end{aligned} \tag{3}$$

where K_h is the horizontal eddy viscosity and K_v represents the vertical eddy viscosity. The stress tensor is calculated using the deformation tensor D_{ij} whose horizontal components are given below,

$$D_{11} = 2m_x m_y\left[\frac{\partial}{\partial x}\left(\frac{u}{m_y}\right) - z_x\frac{\partial}{\partial z}\left(\frac{u}{m_y}\right)\right], \tag{4.1}$$

$$D_{22} = 2m_x m_y\left[\frac{\partial}{\partial y}\left(\frac{v}{m_x}\right) - z_y\frac{\partial}{\partial z}\left(\frac{v}{m_x}\right)\right], \tag{4.2}$$

$$D_{12} = m_x m_y\left[\frac{\partial}{\partial y}\left(\frac{u}{m_y}\right) - z_y\frac{\partial}{\partial z}\left(\frac{u}{m_y}\right) + \frac{\partial}{\partial x}\left(\frac{v}{m_x}\right) - z_x\frac{\partial}{\partial z}\left(\frac{v}{m_x}\right)\right]. \tag{4.3}$$

Considering Equation 1 and the grid size implementations, an anisotropic mixing option (appropriate for Δx and $\Delta y \gg \Delta z$) is used here. In such a condition, K_h is calculated in the code as

$$K_h = C_k L_h \sqrt{e}, \tag{5}$$

where C_k is a constant typically between 0.15 and 0.25, and L_h is the horizontal mixing length defined as

$$L_h = c L_h \sqrt{\Delta x \Delta y}, \tag{6}$$

where cL_h is a new constant that we introduced here to control the mixing length (the default value is 1). Since the vertical diffusion in the current study is handled by the PBL scheme, the K_v formulations are not detailed here.

2.2.2. Two-Dimensional Smagorinsky (Smag2D) Model

The Smag2D model is a first-order closure that determines K_h in terms of the horizontal deformation tensor. The continuous momentum equations, stress, and deformation tensors can be written in the ARW physical space according to Equations 2 and 3 and K_h for Smag2D is given by

$$K_h = C_s^2 L_h^2 \left[0.25(D_{11} - D_{22})^2 + \overline{D_{12}^2}^{xy}\right]^{1/2}, \tag{7}$$

where C_s is the Smagorinsky constant typically equal to 0.25, D_{ij} represents the components of the planar deformation tensor defined previously, and the overline denotes the spatial average around the grid cell. L_h in Equation 7 is the horizontal mixing length scale given by Equation 6, in which we defined the new coefficient cL_h to control the mixing. The vertical diffusion is handled independently by the PBL scheme. Moreover, for temperature mixing, the eddy diffusivity is divided by the Prandtl turbulent number P_r that is typically equal to 1/3 (Deardorff, 1972).

2.3. Suite of WRF Simulations

Two sets of ARW simulations are conducted in this study: one that assesses the horizontal turbulence models' performance over different grid resolutions and another that characterizes the effects of the mixing length variations on the simulated hurricane forecasts. For the first set, three turbulence model configurations were utilized: the NoHorizTurb approach which excludes any horizontal turbulence modeling, Smag2D, and the 1.5-order TKE closure. The three approaches are implemented for five real hurricanes (Katrina, Gustav, Irma, Maria, and Florence) and on five different grid resolutions (2, 4, 8, 16, and 32 km), resulting in $3 \times 5 \times 5 = 75$ ARW simulations.

In the second set, the impacts of the horizontal turbulence parameterization on real hurricane forecasts were investigated by modulating the L_h. This goal was achieved by varying the introduced new constant cL_h (see Equation 6) to 0.25, 0.5, and 1.5. These changes were implemented for five hurricane simulations under two horizontal turbulence models (Smag2D and TKE) over 4 and 32 km grid resolutions, resulting in $3 \times 5 \times 2 \times 2 = 60$ new ARW cases.

Table 2 summarizes all the 135 configurations in terms of grid sizes, turbulence models, and mixing length constants. Each configuration is labeled according to the first two characters of the hurricane name ("Ka" for Katrina, "Ir" for Irma, "Gu" for Gustav, "Ma" for Maria, "Fl" for Florence), the grid size, the horizontal turbulence model (NoHorizTurb, Smag2D, or TKE), and the mixing length L_h. For example, Ka_32km_Smag2D_cL_h0.5 represents a simulation case of Hurricane Katrina in which we used a 32 km grid, Smag2D as the horizontal turbulence model, and changed the default mixing length constant from 1.0 to 0.5. More information about the initial and boundary conditions of the conducted simulations can be found in Supporting Information S1.

2.4. Evaluation Metrics

Various metrics were employed to assess the ARW solver's accuracy by comparing the simulations to a wide range of observations including GPS dropsonde data, National Hurricane Center's best track and wind speed data (Cangialosi et al., 2018; Knabb et al., 2006; Pasch et al., 2017; Pope et al., 2010; Stewart and & Berg, 2019), and Stepped Frequency Microwave Radiometer (SFMR) data provided by NOAA. The first major metric used here evaluates the simulated hurricane intensity. The recorded best track speed from measurements is compared with

Table 2
The Suite of Weather and Research Forecasting Simulations Conducted for the Joint Assessment of Horizontal Turbulence Models, Grid Resolutions, and the Horizontal Mixing Length Scales

Model \ Grid Size			2 km	4 km	8 km	16 km	32 km
NoHorizTurb			X_2km_NoHorizTurb	X_4km_NoHorizTurb	X_8km_NoHorizTurb	X_16km_NoHorizTurb	X_32km_NoHorizTurb
Smag2D	cL_h	0.25	–	X_4km_Smag2D_cL_h0.25	–	–	X_32km_Smag2D_cL_h0.25
		0.5	–	X_4km_Smag2D_cL_h0.5	–	–	X_32km_Smag2D_cL_h0.5
		1*	X_2km_Smag2D	X_4km_Smag2D	X_8km_Smag2D	X_16km_Smag2D	X_32km_Smag2D
		1.5	–	X_4km_Smag2D_cL_h1.5	–	–	X_32km_Smag2D_cL_h1.5
TKE	cL_h	0.25	–	X_4km_TKE_cL_h0.25	–	–	X_32km_TKE_cL_h0.25
		0.5	–	X_4km_TKE_cL_h0.5	–	–	X_32km_TKE_cL_h0.5
		1*	X_2km_TKE	X_4km_TKE	X_8km_TKE	X_16km_TKE	X_32km_TKE
		1.5	–	X_4km_TKE_cL_h1.5	–	–	X_32km_TKE_cL_h1.5

Note. X symbolizes one of the 5 hurricanes, it can either be "Ka" for Katrina, "Ir" for Irma, "Gu" for Gustav, "Ma" for Maria, or "Fl" for Florence. The next variable denotes the grid size which is one of the following: 2, 4, 8, 16 or 32 km. The third variable represents one of the three turbulence models investigated in this study. Finally, cL_h, if written, represents a change of the constant multiplied by the default horizontal mixing length. The default value of cL_h is 1, which gives $L_h = \sqrt{\Delta x \Delta y}$. Please see the text for an example of the case naming convention. In total, $5 \times 27 = 135$ simulations are conducted for the joint investigation of the grid resolution, horizontal turbulence closure, and mixing length scale on the accuracy of real hurricane forecasts in WRF.

the simulated horizontal surface wind speed by defining the following mean absolute error (MAE) criterion for intensity:

$$\mathrm{MAE}_{Intensity} = \frac{1}{N}\sum_{i=1}^{N}\left|M_{Forecast}(i\Delta t) - M_{Best\ Track}(i\Delta t)\right|, \tag{8}$$

where N denotes the total number of samples, Δt is the time interval of the reported best track data (6 hr here), $M_{Forecast}(i\Delta t)$ denotes the maximum simulated horizontal wind speed at 10 m of altitude and at time $t = i\Delta t$, and $M_{Best\,Track}(i\Delta t)$ is the reported best track speed data at $t = i\Delta t$.

The intensity of each hurricane could be different and can also vary with time. To account for all these differences and to provide a standard measure to evaluate the intensity forecasts of WRF simulations among different cases in a non-dimensional way, the mean absolute percentage error (MAPE) is used. MAPE normalizes the error by the observed wind speed data at the defined time step as

$$\mathrm{MAPE}_{Intensity} = \frac{100}{N}\sum_{i=1}^{N}\frac{|M_{Forecast}(i\Delta t) - M_{Best\ Track}(i\Delta t)|}{M_{Best\ Track}(i\Delta t)}. \tag{9}$$

Hence, a $\mathrm{MAPE}_{\mathrm{Intensity}}$ of 20% in one simulation indicates that WRF predicts the intensity of that case with an average error of 20% compared to the observed values.

To assess the track error, the distance between the best track and the forecasted hurricane eye location is evaluated by defining the following MAE for track

$$\mathrm{MAE}_{Track} = \frac{1}{N}\sum_{i=1}^{N}\left|X^{Eye}_{Forecast}(i\Delta t) - X_{Best\ Track}(i\Delta t)\right|, \tag{10}$$

where $\boldsymbol{X}(i\Delta t)$ is the position vector at $t = i\Delta t$, and $\|\cdot\|$ is the Euclidean norm. For the track, we present the errors in dimensional sense (km), since there is no common characteristic length scale for track error normalization, unlike the intensity error.

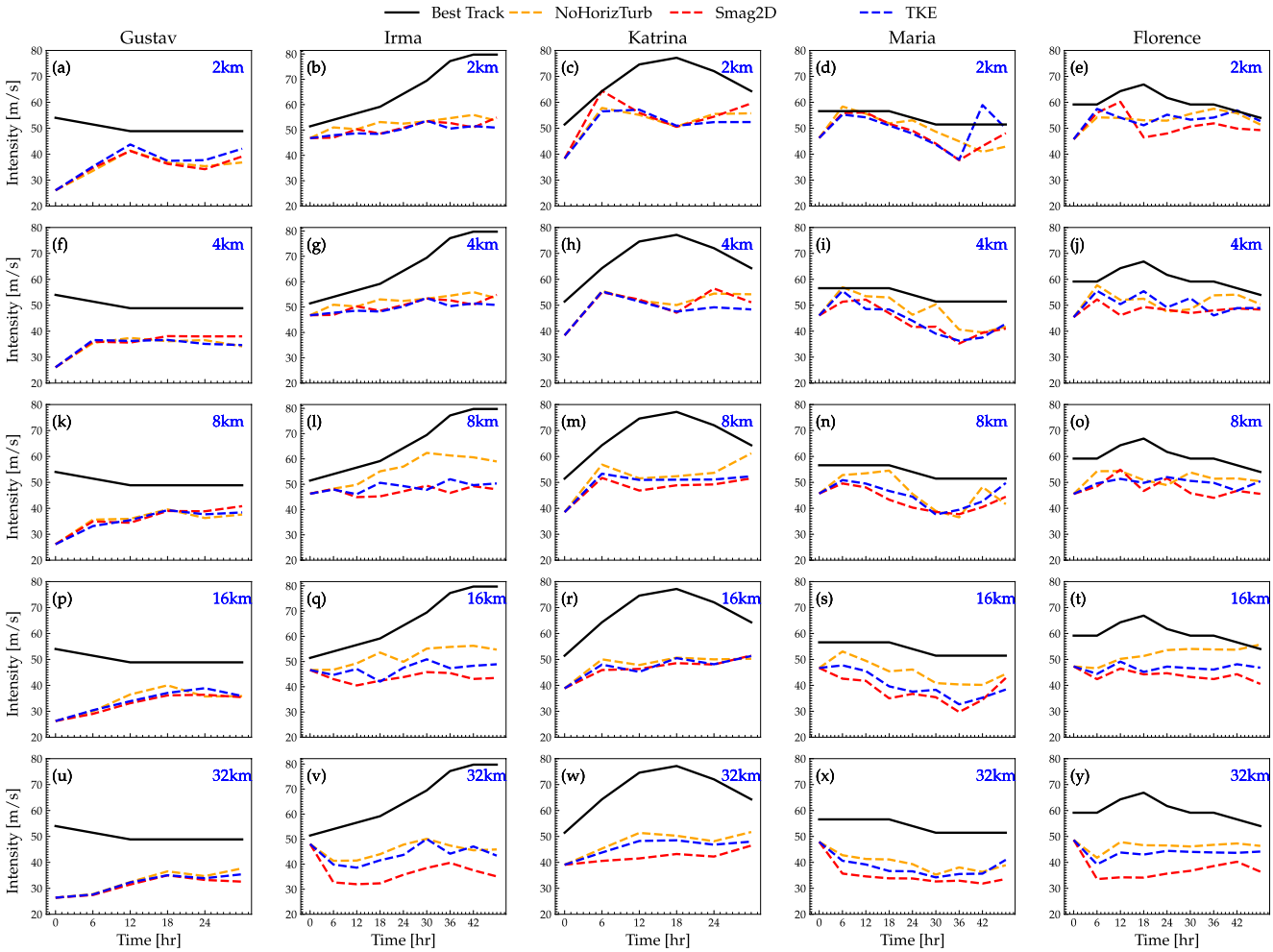

Figure 2. Comparison between the observed best track speed (solid black line) and the forecasted near-surface wind speed (10 m of altitude) for three different turbulence models (NoHorizTurb, Smag2D, turbulent kinetic energy (TKE)) with their default length scales among all hurricanes and grid sizes.

3. Results and Discussion

The results are discussed in three sections. First, the accuracies of the existing horizontal turbulence models are evaluated as a function of grid resolution in Section 3.1. Then, the new ARW simulations by changing the default mixing length values are presented. In Section 3.2, the impacts of the horizontal mixing-length on simulated hurricane dynamics are discussed. In Section 3.3, the forecast improvements of simulated hurricane intensity, track, and wind profiles resulting from changing the default L_h are assessed.

3.1. Assessment of the Default Turbulence Models Under Different Grid Resolutions

In this section, the results of the first set of the ARW simulations (75 cases) with the default turbulence models ($cL_h = 1$) are presented. First, the interacting effects of grid resolution and horizontal turbulence models on hurricane intensity forecasts are examined. Figure 2 displays the timeseries of the best track wind speed in comparison with the forecasted near-surface maximum wind intensity for each hurricane.

The simulations indicate that in general increasing the grid resolution (i.e., reducing the grid size) improves the intensity forecasts in all considered models. For instance, Hurricane Katrina after 24 hr of simulation reached an intensity of ~43 m/s for Smag2D at 32 km grid resolution (red line in Figure 2w). When the grid size was reduced to 2 km, Smag2D's intensity increased to ~51 m/s (red line in Figure 2c), bringing it closer to the best track observed speed of ~77 m/s. Similar trends were noticed for other hurricanes and turbulence models. Note

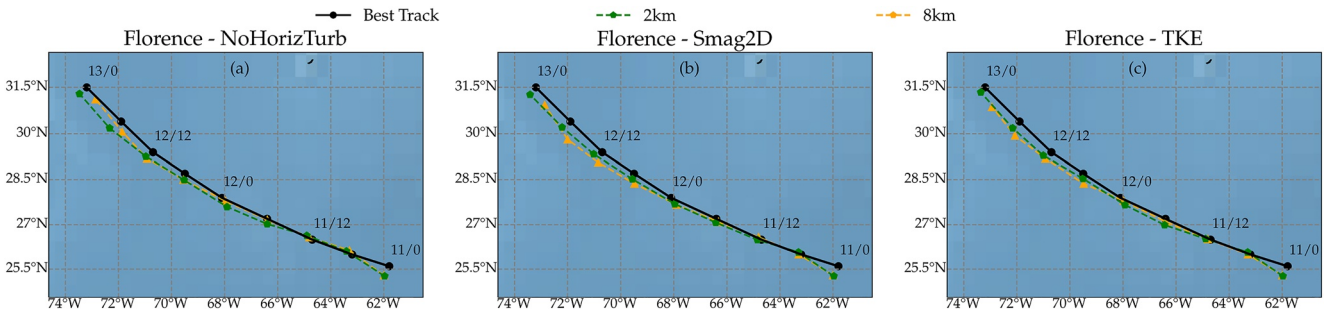

Figure 3. The performance of the three considered horizontal turbulence models in forecasting Hurricane Florence's track for 2 and 8 km grid sizes.

that such improvements were less apparent in Hurricane Gustav's simulations. We attribute this to the poor initial data of Hurricane Gustav's simulations that were obtained from NCEP. The initial gap in Gustav underestimates the intensity by more than 20 m/s, while such gap is ∼10 m/s for other hurricanes.

When comparing the performance of the three considered turbulence models, NoHorizTurb generally appears to be the closest one to the best track speed and Smag2D is often the least accurate one in this sense. This distinction among the three models becomes more apparent as the grid resolution decreases. Figure 2 demonstrates that NoHorizTurb mostly outperformed Smag2D and TKE models for simulating the intensity of the five considered hurricanes especially at low resolutions. This implies that the existing horizontal turbulence models in ARW worsen the intensity of real hurricane simulations compared to a scenario when no horizontal turbulence scheme is turned on. This poor performance of the existing horizontal turbulence models motivates their modifications specifically for hurricane flows. The paper will elucidate the reasons why these models fail to predict proper hurricane intensity (Section 3.2) and will provide an adjustment as a first step to address this deficiency (Section 3.3).

The next important criterion is the accuracy of the forecasted hurricane track. Figure 3 displays the best observed track of Florence (black line) versus the forecasted tracks at two different resolutions for the three considered turbulence models. Please refer to the supplementary Figures S1 and S2 in Supporting Information S1 for the forecasted track in other simulations. The simulations with 2 km grid resolution were generally closer to the best track than the ones with 8 km resolutions in Hurricane Florence. As the figure shows, the considered models decrease the initial gap between the forecasted and real hurricanes tracks from ∼43 km at the starting hour to an average of ∼28 and ∼31 km, for the remaining time steps, for 2 and 8 km grid resolutions, respectively. We note that this is not always the case, and in some cases, the models diverge from the best track as time evolves (see supplementary Figure S1 in Supporting Information S1).

To better understand the overall performance of the horizontal turbulence models in terms of track and intensity forecasts, their statistics during entire simulation periods and among all considered hurricanes are analyzed. Figure 4 presents the average of the two error metrics defined in Equations 9 and 10 over all hurricane forecasts as a function of the grid resolution of the simulations. Each bar plot in this figure shows the average of the five hurricane MAPE or MAE results. As Figure 4a indicates, the average $MAPE_{Intensity}$ decreases from ∼32% at 32 km grid resolution to ∼19% at 2 km grid resolution in all considered turbulence models. In particular, the intensity prediction of the Smag2D simulations with 32 km grid resolution was improved by ∼48.7% when the grid size changed to 2 km ($100\%–100 \times MAPE_{Intensity,2km\_Smag2D}/MAPE_{Intensity,32km\_Smag2D}$). Similarly, the 2 km TKE simulations improved the intensity forecasts of the 32 km TKE cases on average by ∼39.1%, and such intensity prediction improvements were ∼35% on average for the NoHorizTurb cases. As the figure indicates, NoHorizTurb outperforms the other horizontal turbulence models in terms of intensity forecasts in the considered cases. Smag2D has generally the largest $MAPE_{Intensity}$ among the three investigated models.

The average track errors in the conducted simulations are also assessed. Although the MAE_{Track} in Figure 4b shows a slight improvement as the grid resolution increases, this enhancement in the track is not as clear as the intensity error (Figure 4a). Moreover, this trend varies among the employed turbulence models. For Smag2D, the average MAE_{Track} seems to improve with the grid resolution from ∼65.2 km at 32 km grid size to ∼48.6 km at 2 km grid size. The TKE model generally outperforms the NoHorizTurb cases in terms of track forecasts.

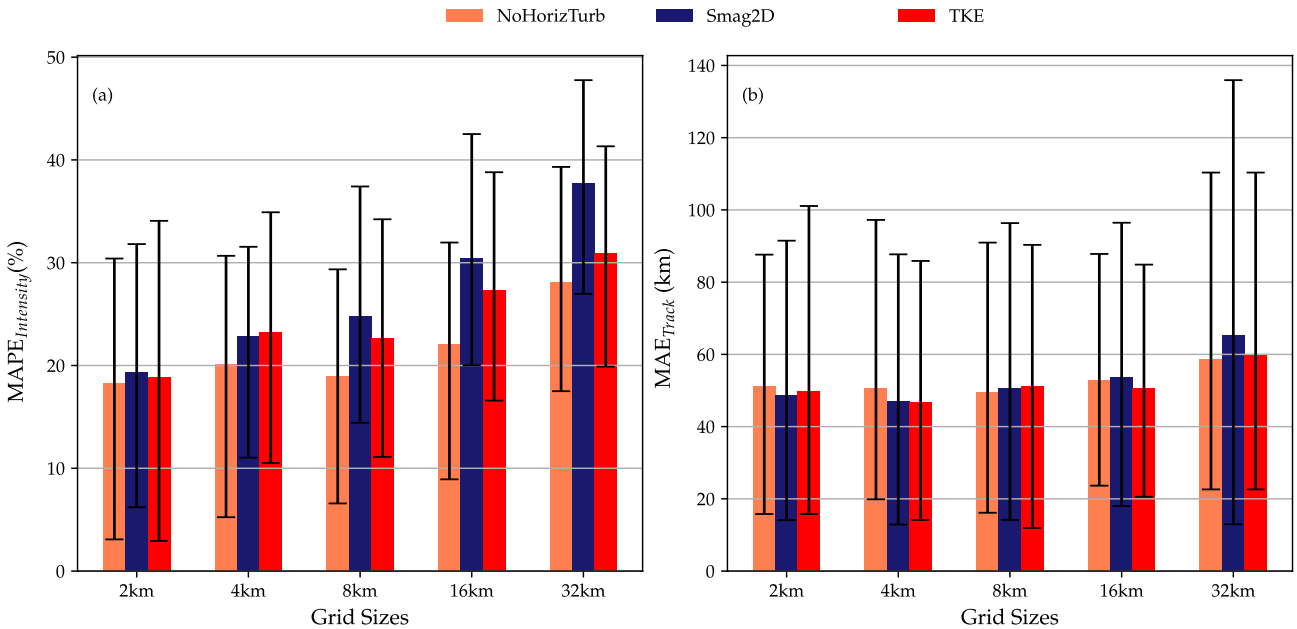

Figure 4. Average error of all considered hurricanes in terms of grid sizes and default turbulence models. The solid black lines represent the 10th and 90th percentile error bars.

Interestingly, despite the poorer performance of Smag2D in intensity, it outperforms the track forecasts of the TKE and NoHorizTurb models at 2 km grid size. In 2 km and 4 km simulations, Smag2D and TKE outperform the overall track forecasts of the NoHorizTurb cases. All the results of the simulation errors are reported in Table S1 in Supporting Information S1.

We also investigated the performance of the Smag3D turbulence model as the third option of ARW. This option is not recommended for grid resolutions greater than 2 km (Wei et al., 2019). Hence, we only conducted the simulations using this model with a 2 km resolution. Smag3D outperforms all the other considered models in terms of intensity predictions by achieving the lowest MAPE of ∼16% (see Figure S4 in Supporting Information S1). However, it predicts the hurricane tracks inferior to the other considered models. Since Smag3D is not recommended for grid resolutions greater than 2 km, we have not included it in the main paper and the details of these simulations can be found in Supporting Information S1.

To quantify the relationship between the intensity and track errors with the grid resolution, a linear regression analysis of the data is conducted. Table 3 presents the results of the linear regression applied to the normalized errors for each turbulence model versus the grid resolution (normalized by 32 km). The linear regression curves are shown in Figure S3 in Supporting Information S1. In this analysis, the MAE_{Track} data are normalized by their maximum track error for each turbulence model. The table corroborates the previous qualitative observation of the relationship between the grid size and the forecast errors. High R^2 values ($\gtrsim$0.95) and low P values ($\lesssim$0.05) in the first row of Table 3 indicate a strong relationship between $MAPE_{Intensity}$ and grid resolution in all considered turbulence models. Furthermore, Smag2D seems to be more sensitive to grid resolution changes in terms of forecasting intensity by having a larger slope (∼19) than the other models. These results are consistent with

Table 3
Linear Regression Parameters That Model the Relationship Between the Normalized Grid Size (Δx/32 km) and the Normalized Errors for Each Turbulence Model

	NoHorizTurb			Smag2D			TKE		
	Slope	R^2	P	Slope	R^2	P	Slope	R^2	P
$\overline{MAPE}_{Intensity}$ versus Grid Sizes/32 km	10.1	0.94	0.017	18.6	0.98	10^{-7}	11.5	0.97	0.001
$\overline{Normalized\ MAE}_{Track}$ versus Grid Sizes/32 km	0.08	0.18	0.36	0.17	0.22	0.4	0.1	0.26	0.28

Note. In this table, P refers to the p-value and tests the null hypothesis. See Figure S3 in Supporting Information S1 for more details

previous findings of the ARW simulations of hurricanes that showed improved intensity forecasts with increasing grid resolution (Davis et al., 2010). Furthermore, increasing the grid resolution appears to have a more significant impact on intensity forecast improvements than choosing one of the considered horizontal turbulence models.

The results in Table 3 do not show a statistically significant dependence of the track error on the grid resolution for all models by yielding low R^2 values (<0.3). Once more Smag2D appears to be more sensitive to the grid resolution for track forecasts than the other models (higher slope values as shown on the second row of Table 3). Previous studies also showed no statistically significant differences regarding track error among high and low grid resolution ARW simulations (Davis et al., 2010; Fierro et al., 2009). This was partly attributed to the fact that, unlike intensity, track forecasts depend on large-scale processes that can be resolved with coarse grid resolutions as well (Fierro et al., 2009).

3.2. The Impacts of Varying the Horizontal Mixing Length on the Dynamics of Hurricanes

In the previous section, we assessed the performance of the default turbulence closures at different resolutions. In this section, we characterize the impacts of changing the horizontal mixing length scale on the dynamics of hurricanes in order to improve the accuracy of the forecasts. Turbulent transport processes play a vital role in TCs maintenance and intensification (Emanuel, 1995; Persing & Montgomery, 2003). This is partly because enthalpy uptake from the ocean and absolute angular momentum loss in hurricanes are mainly induced by turbulent fluxes (Zhang & Montgomery, 2012). In ARW, the horizontal turbulent fluxes are modeled using a mixing length parameterization that does not vary in time or space. However, the mixing length in Prandtl's hypothesis, which forms the basis of these models, is a variable that depends on the nature of the flow (Pope, 2000). This lack of hurricane physics considerations in existing turbulence closures primarily explains the poor performance of both Smag2D and TKE schemes in simulating real hurricanes intensity compared to the NoHorizTurb configuration.

Recent numerical (Momen et al., 2021) and observational studies (Zhang, 2010; Zhang & Montgomery, 2012) have shown that the size of the most energetic turbulent eddies in hurricane BLs are smaller than in typical ABLs due to the strong rotations in hurricanes. Based on these findings, we hypothesize that the mixing length scale in hurricanes should be reduced compared to typical ABLs to improve the accuracy of hurricane forecasts. This improvement is especially expected for wind intensity predictions since the mixing length directly modulates the diffusion in hurricanes. To investigate this hypothesis and to comprehensively characterize the impacts of mixing-length values on real hurricane simulations, we performed a sensitivity analysis of L_h by varying the newly defined coefficient, cL_h, in Equation 6 to 0.25, 0.5, and 1.5. This analysis provides notable insights into the impacts of this parameter on hurricane dynamics. Furthermore, it provides guidance on how L_h should be modified in operational models to improve the forecasts. This section presents the results of the second set of WRF simulations and compares them with the default parameterization in ARW ($cL_h = 1$) for Smag2D and TKE turbulence models.

The horizontal mixing length significantly impacts the simulated hurricane structure and intensity. Figure 5 displays the wind speed and pressure contour maps at 500 m elevation for decreased (Figures 5a and 5c) and increased L_h values (Figures 5b and 5d) in Hurricanes Katrina and Maria. As Figure 5 shows, the simulations with reduced L_h produce stronger hurricanes in comparison to larger L_h cases (compare the darker red colors in Figures 5a and 5c with lighter colors in Figures 5b and 5d). The contours for the small L_h Smag2D simulation (Figure 5a) exhibit regions with an average velocity of ∼67 m/s and the sea-level pressure (SLP) at the eye of the hurricane is as low as ∼936 hPa, while the contour with the large L_h (Figure 5b) has an average velocity of ∼52 m/s and the eye SLP of ∼956 hPa. The results indicate almost 15 m/s difference in terms of maximum intensity and ∼20 hPa in terms of SLP between these two cases, underscoring the significant impact of L_h on simulated hurricane intensity and structure. This effect is also visible in higher resolution runs of the TKE turbulence model (Figures 5c and 5d), even though the difference is reduced to ∼6 m/s in wind intensity and ∼8 hPa in SLP. The comparison between the rest of the cases can be found in Figures S5.1 and S5.2 in Supporting Information S1.

The hurricane intensification in decreased L_h cases is expected since reducing L_h leads to a decreased horizontal diffusion that spins up the hurricane vortex. Our findings are thus consistent with other studies that conducted idealized hurricane simulations using other NWP models (Bryan & Rotunno, 2009; Zhang & Marks, 2015). Furthermore, a lower eye SLP is observed when L_h is reduced. This is also expected since a lower SLP creates a

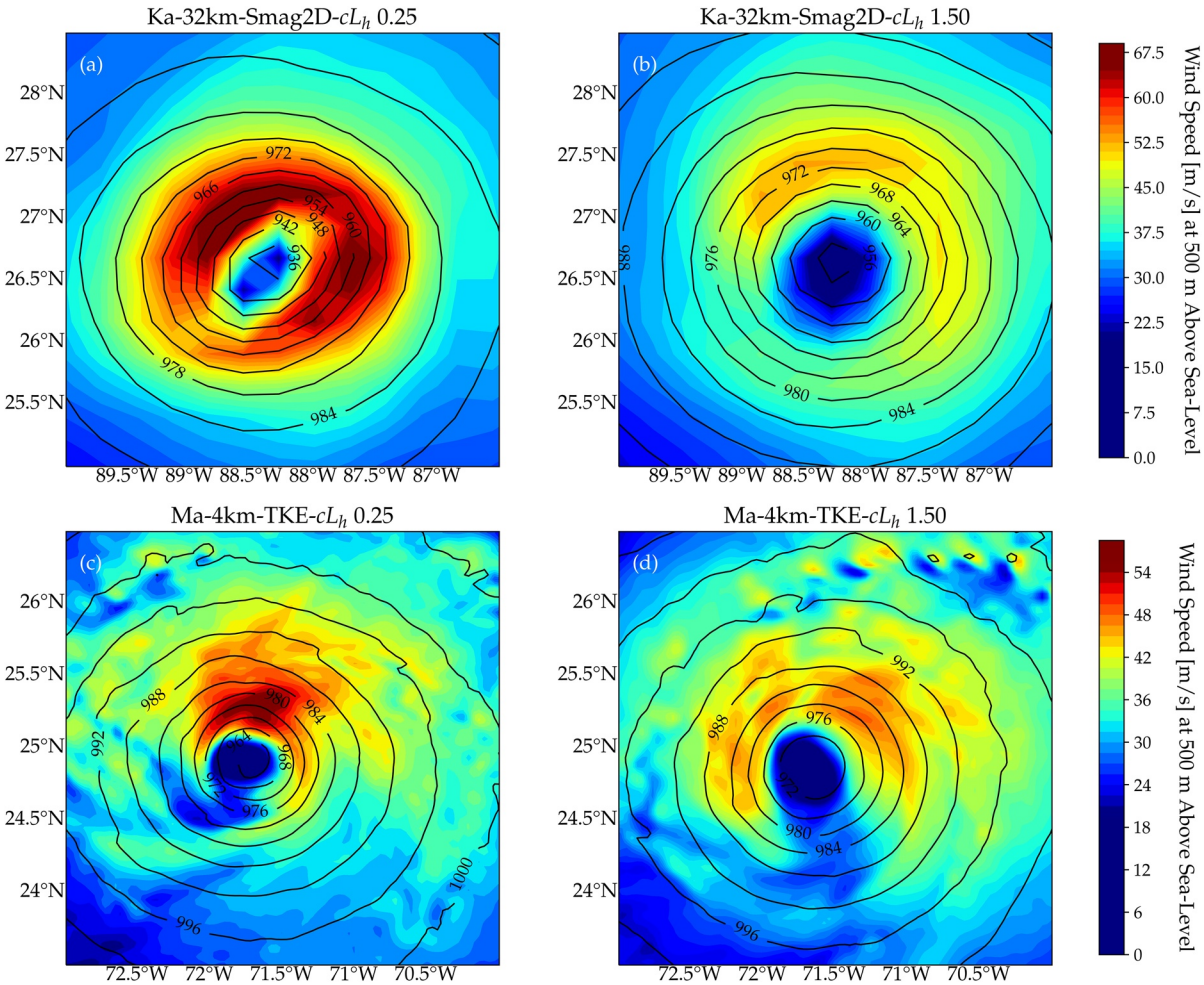

Figure 5. Contour maps depicting the wind speed at 500 m of altitude overlayed by the sea-level pressure isolines for hurricanes Katrina (after 9 hr of simulation) and Maria (after 30 hr of simulation) respectively for 32 and 4 km grid sizes: (a and b) are the maps for Smag2D turbulence model for respectively a cL_h of 0.25 and 1.5, (c and d) are the maps for turbulent kinetic energy (TKE) turbulence model for respectively a cL_h of 0.25 and 1.5.

larger pressure gradient between the eye and the eyewall and thus generates higher wind speeds (this can be seen from the simplified momentum equation in the inner eye region: $\frac{1}{\rho}\frac{\partial p}{\partial r} \sim \frac{u_\theta^2}{r}$).

The impact of changing L_h is not expected to be the same on different turbulence models as Equations 5 and 7 indicate. To assess the sensitivity of each model, the wind speed versus the distance from the hurricane center and SLP are depicted in Figure 6 after 22 hr of simulation using the Smag2D and TKE schemes. While both models start with the same initial conditions (black line), they evolve differently after 22 hr of simulations (color lines) for Hurricane Gustav. Although both models are highly sensitive to the L_h values, the figure shows that Smag2D is more sensitive to cL_h changes than TKE. The difference in maximum wind speed between cL_h = 0.25 and cL_h = 1 (the default value) is ~4 m/s for Smag2D compared to ~1.5 m/s for TKE. Moreover, Smag2D resulted in ~8 hPa SLP difference at the hurricane center between cL_h = 0.25 and cL_h = 1, almost 2 times the SLP difference for the TKE. These results suggest that the eddy viscosity formulations in turbulence models modulate their responses to changes in the mixing length values and this should be considered for designing hurricane specific turbulence closures.

To comprehensively examine the impact of L_h among different turbulence models and grid resolutions, all conducted cases were examined. The wind speed versus radius and SLP for Hurricane Katrina and Irma are shown in Figure 7 and for the rest of the cases are depicted in Figures S6.1–S6.8 in Supporting Information S1. In general, reducing L_h increases the average maximum wind intensity in simulated hurricanes and decreases their

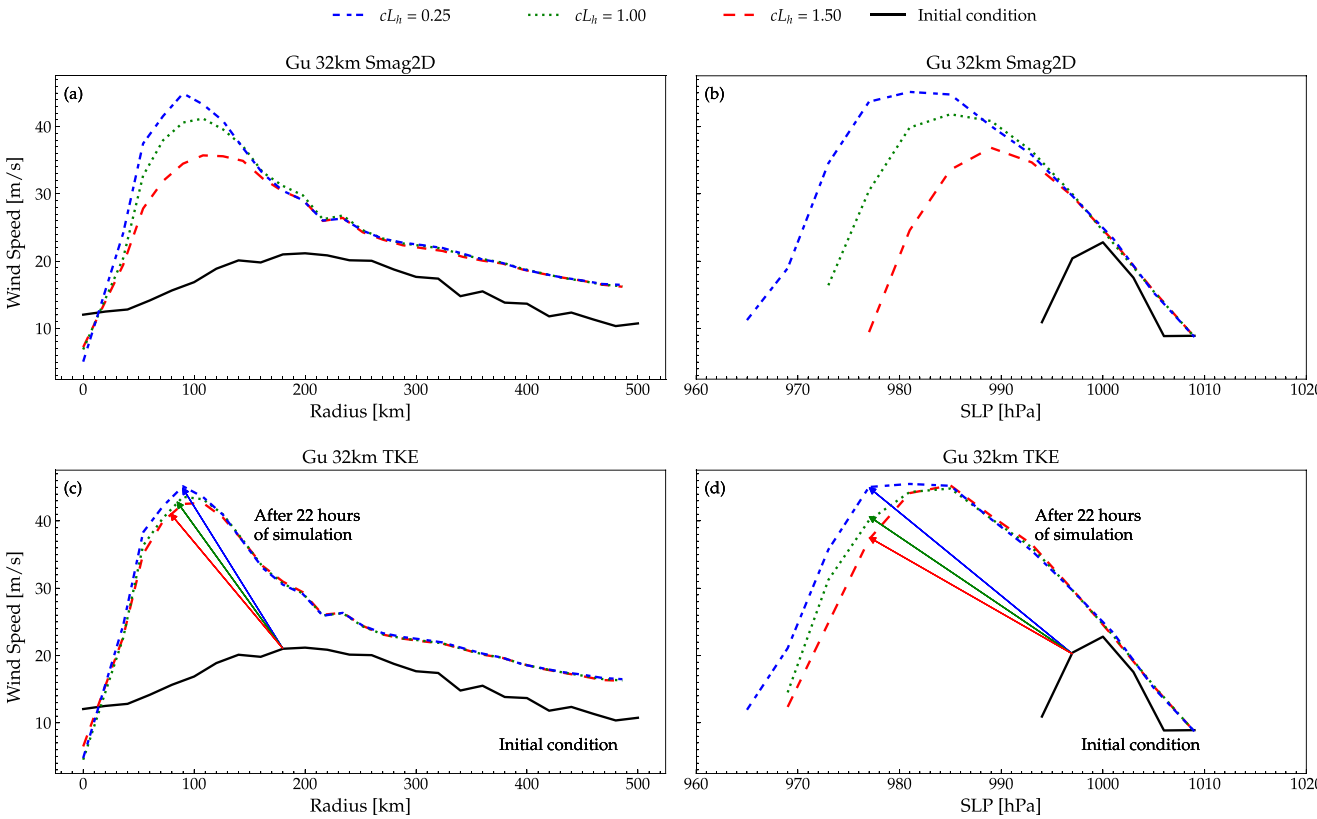

Figure 6. Temporal evolution of Hurricane Gustav's intensity according to three horizontal mixing length scales cL_h = 0.25, cL_h = 1.0 (default) and cL_h = 1.5. The straight black line (—) represents the initial condition of all the simulations. Other lines represent a one-hour temporal average of the data after 22 hr of simulation for hurricane Gustav in terms of the radius (a and c) and the sea-level pressure (SLP) (b and d). The *y*-axis of (b) is the same as (a) and the *y*-axis of (d) is the same as (c).

size (see Figure S.7 in Supporting Information S1). Figure 7 also confirms the previous finding that Smag2D is more sensitive to L_h changes than TKE. In fact, the higher sensitivity of Smag2D to L_h changes is also expected from its horizontal diffusion formulation. The horizontal eddy viscosity (K_h) is linearly related to L_h in the TKE model (Equation 5), while K_h is related to the square of L_h in the Smag2D model (Equation 7).

Another common pattern that can be observed from Figure 7 is that the wind speed values for all L_h values converge far from the hurricane center (radius $\gtrsim$ 200 km). This indicates that the major impact of changing L_h is on high wind speeds near the eyewall region. As we go far from the hurricane center both wind speeds and horizontal diffusion impacts significantly decrease. This result is expected since the horizontal diffusion depends on the horizontal gradient of the wind speed and such gradients are typically larger near the eyewall region. Therefore, modulating L_h at large distances from the hurricane center will have smaller effects on the wind speed magnitudes.

Next, the impacts of horizontal mixing length on the vertical profiles of radial and tangential wind in hurricanes are investigated. Each hurricane case has its own parameter space that evolves nonlinearly in time. Hence, choosing one radius for plotting vertical profiles of radial and tangential wind in all cases is inappropriate since for one case that radius can be at the outer eye region whereas for another case that could be at the eyewall. For this reason, we normalized the radius and the SLP respectively by the radius of the maximum wind (R_{MW}) and SLP of the maximum wind (SLP_{MW}). The third and fourth columns of Figure 7 depict the normalized data of the first two columns. These data were employed to choose the proper distance in order to examine the vertical profiles of radial and tangential wind in different cases using non-dimensional variables. Figure 8 displays the vertical profiles of radial and tangential wind velocity for Hurricanes Katrina 32 km and Irma 4 km averaged over 1 hour time. The results for other cases are shown in Figures S8.1–S8.5 in Supporting Information S1. In the first column of Figure 8, the wind profiles of the inner eyewall region are displayed, and in the second column, the wind profiles of the outer eyewall are shown.

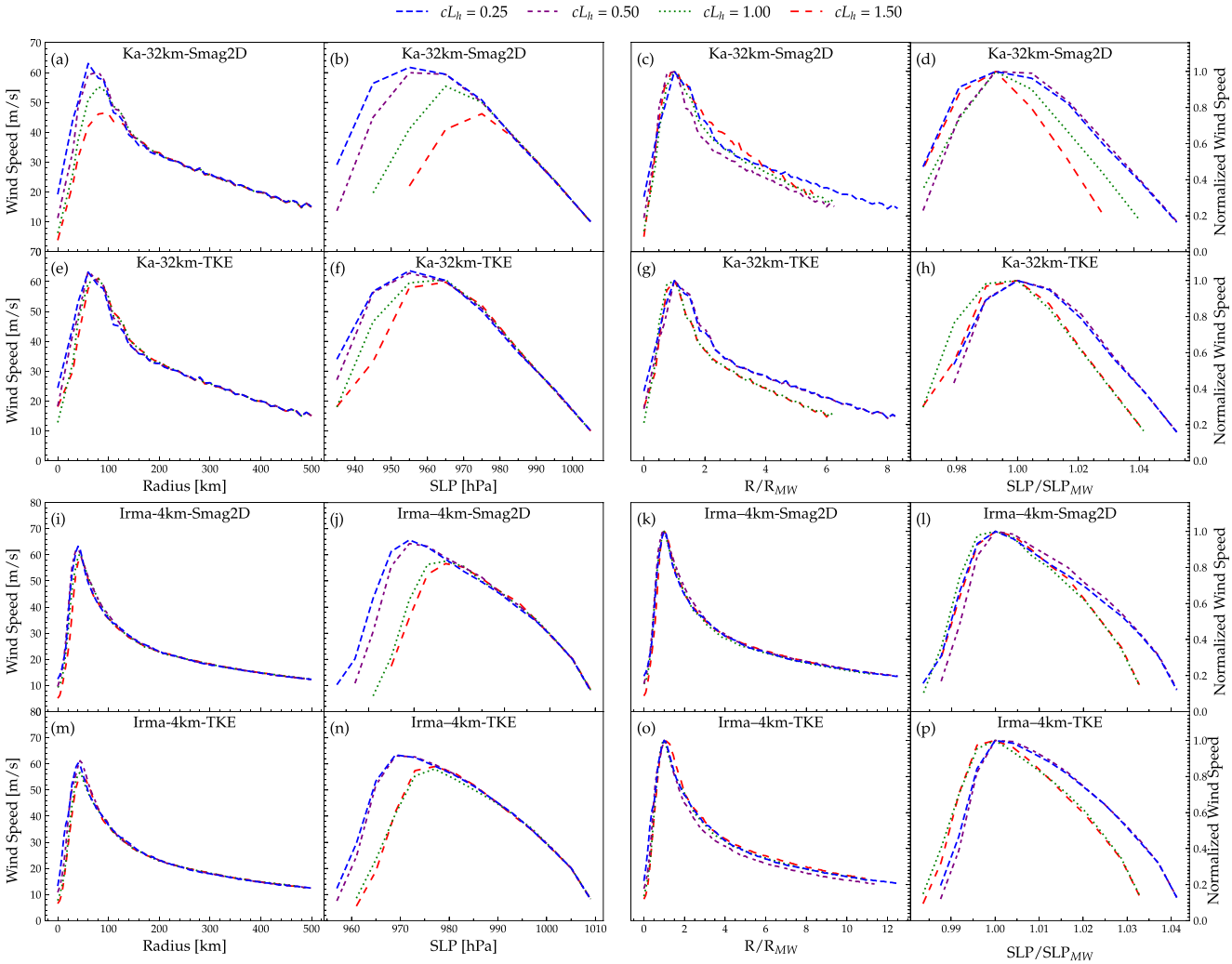

Figure 7. A one-hour temporal average of 500 m above sea-level wind speed in terms of both radial distance from the hurricane's eye and sea-level-pressure for hurricanes (a, b, e and f) Katrina (32 km grid) and (i, j, m and n) Irma (4 km grid). 500-m wind speed is normalized for hurricanes (c, d, g and h) Katrina and (k, l, o and p) Irma with respect to the maximum wind speed (V/Vmax), the radius is normalized with respect to the radius of the maximum wind (R/R$_{MW}$) and sea-level pressure is normalized with respect to the maximum wind speed's sea-level pressure (SLP/SLP$_{MW}$).

In general, the wind profiles corresponding to $cL_h = 0.25$ presented larger maximum wind values both in the inner and outer eyewall regions. Similarly, increasing the cL_h value to 1.5 resulted in typically lower wind speeds (red lines in Figure 8). The maximum wind speed occurs at ∼ 500–1500 m height in most cases. Furthermore, these results indicate that changing the horizontal mixing length can cause a remarkable impact on both the magnitudes and shapes of the vertical profiles of radial and tangential wind. Changing the horizontal mixing length impacts both radial and tangential wind velocities. Decreasing the cL_h, increases the inflow; this intensifies the hurricane vortex and the tangential velocity component (see the wind contours in Section S9 in Supporting Information S1).

To better elucidate why decreasing L_h results in intensified winds, we examined the turbulent stresses of each case. To achieve this goal, we implemented an in-house module for online calculation of the total Reynolds stresses (resolved + SGS) of all cases in ARW using a cylindrical transformation similar to (Momen & Bou-Zeid, 2017). Figure 9 depicts the Reynolds stresses of all turbulence models for Hurricane Katrina 32 km and Irma 4 km runs. The stress profiles for other cases are shown in Figures S11.1–S11.5 in Supporting Information S1. Reducing L_h decreases the horizontal diffusion and the characteristic size of turbulent eddies. This decrease in the size of turbulent mixing eddies also leads to a decrease in the vertical momentum diffusion. The results show that an increase in L_h consistently increases the $\overline{u'_r u'_\theta}$, $\overline{u'_r u'_z}$ and $\overline{u'_\theta u'_z}$ stresses inside and outside the eyewall for Smag2D and TKE. This increase in the Reynolds stresses magnitudes increases their vertical gradients ($\partial \overline{u'_\theta u'_z}/\partial z$ and

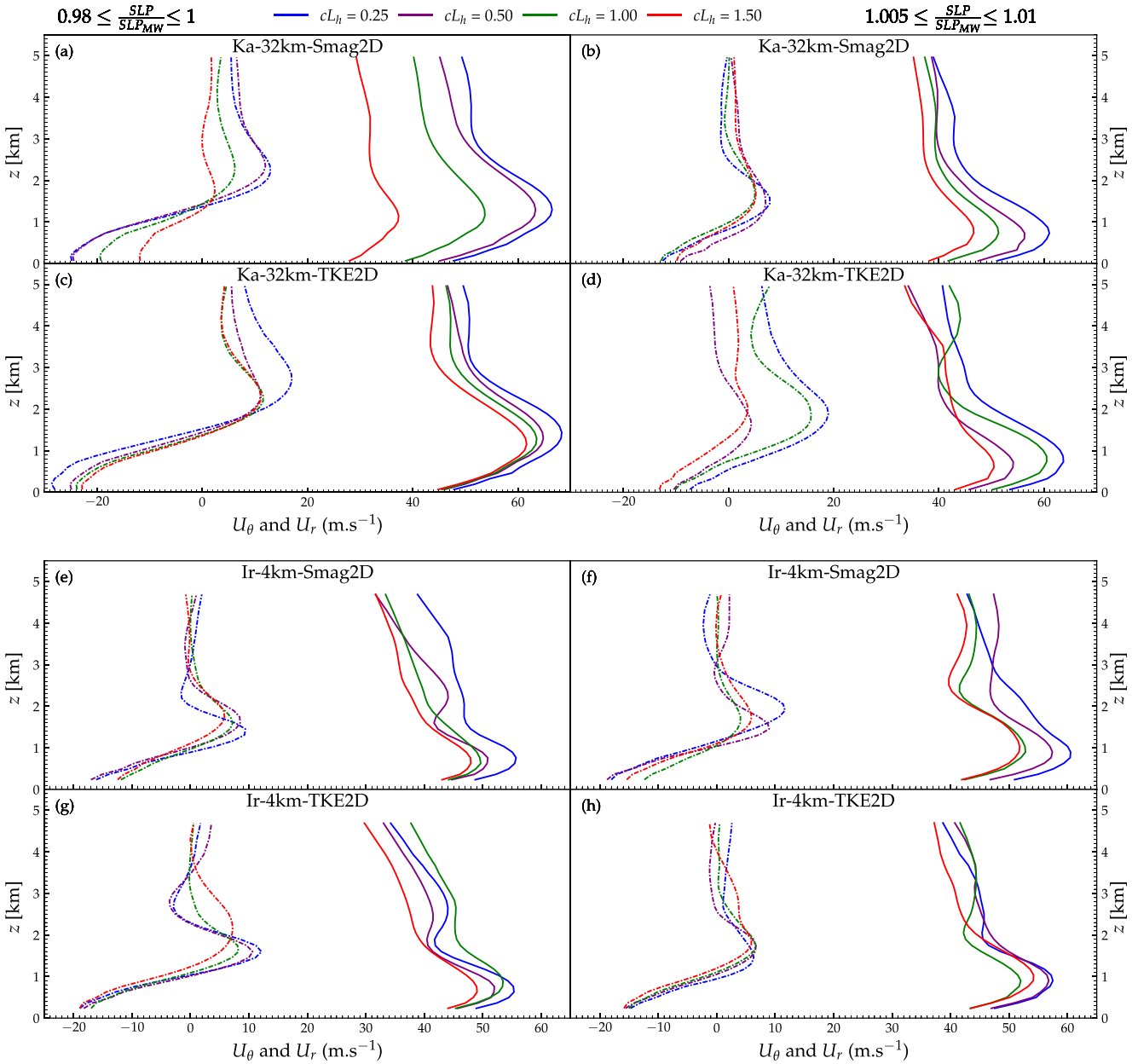

Figure 8. One-hour temporal average of the radial (U_r) and tangential (U_θ) wind profiles for different horizontal turbulence models and mixing length scales. The dashed-dotted lines represent the vertical profiles of the radial wind, and the solid lines depict the vertical profiles of the tangential wind velocity. The exhibited results are for Hurricanes Katrina (32 km resolution) and Irma (4 km resolution).

$\partial\overline{u_r'u_z'}/\partial z$) as can be seen in Figure 9 (slope of the lines). Hence, the flow tends to be more dissipative as the L_h increases and additional generated frictional forces (e.g., via $-\rho\left[\partial\overline{u_\theta'u_z'}/\partial z + \overline{u_r'u_\theta'}/r + \partial r\overline{u_r'u_\theta'}/r\partial r\right]$ in the tangential momentum equation) decrease the intensity of the simulated hurricanes. Furthermore, changing the L_h not only impacts the horizontal diffusion but also significantly modulates the vertical diffusion magnitudes and distributions ($\overline{u_r'u_z'}$ and $\overline{u_\theta'u_z'}$).

The intensification of the simulated hurricanes can also be explained by examining the radial momentum budget balance. When L_h is reduced, the diffusion term decreases as expected from Equations 3, 5 and 7 and Figure 9. For relatively the same pressure gradient force, the centrifugal and Coriolis forces need to increase to compensate

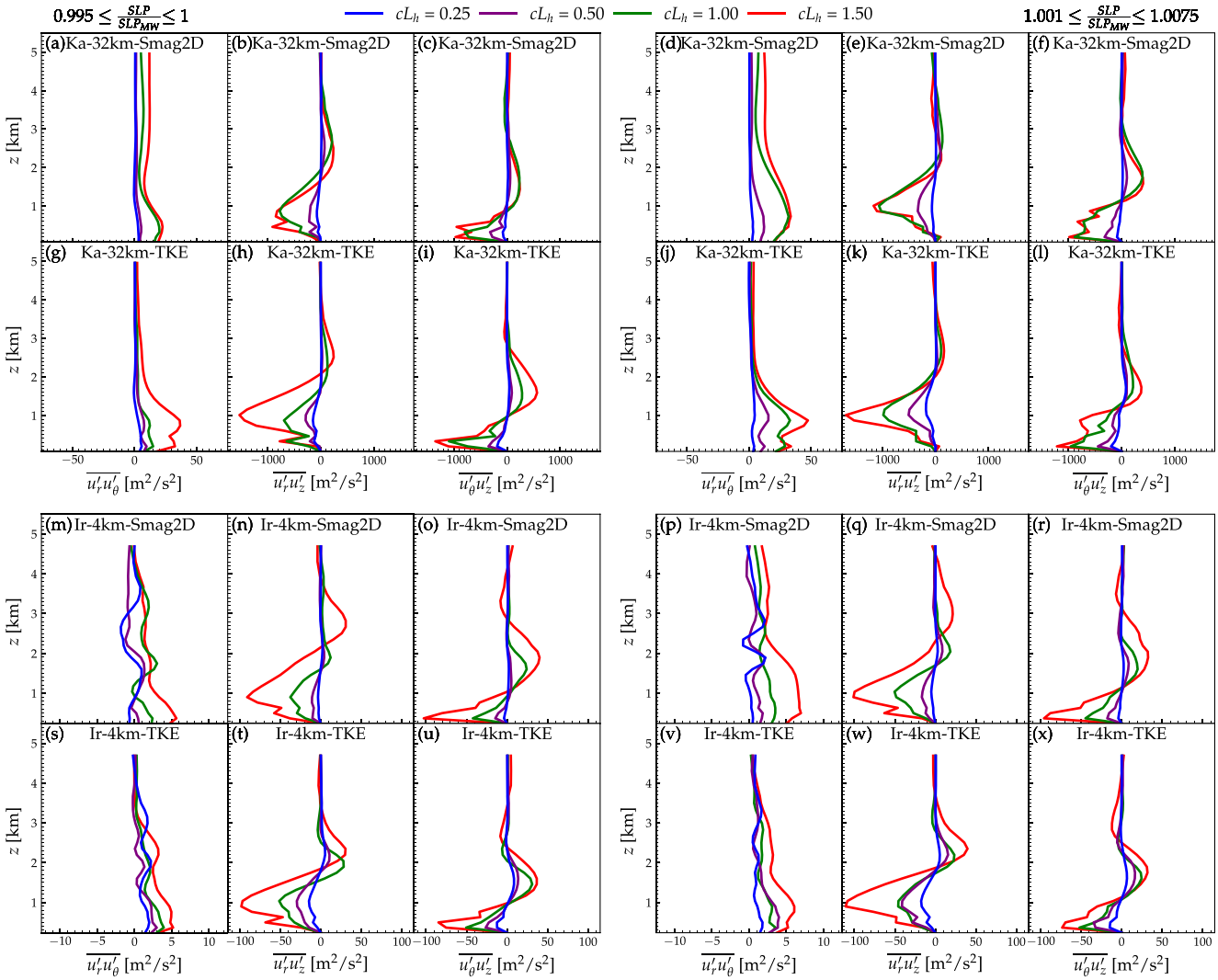

Figure 9. One-hour temporal average of the resolved + SGS stresses for Hurricanes Katrina (top) and Irma (bottom), simulated with the Smag2D and turbulent kinetic energy (TKE) turbulence models.

for this decrease in the diffusion. This leads to higher wind intensities for decreased L_h as shown in Section S10 in Supporting Information S1.

The magnitude of the Reynolds stresses considerably decreases as we increase the grid resolution. This means changing L_h in low-resolution simulations causes larger absolute changes in the Reynolds stress profiles. Moreover, this trend indicates that lower-resolution cases are more dissipative than high-resolution cases. Therefore, their wind speeds should be generally lower than in high-resolution cases (as can also be seen in Figure 2). This is attributed to the way the mixing length is formulated in ARW as a function of grid spacings (Equations 5–7). As the grid resolution increases, the L_h grows yielding a larger diffusion both horizontally and vertically. This improper formulation can lead to over-dissipative hurricane simulations in low-resolution configurations and result in inaccurate hurricane intensity forecasts.

3.3. The Impacts of the Horizontal Mixing Length on the Accuracy of WRF Simulations

In the previous section, the impacts of the horizontal mixing length on hurricane structure and dynamics were assessed. In this section, we examine the effects of varying the L_h on the accuracy of real hurricane simulations by comparing wind profiles, intensity, and track results with observations. To this end, first we compare the forecasted wind profiles with the GPS dropsonde data. Six hours of ARW simulations that coincided with the

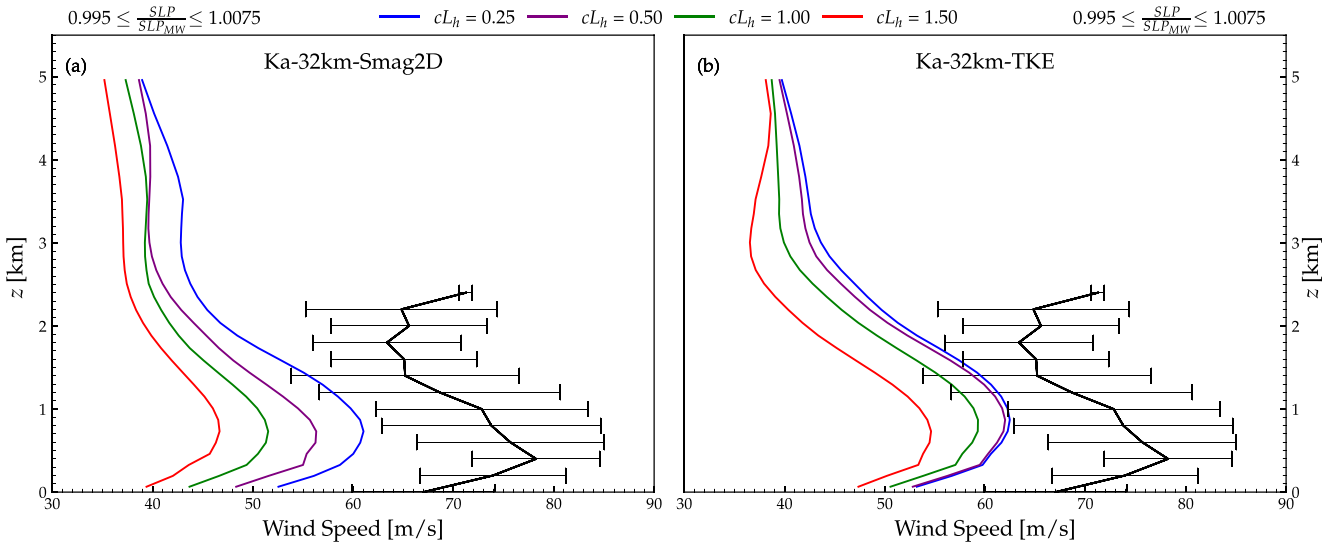

Figure 10. Comparison of the 6-hr average vertical profiles of wind speed with the GPS dropsonde data for Hurricane Katrina. The forecasted data for both Smag2D and turbulent kinetic energy (TKE) turbulence models are averaged at the eye wall, around 50 km from the eye and the dropsondes were dropped at the same distance from the hurricane's eye. The horizontal black error bars represent the standard deviation of the dropsonde data for each recording altitude.

available GPS dropsonde data for Hurricane Katrina were analyzed. The average wind profiles at the hurricane's eyewall are compared in Figure 10. In both Smag2D and TKE, decreasing L_h increases the wind speed as expected. Furthermore, decreasing L_h from its default value improves the accuracy of wind speed forecasts when compared to observations. As the figure indicates, at 1000 m of altitude, the default mixing length parametrization predicted a wind speed of ∼50 m/s for the Smag2D model and ∼58 m/s for the TKE. Once we decrease the horizontal mixing length to 1/4th of its default value, the wind speed at 1000 m increased respectively to ∼59 m/s and ∼63 m/s, bringing it closer to the average measured value of ∼73 m/s.

To evaluate the ARW's capability to forecast the correct maximum wind speeds in hurricanes, we compared the histograms of extreme winds of simulations and observations. Figure 11 displays the probability density histograms (PDH) of the forecasted wind speed at all points below 1000 m against dropsonde data recorded during the same time for Hurricane Katrina. For this comparison, we used a 5 m/s wind speed range for each bar starting from 40 m/s, and considered 6 hours of simulations that coincided with the observations. As the figure shows, the green bars reflecting the PDH of the default ARW's mixing length parametrization tend to shrink as the wind speed increases, and they disappear at 65 m/s for Smag2D and 70 m/s for TKE. However, when we set cL_h to 0.25, the blue bars continue alongside the same axis and reach 70 m/s for Smag2D and 75 m/s for TKE

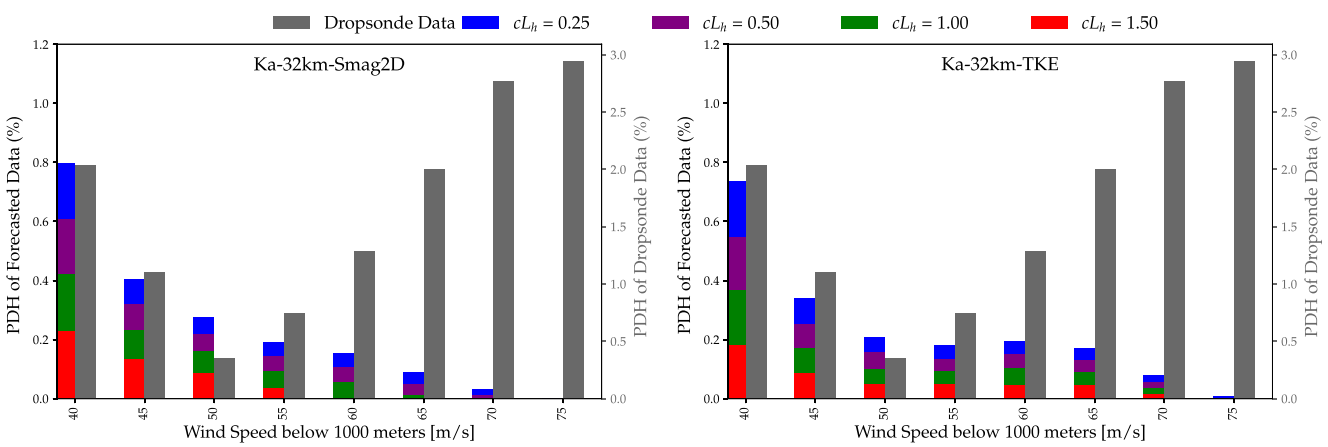

Figure 11. Comparison of the probability density histogram of Hurricane Katrina's wind speed below 1000 m of altitude during 6 hr of simulation for different mixing length scales versus the GPS dropsonde data.

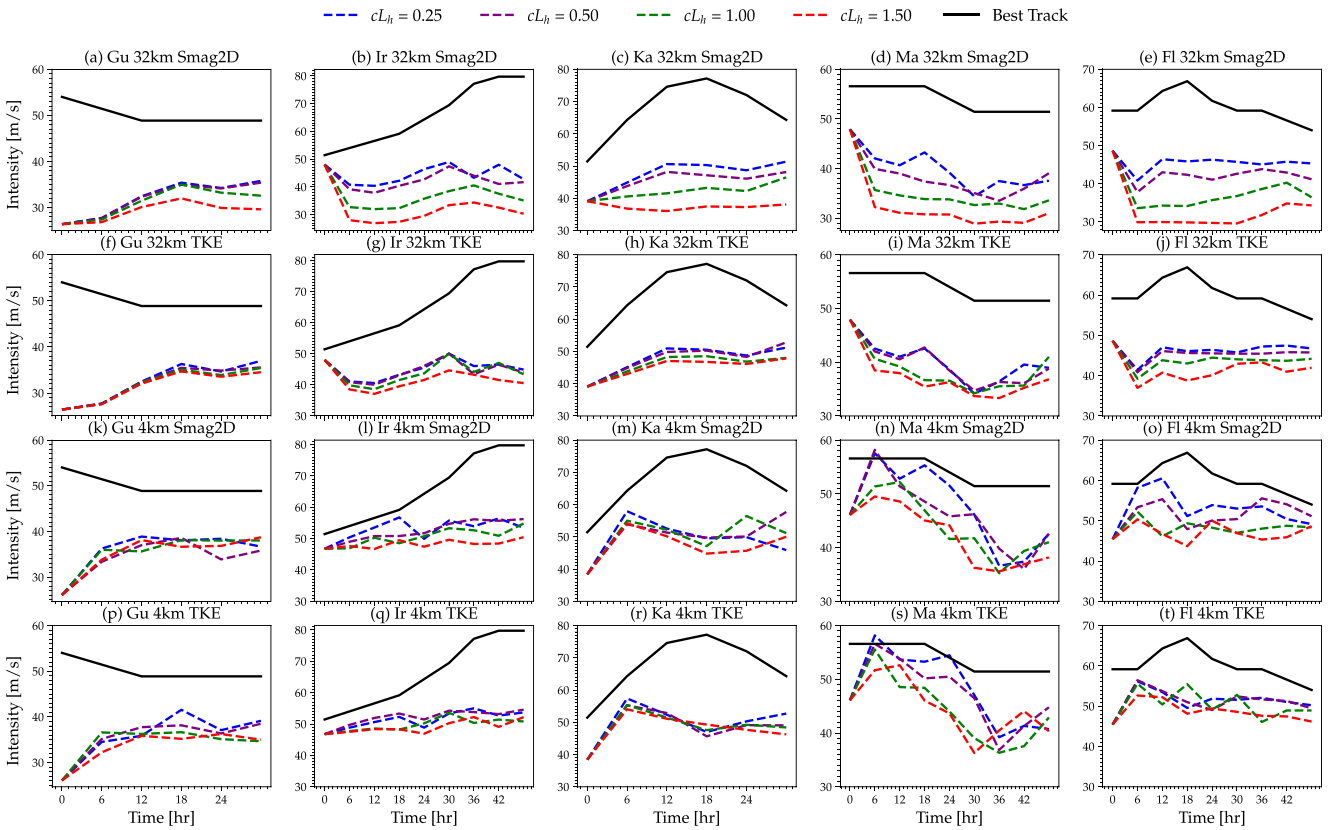

Figure 12. Smag2D and turbulent kinetic energy (TKE) forecasted near-surface wind speed (10 m of altitude) among all considered hurricanes, for 32 and 4 km grid sizes and for different L_h values. The solid black lines represent the best track wind speeds.

more consistent with the observed data. This trend demonstrates that decreasing the default mixing length scale in ARW, improves its capability to predict extreme winds in hurricanes.

To corroborate the generality of the previous findings, the forecasted surface wind intensities of all cases were compared with the observed best track speed data. Figure 12 displays the wind intensity time series of WRF simulations for the five considered hurricanes using a coarse (32 km) and fine (4 km) grid resolution. As the figure indicates, the wind intensity forecasts generally improve as the default L_h in WRF is reduced in all hurricanes. For instance, case Ka_32km_Smag2D had a maximum speed of ∼43 m/s after 24 hr of simulation using the default configuration (green line in Figure 12c). However, when we decreased cL_h from 1.0 to 0.25, the hurricane intensity increased to ∼50 m/s (blue line), bringing it closer to the observed data (black line). This improvement is also noticed in other hurricanes and the TKE model simulations. Nonetheless, the TKE turbulence model appears to be less sensitive to the horizontal mixing length changes than the Smag2D model.

To gain a comprehensive understanding of the impacts of the horizontal mixing length variations on the simulated hurricane accuracy, the errors of all new conducted simulations were averaged. Figure 13 summarizes the forecasted intensity and track errors for all considered hurricanes and turbulence models. The figure clearly indicates that the hurricane wind intensity forecast improves by decreasing the default L_h value. Smag2D is more sensitive to these variations than TKE consistent with previous findings. For Smag2D, $\text{MAPE}_{\text{Intensity}}$ was improved by 22.7% for the 32 km grid size and 14.5% for the 4 km grid size compared to the default WRF simulations ($100\%-100\% \times \text{MAPE}_{\text{Intensity},cLh0.25}/\text{MAPE}_{\text{Intensity},cLh1.0}$). The highest wind intensity improvement for Smag2D's 32 km resolution simulations was for Irma by ∼27.5% and the lowest was for Gustav by ∼5.5%. For the TKE model, $\text{MAPE}_{\text{Intensity}}$ was improved by ∼8.16% for the 32 km grid size and ∼10.7% for the 4 km grid size compared to the default WRF simulations. Another remark from this figure is that for $cL_h = 1.5$ the intensity error of Smag2D is higher than TKE; however, as L_h decreases, Smag2D and TKE intensity errors converge to approximately a similar value.

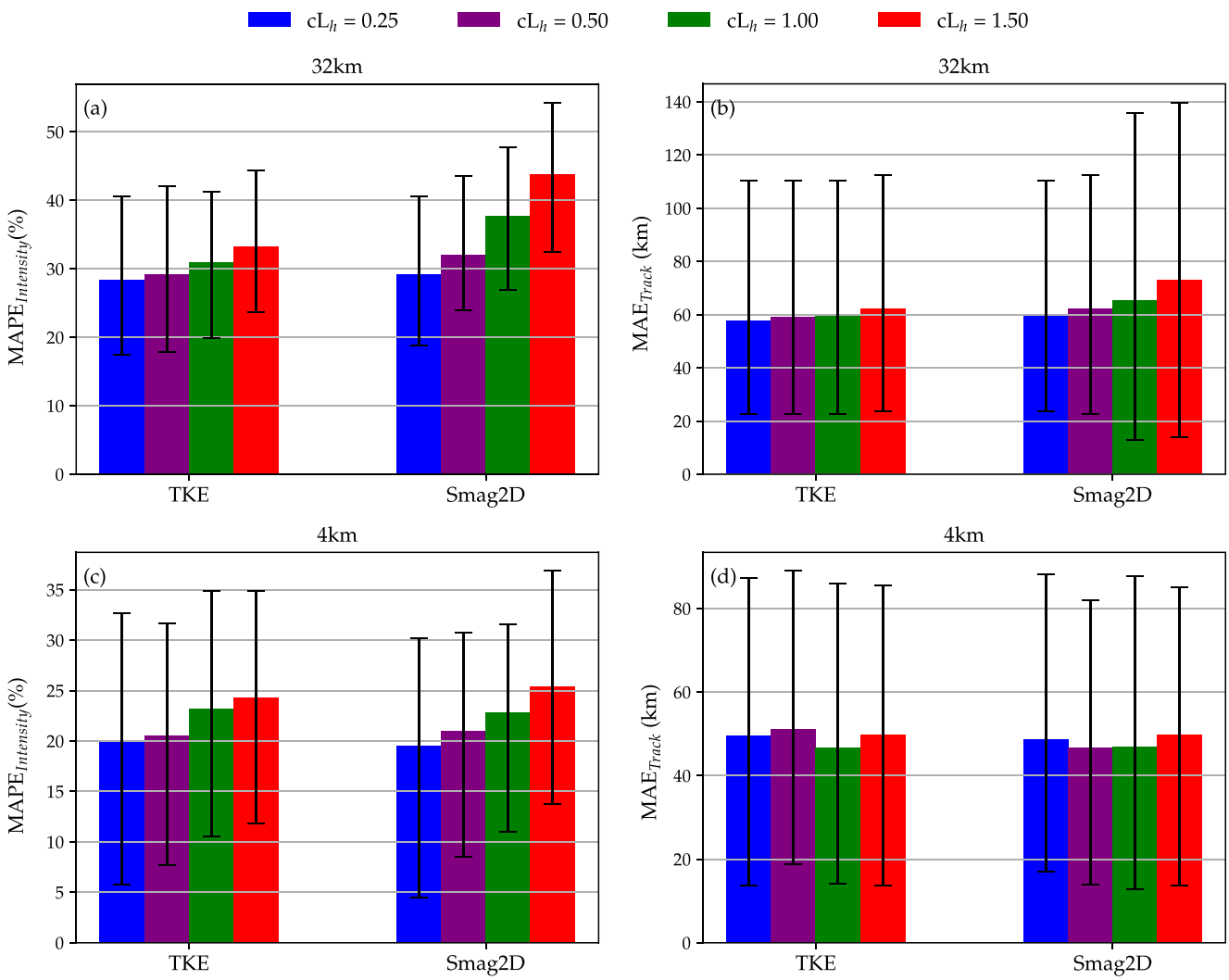

Figure 13. Average error of all hurricanes in terms of horizontal mixing length scales and turbulence models for 32 and 4 km grid sizes. The solid black error bars represent the 10th and 90th percentile intervals.

The final metric that we investigate is the track error. As Figures 13b and 13d exhibit, track error improvements with changing the horizontal mixing length are not as consistent as the intensity error. In particular, clear improvements in the MAE_{track} for Smag2D and TKE with 32 km grid resolution were noticed. However, no clear statement can be made about the MAE_{track} of Smag2D and TKE runs with 4 km grid resolution. This can be attributed to the fact that the impact of small-scale processes on hurricane track prediction is complex since track forecasts depend on many large-scale processes (Fierro et al., 2009).

The impacts of adjusting the horizontal mixing length were also tested for weak hurricanes (category 1–2). We simulated Hurricanes Sally (2020), Jerry (2019), Helene (2018), Gordon (2012), and Nadine (2012). Please refer to Section 12 of the Supporting Information S1 for further details about these simulations. The results suggest a similar trend for both models across coarse and fine grid resolutions. By decreasing the default L_h ($cL_h = 0.25$), the intensity forecasts improved by ~5%–26% on average for the TKE and Smag2D models among the considered weak hurricanes.

To sum up, it was shown that the existing horizontal turbulence models in ARW perform poorly in hurricane intensity forecasts. On average, they even underperformed the NoHorizTurb cases for intensity predictions among the five considered hurricanes. This poor performance is associated with the inaccurate parameterization of horizontal diffusion in ARW that is neither designed nor adjusted for hurricane flows. In particular, none of these models consider the impacts of strong rotation in hurricanes in their parameterizations. In fact, rotation in

turbulent flows can influence the transfer of energy from the mean to the turbulence. Prior studies have shown that rotation can suppress or enhance turbulence production in rotating shear flows (Arolla & Durbin, 2013; Cazalbou et al., 2005; Durbin, 2011; Tritton, 1992). These studies demonstrate that turbulence stabilization in rotating channel flows occurs when the background vorticity is larger than the local shear vorticity or when they are in the same direction. In hurricanes, we can make an analogy with such flow systems to determine the turbulence suppression or growth regions due to the rotation. If we model a hurricane using a circulating Rankine vortex flow, the background vorticity will be typically larger than the local shear vorticity and thus the turbulence is expected to be suppressed in such flows. Furthermore, previous numerical and observations studies (Momen et al., 2021; Zhang, 2010; Zhang & Montgomery, 2012) showed that the energy-containing turbulent eddy sizes in hurricane BLs are smaller than regular ABLs which indicates their suppression by strong rotation in hurricanes. Building upon these studies, we found and demonstrated that one of the major deficiencies of the existing turbulence models is their overestimated horizontal mixing length for real hurricane forecasts. By reducing the default L_h values when setting $cL_h = 0.25$, the adjusted Smag2D and TKE models improved both the intensity and track forecasts of the NoHorizTurb cases by ∼2%–5% on average for the 4 km simulations. Furthermore, we managed to decrease the $\text{MAPE}_{\text{intensity}}$ of the default Smag2D and TKE models between ∼8%–23% on average among the considered hurricanes.

Our results indicate that adjusting the horizontal mixing length in the existing turbulence models can remarkably improve their simulated hurricane intensities. In fact, the current study motivates further research on developing more appropriate horizontal turbulence models for hurricane flows that include structural changes in the existing models. For instance, the impacts of rotation or anisotropy in hurricane flows can be represented in a new horizontal turbulence parameterization. Moreover, vertical diffusion plays an essential role for improving the simulated storm size and intensity (Zhang et al., 2015). Indeed, our preliminary tests indicate that modifying the vertical diffusion in PBL schemes (not shown) can remarkably impact the intensity predictions in WRF. These efforts will be shown in future work as the focus of this paper is to delineate and characterize the deficiencies in existing horizontal turbulence models for simulating real hurricanes.

It is important to note that these findings regarding turbulence modeling in hurricane simulations are general and can be applied to improve any NWP model (e.g., Hurricane WRF). Although we relied on WRF-ARW to conduct this study, the main findings are valid and applicable to other NWP models as the physics behind these models remains the same. Therefore, this paper could potentially lead to further improvements in the operational models such as Hurricane Analysis and Forecast System (HAFS). In particular, the present work would provide some guidance for the improvement of other hurricane models that implement Smagorinsky (Zhang et al., 2018) and TKE type turbulence closures (Gopalakrishnan et al., 2021).

4. Conclusion

This study characterizes the interacting effects of grid resolution, horizontal turbulence models, and horizontal diffusion length scales on the accuracy of real hurricane simulations. Five category 4 or 5 hurricanes over the Atlantic Ocean (Katrina, Gustav, Irma, Maria, and Florence) were simulated using the WRF-ARW code version 4.1. In total, 135 WRF simulations were conducted to examine the parameter space of the problem. 75 runs investigated the combined effects of grid resolution and default turbulence models (NoHorizTurb, Smag2D, and TKE) on real hurricane prediction accuracy. 60 cases assessed the impacts of changing the horizontal mixing length scales on hurricane dynamics and forecasts. Two major metrics (intensity and track errors) were primarily employed to evaluate the accuracy of the forecasts. In summary, the key findings of this study are:

1. The considered default horizontal turbulence models underpredict the intensity of the studied hurricanes in comparison to the observed best track wind speed data. As the grid resolution increases, the intensity forecasts of all considered turbulence models improve. The intensity predictions of the 32 km grid resolution cases were improved by an average of ∼42% when the grid size was changed to 2 km among the three considered turbulence models. Such a consistent trend was not found for the track error similar to a handful of previous studies.
2. The default horizontal turbulence models (Smag2D and TKE) yield higher hurricane intensity errors compared to the NoHorizTurb method on all grid sizes. This indicates that the default horizontal turbulence models in WRF are not appropriately designed for hurricane flows such that turning them off generally improves the intensity forecasts. One of the major drawbacks of horizontal turbulence models in WRF is the

implementation of the horizontal mixing length scale, which does not consider the physics of the flow such as the strong rotation in hurricanes.

3. Our simulations showed that decreasing the default horizontal mixing length scale in ARW significantly improves the intensity forecasts of the considered real hurricane cases. This confirms our hypothesis based on previous studies that showed the characteristic horizontal eddy size in hurricane BLs should be smaller than typical ABLs (Momen et al., 2021; Zhang, 2010) since strong rotation in hurricane flows suppresses turbulence production. Furthermore, decreasing the default L_h led to enhanced vertical profiles of radial and tangential winds and extreme wind predictions in Hurricane Katrina simulations when compared to the GPS dropsonde data.
4. Reducing the default horizontal mixing length strengthens the hurricane vortex, results in lower hurricane center SLP, and decreases the radius of maximum wind. On the other hand, larger L_h values yield a smaller storm intensity due to their excessive horizontal and vertical Reynolds stresses ($\overline{u_r'u_\theta'}$, $\overline{u_r'u_z'}$ and $\overline{u_\theta'u_z'}$) that dissipate the hurricane's motion.
5. The results indicate that the Smag2D model is more sensitive to the horizontal mixing length scale variations than the TKE scheme. This higher sensitivity of Smag2D is also expected from its K_h formulation in which it is related to the square of L_h while in the TKE model K_h is linearly related to the L_h. By reducing the default L_h, we managed to decrease the $\text{MAPE}_{\text{intensity}}$ on average between ~8–23% for the Smag2D and TKE models for both low and high resolutions. Similar to the grid resolution analysis, no consistent trend for track error was observed by changing the L_h among the considered cases.

Our results demonstrate that the default horizontal mixing length scales in ARW need to be revisited for hurricane simulations as they are currently over-dissipative and significantly underpredict the intensity of real hurricanes. The paper provides notable insights into the role of this parameter on hurricane dynamics and some guidance to enhance its parameterizations in operational NWPs. Hence, the findings are useful to improve the horizontal diffusion parameterizations in weather and climate models, especially for hurricane forecasts.

Data Availability Statement

The data for the 32-km 3D snapshots of the velocity field, pressure, and temperature in NetCDF format can be found in (Romdhani et al., 2022). The readers can access the simulation data of this paper and an instruction to load them via Python codes in this repository.

Acknowledgments
The authors acknowledge the Department of Civil and Environmental Engineering at the University of Houston for support via startup funds. Jun Zhang was supported by NOAA Grant NA19OAR4590239 and ONR grant N00014-20-1-2071. The simulations were performed on the computing clusters of the University of Houston (Sabine and Carya). The authors also acknowledge Dr. Mohamad Ibrahim Cheikh (now at Tesla, Inc.) for his assistance with the initial runs of the ARW simulations and preliminary results of the current work.

References

Abarca, S., & Corbosiero, K. (2011). Secondary eyewall formation in WRF simulations of hurricanes Rita and Katrina (2005). *Geophysical Research Letters*, *38*(7). https://doi.org/10.1029/2011GL047015

Alvey, G. R., Zipser, E., & Zawislak, J. (2020). How does hurricane Edouard (2014) evolve toward symmetry before rapid intensification? A high-resolution ensemble study. *Journal of the Atmospheric Sciences*, *77*(4), 1329–1351. https://doi.org/10.1175/JAS-D-18-0355.1

Al-Yahyai, S., Charabi, Y., & Gastli, A. (2010). Review of the use of numerical weather prediction (NWP) models for wind energy assessment. *Renewable and Sustainable Energy Reviews*, *14*(9), 3192–3198. https://doi.org/10.1016/j.rser.2010.07.001

Anderson, W., & Chamecki, M. (2014). Numerical study of turbulent flow over complex Aeolian dune fields: The White Sands National Monument. *Physical Review E—Statistical, Nonlinear and Soft Matter Physics*, *89*(1), 1–14. https://doi.org/10.1103/PhysRevE.89.013005

Arguez, A., & Elsner, J. (2001). *Trends in US tropical cyclone mortality during the past century* (Vol. 32). The Florida Geographer.

Arolla, S. K., & Durbin, P. A. (2013). Modeling rotation and curvature effects within scalar eddy viscosity model framework. *International Journal of Heat and Fluid Flow*, *39*, 78–89. https://doi.org/10.1016/j.ijheatfluidflow.2012.11.006

Bender, M. A., Knutson, T. R., Tuleya, R. E., Sirutis, J. J., Vecchi, G. A., Garner, S. T., & Held, I. M. (2010). Modeled impact of anthropogenic warming on the frequency of intense Atlantic hurricanes. *Science*, *327*(5964), 454–458. https://doi.org/10.1126/science.1180568

Beven, J. L., II, & Kimberlain, T. B. (2009). *Tropical cyclone report: Hurricane Gustav (2008)*. National Hurricane Center.

Blake, E. S., Landsea, C. W., Miami, N., Gibney, E. J., & Others (2021). *The Deadliest, costliest and most intense U.S. Tropical cyclones—Tables updated*. NOAA/National Weather Service. National Centers for Environmental Prediction. Retrieved from https://www.ncdc.noaa.gov/billions/dcmi.pdf

Bou-Zeid, E., Anderson, W., Katul, G. G., & Mahrt, L. (2020). The persistent challenge of surface heterogeneity in boundary-layer meteorology: A review. *Boundary-Layer Meteorology*, *177*(2), 227–245. https://doi.org/10.1007/s10546-020-00551-8

Braun, S. A., & Tao, W. K. (2000). Sensitivity of high-resolution simulations of Hurricane Bob (1991) to planetary boundary layer parameterizations. *Monthly Weather Review*, *128*(12), 3941–3961. https://doi.org/10.1175/1520-0493(2000)129<3941:sohrso>2.0.co;2

Bryan, G. H., & Rotunno, R. (2009). The maximum intensity of tropical cyclones in axisymmetric numerical model simulations. *Monthly Weather Review*, *137*(6), 1770–1789. https://doi.org/10.1175/2008MWR2709.1

Calaf, M., Meneveau, C., & Meyers, J. (2010). Large eddy simulation study of fully developed wind-turbine array boundary layers. *Physics of Fluids*, *22*(1), 1–16. https://doi.org/10.1063/1.3291077

Cangialosi, J. P., Latto, A. S., & Berg, R. (2018). Hurricane Irma. *National Hurricane Center Tropical Cyclone Report: Hurricane Irma*, 1–111. https://www.nhc.noaa.gov/data/tcr/AL112017_Irma.pdf

Cavallo, S., Torn, R., Snyder, C., Davis, C., Wang, W., & Done, J. (2013). Evaluation of the Advanced Hurricane WRF data assimilation system for the 2009 Atlantic hurricane season. *Monthly Weather Review*, *141*(2), 523–541. https://doi.org/10.1175/mwr-d-12-00139.1

Cazalbou, J. B., Chassaing, P., Dufour, G., & Carbonneau, X. (2005). Two-equation modeling of turbulent rotating flows. *Physics of Fluids*, *17*(5), 1–14. https://doi.org/10.1063/1.1920630

Cheikh, M. I., & Momen, M. (2020). The interacting effects of storm surge intensification and sea-level rise on coastal resiliency: A high-resolution turbulence resolving case study. *Environmental Research Communications*, *2*(11), 115002. https://doi.org/10.1088/2515-7620/abc39e

Chen, C., Chen, Y., Nguyen, H. V., Setup, M., & Scheme, V. S. (2006). The spin-up process of a cyclone vortex in a tropical cyclone initialization scheme and its impact on the initial TC structure. *SOLA*, *10*(0), 93–97. https://doi.org/10.2151/sola.2014-019

Davis, C., Wang, W., Chen, S. S., Chen, Y., Corboseiro, K., Demaria, M., et al. (2008). Prediction of landfalling hurricanes with the advanced hurricane WRF model. *Monthly Weather Review*, *136*(6), 1990–2005. https://doi.org/10.1175/2007MWR2085.1

Davis, C., Wang, W., Dudhia, J., & Torn, R. (2010). Does increased horizontal resolution improve hurricane wind forecasts? *Weather and Forecasting*, *25*(6), 1826–1841. https://doi.org/10.1175/2010WAF2222423.1

Deardorff, J. W. (1972). Numerical investigation of neutral and unstable planetary boundary layers. *Journal of the Atmospheric Sciences*, *29*(1), 91–115. https://doi.org/10.1175/1520-0469(1972)029<0091:nionau>2.0.co;2

Durbin, P. (2011). Review: Adapting scalar turbulence closure models for rotation and curvature. *Journal of Fluids Engineering, Transactions of the ASME*, *133*(6), 1–8. https://doi.org/10.1115/1.4004150

Emanuel, K. (1995). Sensitivity of tropical cyclones to surface exchange coefficients and a revised steady-state model incorporating eye dynamics. *Journal of the Atmospheric Sciences*, *52*(22), 3969–3976. https://doi.org/10.1175/1520-0469(1995)052<3969:sotcts>2.0.co;2

Emanuel, K. (2005). Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, *436*, 686–688. https://doi.org/10.1038/nature03906

Emanuel, K. (2017). Will global warming make hurricane forecasting more difficult? *Bulletin of the American Meteorological Society*, *98*(3), 495–501. https://doi.org/10.1175/bams-d-16-0134.1

Emanuel, K. A. (1991). The theory of hurricanes. *Annual Review of Fluid Mechanics*, *23*(1), 179–196. https://doi.org/10.1146/annurev.fl.23.010191.001143

Fierro, A. O., Rogers, R. F., Marks, F. D., & Nolan, D. S. (2009). The Impact of horizontal grid spacing on the microphysical and kinematic structures of strong tropical cyclones simulated with the WRF-ARW model. *Monthly Weather Review*, *137*(11), 3717–3743. https://doi.org/10.1175/2009MWR2946.1

Floors, R., Peña, A., & Gryning, S. E. (2015). The effect of baroclinicity on the wind in the planetary boundary layer. *Quarterly Journal of the Royal Meteorological Society*, *141*(687), 619–630. https://doi.org/10.1002/qj.2386

Foster, R. C. (2009). Boundary-layer similarity under an axisymmetric, gradient wind vortex. *Boundary-Layer Meteorology*, *131*(3), 321–344. https://doi.org/10.1007/s10546-009-9379-1

Garratt, J. (1994). Review: The atmospheric boundary layer. *Earth-Science Reviews*, *37*(1–2), 89–134. https://doi.org/10.1016/0012-8252(94)90026-4

Ginzburg, E., Zakrison, T. L., Pust, G. D., Grant, A. A., Lu, N., Rattan, R., & Namias, N. (2018). Hurricane Irma. *Journal of Trauma and Acute Care Surgery*, *85*(3), 635–636. https://doi.org/10.1097/ta.0000000000002006

Gopalakrishnan, S., Hazelton, A., & Zhang, J. A. (2021). Improving hurricane boundary layer parameterization scheme based on observations. *Earth and Space Science*, *8*(3), 1–13. https://doi.org/10.1029/2020EA001422

Guha-Sapir, D., Vos, F., & Below, R. (2011). Annual disaster statistical review 2011 the numbers and trends.

Houze, R. A., Chen, S. S., Lee, W. C., Rogers, R. F., Moore, J. A., Strossmeister, G. J., et al. (2006). The hurricane rainbrand and intensity change experiment. *Bulletin of the American Meteorological Society*, *87*(11), 1503–1521. https://doi.org/10.1175/BAMS-87-11-1503

IPCC. (2007). *Climate change 2007* (Vol. 4). Cambridge University Press. https://doi.org/10.1256/wea.58.04

Islam, T., Srivastava, P. K., Rico-Ramirez, M. A., Dai, Q., Gupta, M., & Singh, S. K. (2015). Tracking a tropical cyclone through WRF–ARW simulation and sensitivity of model physics. *Natural Hazards*, *76*(3), 1473–1495. https://doi.org/10.1007/s11069-014-1494-8

Kepert, J. D. (2012). Choosing a boundary layer parameterization for tropical cyclone modeling. *Monthly Weather Review*, *140*(5), 1427–1445. https://doi.org/10.1175/MWR-D-11-00217.1

Knabb, R. D., Rhome, J. R., & Brown, D. P. (2006). Tropical cyclone report: Hurricane Katrina, August 23–30, 2005. National Hurricane Center. Retrieved from https://www.nhc.noaa.gov/data/tcr/AL122005_Katrina.pdf

Lemone, M. A., Angevine, W. M., Bretherton, C. S., Chen, F., Dudhia, J., Fedorovich, E., et al. (2019). 100 Years of progress in boundary layer meteorology. *Meteorological Monographs*, *59*, 9.1–9.85. https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0013.1

Mahrt, L. (1999). Stratified atmospheric boundary layers. *Boundary-Layer Meteorology*, *90*(3), 375–396. https://doi.org/10.1023/A:1001765727956

Marsooli, R., Lin, N., Emanuel, K., & Feng, K. (2019). Climate change exacerbates hurricane flood hazards along US Atlantic and Gulf Coasts in spatially varying patterns. *Nature Communications*, *10*(1), 1–9. https://doi.org/10.1038/s41467-019-11755-z

Mei, W., & Xie, S. P. (2016). Intensification of landfalling typhoons over the northwest Pacific since the late 1970s. *Nature Geoscience*, *9*(10), 753–757. https://doi.org/10.1038/ngeo2792

Mirocha, J. D., Lundquist, J. K., & Kosović, B. (2010). Implementation of a nonlinear subfilter turbulence stress model for large-eddy simulation in the advanced research WRF model. *Monthly Weather Review*, *138*(11), 4212–4228. https://doi.org/10.1175/2010MWR3286.1

Moeng, C. H., Dudhia, J., Klemp, J., & Sullivan, P. (2007). Examining two-way grid nesting for large eddy simulation of the PBL using the WRF model. *Monthly Weather Review*, *135*(6), 2295–2311. https://doi.org/10.1175/MWR3406.1

Momen, M. (2022). Baroclinicity in stable atmospheric boundary layers: Characterizing turbulence structures and collapsing wind profiles via reduced models and large-eddy simulations. *Quarterly Journal of the Royal Meteorological Society*, *148*(742), 76–96. https://doi.org/10.1002/qj.4193

Momen, M., & Bou-Zeid, E. (2016). Large eddy simulations and damped-oscillator models of the unsteady Ekman boundary layer. *Journal of the Atmospheric Sciences*, *73*(1), 25–40. https://doi.org/10.1175/JAS-D-15-0038.1

Momen, M., & Bou-Zeid, E. (2017). Mean and turbulence dynamics in unsteady Ekman boundary layers. *Journal of Fluid Mechanics*, *816*, 209–242. https://doi.org/10.1017/jfm.2017.76

Momen, M., Bou-Zeid, E., Parlange, M. B., & Giometto, M. (2018). Modulation of mean wind and turbulence in the atmospheric boundary layer by baroclinicity. *Journal of the Atmospheric Sciences*, *75*(11), 3797–3821. https://doi.org/10.1175/JAS-D-18-0159.1

Momen, M., Parlange, M. B., & Giometto, M. G. (2021). Scrambling and reorientation of classical boundary layer turbulence in hurricane winds. *Geophysical Research Letters*, *48*(7). https://doi.org/10.1029/2020GL091695

Montgomery, M. T., & Smith, R. K. (2017). Recent developments in the fluid dynamics of tropical cyclones. *Annual Review of Fluid Mechanics*, *49*(1), 541–574. https://doi.org/10.1146/annurev-fluid-010816-060022

Nasrollahi, N., AghaKouchak, A., Li, J., Gao, X., Hsu, K., & Sorooshian, S. (2012). Assessing the impacts of different WRF precipitation physics in hurricane simulations. *Weather and Forecasting*, *27*(4), 1003–1016. https://doi.org/10.1175/waf-d-10-05000.1

Pasch, R. J., Penny, A. B., & Berg, R. (2017). *Hurricane Maria*. National Hurricane Center. (Vol. 5, 16–30).
Pattanayak, S., Mohanty, U. C., Rizvi, S. R., Huang, X., & Ratna, K. N. (2006). A comparative study on performance of MM5 and WRF (ARW & NNM) models in simulation of tropical cyclone over Bay of Bengal. *Current Science*, *95*(7), 923–936.
Peduzzi, P., Chatenoux, B., Dao, H., De Bono, A., Herold, C., Kossin, J., et al. (2012). Global trends in tropical cyclone risk. *Nature Climate Change*, *2*(4), 289–294. https://doi.org/10.1038/nclimate1410
Persing, J., & Montgomery, M. (2003). Hurricane superintensity. *Journal of the Atmospheric Sciences*, *60*(19), 2349–2371. https://doi.org/10.1175/1520-0469(2003)060<2349:hs>2.0.co;2
Pita, G., Pinelli, J., Gurley, K., & Mitrani-Reiser, J. (2015). State of the art of hurricane vulnerability estimation methods: A review. *Natural Hazards Review*, *16*(2), 4014022. https://doi.org/10.1061/(asce)nh.1527-6996.0000153
Pope, S. B. (2000). *Turbulent flows* (1st edn). Cambridge University Press.
Pope, S. B., Wicker, L. J., Skamarock, W. C. W. C., Soloviev, A. V., Lukas, R., Donelan, M. A., et al. (2010). Tropical cyclone report: Hurricane Gustav (2008). *Monthly Weather Review*, *136*(1), 91–115. https://doi.org/10.1175/2007MWR2085.1
Ramamurthy, P., Pardyjak, E. R., & Klewicki, J. C. (2007). Observations of the effects of atmospheric stability on turbulence statistics deep within an Urban Street Canyon. *Journal of Applied Meteorology and Climatology*, *46*(12), 2074–2085. https://doi.org/10.1175/2007JAMC1296.1
Romdhani, O., Zhang, J., & Momen, M. (2022). Characterizing the impacts of turbulence closures on real hurricane forecasts: A comprehensive joint assessment of grid resolution, horizontal turbulence models, and horizontal mixing length [Dataset]. Harvard Dataverse, V1. https://doi.org/10.7910/DVN/D6P6VC .
Rotunno, R., & Bryan, G. H. (2012). Effects of parameterized diffusion on simulated hurricanes. *Journal of the Atmospheric Sciences*, *69*(7), 2284–2299. https://doi.org/10.1175/JAS-D-11-0204.1
Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Zhiquan, L., Berner, J., et al. (2019). A description of the advanced research WRF model version 4. NCAR Technical Note NCAR/TN-475+STR.
Stewart, S. R., & Berg, R. (2019). *Hurricane Florence*. National Hurricane Center Tropical Cyclone, (Vol. 1).
Stull, R. (1988). *An introduction to boundary layer meteorology* (1st edn). Kluwer Academic Publishers. http://dx.doi.org/10.1007/978-94-009-3027-8
Tritton, D. J. (1992). Stabilization and destabilization of turbulent shear flow in a rotating fluid. *Journal of Fluid Mechanics*, *241*(21), 503–523. https://doi.org/10.1017/S0022112092002131
van de Wiel, B. J. H., Moene, A. F., & Jonker, H. J. J. (2012). The cessation of continuous turbulence as precursor of the very stable nocturnal boundary layer. *Journal of the Atmospheric Sciences*, *69*(11), 3097–3115. https://doi.org/10.1175/JAS-D-12-064.1
Wei, W., Bruyère, C., Duda, M., Dudhia, J., Gill, D., Kavulich, M., et al. (2019). WRF ARW version 4 modeling system user's guide.
Weinkle, J., Landsea, C., Collins, D., Musulin, R., Crompton, R. P., Klotzbach, P. J., & Pielke, R. (2018). Normalized hurricane damage in the continental United States 1900–2017. *Nature Sustainability*, *1*(12), 808–813. https://doi.org/10.1038/s41893-018-0165-2
Worsnop, R. P., Bryan, G. H., Lundquist, J. K., & Zhang, J. A. (2017). Using large-eddy simulations to define spectral and coherence characteristics of the hurricane boundary layer for wind-energy applications. *Boundary-Layer Meteorology*, *165*(1), 55–86. https://doi.org/10.1007/s10546-017-0266-x
Xue, M., Schleif, J., Kong, F., Thomas, K. W., Wang, Y., & Zhu, K. (2013). Track and intensity forecasting of hurricanes: Impact of convection-permitting resolution and global ensemble Kalman filter analysis on 2010 Atlantic season forecasts. *Weather and Forecasting*, *28*(6), 1366–1384. https://doi.org/10.1175/WAF-D-12-00063.1
Zhang, J. A. (2010). Spectral characteristics of turbulence in the hurricane boundary layer over the ocean between the outer rain bands. *Quarterly Journal of the Royal Meteorological Society*, *136*(649), 918–926. https://doi.org/10.1002/qj.610
Zhang, J. A., & Marks, F. D. (2015). Effects of horizontal diffusion on tropical cyclone intensity change and structure in idealized three-dimensional numerical simulations. *Monthly Weather Review*, *143*(10), 3981–3995. https://doi.org/10.1175/MWR-D-14-00341.1
Zhang, J. A., Marks, F. D., Sippel, J. A., Rogers, R. F., Zhang, X., Gopalakrishnan, S. G., et al. (2018). Evaluating the impact of improvement in the horizontal diffusion parameterization on hurricane prediction in the operational Hurricane Weather Research and Forecast (HWRF) Model. *Weather and Forecasting*, *33*(1), 317–329. https://doi.org/10.1175/WAF-D-17-0097.1
Zhang, J. A., & Montgomery, M. T. (2012). Observational estimates of the horizontal eddy diffusivity and mixing length in the low-level region of intense hurricanes. *Journal of the Atmospheric Sciences*, *69*(4), 1306–1316. https://doi.org/10.1175/JAS-D-11-0180.1
Zhang, J. A., Nolan, D. S., Rogers, R. F., & Tallapragada, V. (2015). Evaluating the impact of improvements in the boundary layer parameterization on hurricane intensity and structure forecasts in HWRF. *Monthly Weather Review*, *143*(8), 3136–3155. https://doi.org/10.1175/MWR-D-14-00339.1
Zhang, J. A., Rogers, R. F., & Tallapragada, V. (2017). Impact of parameterized boundary layer structure on tropical cyclone rapid intensification forecasts in HWRF. *Monthly Weather Review*, *145*(4), 1413–1426. https://doi.org/10.1175/MWR-D-16-0129.1
Zhu, P., Hazelton, A., Zhang, Z., Marks, F., & Tallapragada, V. (2021). The role of eyewall turbulent transport in the pathway to intensification of tropical cyclones. *Journal of Geophysical Research: Atmospheres*, *126*(17), e2021JD034983.
Zhu, P., Menelaou, K., & Zhu, Z. (2014). Impact of subgrid-scale vertical turbulent mixing on eyewall asymmetric structures and mesovortices of hurricanes. *Quarterly Journal of the Royal Meteorological Society*, *140*(679), 416–438. https://doi.org/10.1002/qj.2147

References From the Supporting Information

Beven, II., J. L., & Kimberlain, T. B. (2009). Tropical Cyclone Report: Hurricane Gustav (2008). *Tropical Cyclone Report Hurricane Gustav*, *1–38*.
Black, P. G., D'Asaro, E. A., Drennan, W. M., French, J. R., Niiler, P. P., Sanford, T. B., et al. (2007). Air-sea exchange in hurricanes. *Bulletin of the American Meteorological Society*, *88*(3), 357–374. https://doi.org/10.1175/bams-88-3-357
Bryan, G. H., Worsnop, R. P., Lundquist, J. K., & Zhang, J. A. (2017). A simple method for simulating wind profiles in the boundary layer of tropical cyclones. *Boundary-Layer Meteorology*, *162*(3), 475–502. https://doi.org/10.1007/s10546-016-0207-0
Cangialosi, J. P., Latto, A. S., & Berg, R. (2018). Hurricane Irma. National Hurricane Center Tropical Cyclone Report: *Hurricane Irma*, *1–111*.
Cavallo, S., Torn, R., Snyder, C., Davis, C., Wang, W., & Done, J. (2013). Evaluation of the advanced hurricane WRF data assimilation system for the 2009 Atlantic hurricane season. *Monthly Weather Review*, *141*(2), 523–541, https://doi.org/10.1175/mwr-d-12-00139.1
Davis, C., Wang, W., Chen, S. S., Chen, Y., Corboseiro, K., Demaria, M., et al. (2008). Prediction of Landfalling Hurricanes with the Advanced Hurricane WRF Model. *Monthly Weather Review*, *136*, (6), 1990–2005. https://doi.org/10.1175/2007MWR2085.1
Donelan, M. A., Haus, B. K., Reul, N., Plant, W. J., Stiassnie, M., Graber, H. C., et al. (2004). On the limiting aerodynamic roughness of the ocean in very strong winds. *Geophysical Research Letters*, *31*(18), 1-5. https://doi.org/10.1029/2004GL019460

Dudhia, J. (1989). Numerical study of convection observed during the winter monsoon experiment using a mesoscale two dimensional model. *Journal of the Atmospheric Sciences*, *46*(20), 3077. https://doi.org/10.1175/1520-0469(1989)046<3077:nsocod>2.0.co;2

Garratt, J. R. (1994). Review: the atmospheric boundary layer. *Earth Science Reviews*, *37*(1–2), 89–134. https://doi.org/10.1016/0012-8252(94)90026-4

Hong, S. (2010). A new stable boundary-layer mixing scheme and its impact on the simulated East Asian summer monsoon. *Quarterly Journal of the Royal Meteorological Society*, *136*(651), 1481-1496, https://doi.org/10.1002/qj.665

Hong, S. Y., Dudhia, J., & Chen, S. H. (2004). A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Monthly Weather Review*, *132*(1), 103-120. https://doi.org/10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2

Hong, S.-Y., Lim, K.-S., Kim, J.-H., Ock, J.-, Lim, J., & Dudhia, J. WRF Single-Moment 6-Class Microphysics Scheme (WSM6).

Hong, S. Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*(9), 2318-2341. https://doi.org/10.1175/MWR3199.1

Hu, X.-M., Nielsen-Gammon, J. W., & Zhang, F. (2010). Evaluation of three planetary boundary layer schemes in the WRF model. *Journal of Applied Meteorology and Climatology*, *49*(9), 1831–1844. https://doi.org/10.1175/2010JAMC2432.1

Jarosz, E., Mitchell, D. A., Wang, D. W., & Teague, W. J. (2007). Major Tropical Cyclone, *557*, 2005–2007.

Knabb, R. D., Rhome, J. R., & Brown, D. P. (2006). *Tropical cyclone report: Hurricane Katrina, August 23–30, 2005*. Tropical Cyclone Report Hurricane Katrina, 1–43.

Lacis, A., & Hansen, J. (1974). A parameterization for the absorption of solar radiation in the Earth's atmosphere. *Journal of Atmospheric Sciences*, *31*(1), 118-133, https://doi.org/10.1175/1520-0469(1974)031<0118:apftao>2.0.co;2

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research Atmospheres*, *102*(14), 16663–16682. https://doi.org/10.1029/97jd00237

Momen, M., Parlange, M. B., & Giometto, M. G. (2021). Scrambling and reorientation of classical atmospheric boundary layer turbulence in hurricane winds. *Geophysical Research Letters*, *48*(7). https://doi.org/10.1029/2020GL091695

Mooney, P. A., Mulligan, F. J., Bruyère, C. L., Parker, C. L., & Gill, D. O. (2019). Investigating the performance of coupled WRF-ROMS simulations of Hurricane Irene (2011) in a regional climate modeling framework. *Atmospheric Research*, *215*, 57-74. https://doi.org/10.1016/j.atmosres.2018.08.017

NCAR. (2019). User's guides for the advanced research WRF (ARW) modeling system, Version 4, 456.

NCEP, & GDAS. (2015). *FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids*, Research Data Archive at the National Center for Atmospheric Research; Computational and Information Systems Laboratory.

NCEP National Centers for Environmental Prediction/National Weather Service/NOAA/US Department of Commerce. (2000). *NCEP FNL operational model global tropospheric analyses, continuing from July 1999*, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory.

Nolan, D. S., Stern, D. P., & Zhang, J. A. (2009). Evaluation of planetary boundary layer parameterizations in tropical cyclones by comparison of in situ observations and high-resolution simulations of Hurricane Isabel (2003). Part II: Inner-core boundary layer and eyewall structure. *Monthly Weather Review*, *137*(11), 3675-3698. https://doi.org/10.1175/2009MWR2786.1

Pasch, R. J., Penny, A. B., & Berg, R. (2019). Hurricane Maria. National Hurricane Center Tropical Cyclone Report: Hurricane Maria, 1–69. Retrieved from https://www.nhc.noaa.gov/data/tcr/AL152017_Maria.pdf

Powell, M. D., Vickery, P. J., & Reinhold, T. A. (2003). Reduced drag coefficient for high wind speeds in tropical cyclones. *Nature*, *422*(6929), 279–283. https://doi.org/10.1038/nature01481

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Zhiquan, L., Berner, J., et al. (2019). A description of the advanced research WRF model version 4. NCAR technical note NCAR/TN-475+STR. Retrieved from http://library.ucar.edu/research/publish-technote

Soloviev, A. V., Lukas, R., Donelan, M. A., Haus, B. K., & Ginis, I. (2014). The air-sea interface and surface stress under tropical cyclones. *Scientific Reports*, *4*, 1-6. https://doi.org/10.1038/srep05306

Wei, W., Bruyère, C., Duda, M., Dudhia, J., Gill, D., Kavulich, M., et al. (2019). WRF ARW version 4 modeling system user's guide.

Wicker, L. J., & Skamarock, W. C. (2002). Time-splitting methods for elastic models using forward time schemes. *Monthly Weather Review*, *130*(8), 2088–2097. https://doi.org/10.1175/1520-0493(2002)130⟨2088:TSMFEM⟩2.0.CO;2

Zhang, J. A. (2010). Spectral characteristics of turbulence in the hurricane boundary layer over the ocean between the outer rain bands. *Quarterly Journal of the Royal Meteorological Society*, *136*(649), 918-926. https://doi.org/10.1002/qj.610