Evaluation of Atmosphere and Ocean Initial Condition Uncertainty and
Stochastic Exchange Coefficients on Ensemble Tropical Cyclone Intensity
Forecasts

Ryan D. Torn *

Department of Atmospheric and Environmental Sciences, University at Albany, State University
of New York

*Corresponding author address: Ryan Torn, University at Albany, ES 351, 1400 Washington Avenue, Albany, NY 12222.
E-mail: rtorn@albany.edu
ABSTRACT

Tropical Cyclone (TC) intensity forecasts are impacted by errors in atmosphere and ocean initial conditions and the model formulation, which motivates using an ensemble approach. This study evaluates the impact of uncertainty in atmospheric and oceanic initial conditions, as well as stochastic representations of the drag ($C_d$) and enthalpy ($C_k$) exchange coefficients on ensemble Advanced Hurricane WRF (AHW) TC intensity forecasts via a series of experiments that consist of 35 initialization times over 20 Atlantic TCs from 2008-2011. Each ensemble experiment consists of different combinations of having deterministic, or ensemble atmospheric and/or oceanic initial conditions, as well as fixed, or stochastic representations of $C_d$ or $C_k$. Among those experiments with a single uncertainty source, atmospheric uncertainty produces the largest standard deviation in TC intensity. While ocean uncertainty leads to continuous growth in ensemble standard deviation, the ensemble standard deviation in the experiments with $C_d$ and $C_k$ uncertainty levels off by 48 h. Combining atmospheric and oceanic uncertainty leads to larger intensity standard deviation than atmosphere or ocean uncertainty alone and preferentially adds variability outside of the TC core. By contrast, combining $C_d$ or $C_k$ uncertainty with any other source leads to negligible increases in standard deviation. All of the ensemble experiments are deficient in ensemble standard deviation; however, the experiments with combinations of uncertainty sources generally have an ensemble standard deviation closer to the ensemble-mean errors.
1. Introduction

Despite continuing advances in numerical weather prediction (NWP), tropical cyclone (TC) intensity change remains a significant challenge. TC intensity changes can result from a variety of processes and scales; therefore, the predictability timescale for TC intensity remains elusive. Moist convection, which is thought to have a predictability timescale of a few hours, drives some of the internal dynamics present in mature systems, which appear to limit TC intensity predictability in idealized scenarios to 2 days or less (e.g., Hakim 2013; Brown and Hakim 2013); however, Judt et al. (2015) suggest that Hurricane Earl (2010) intensity forecasts are relatively insensitive to high-wavenumber errors. In addition, large-scale environmental factors, such as sea-surface temperature, vertical wind shear, which tend to have slower error growth, strongly modulate TC intensity, which could provide longer-range predictability (e.g., Emanuel et al. 2004; Wang and Wu 2004; Komaromi and Majumdar 2014). Moreover, 3D idealized simulations suggest that TC intensity predictability is strongly modulated by the environmental forcing, such that moderate vertical wind shear and sea surface temperature can lead increase or decrease the predictability timescale (e.g., Zhang and Tao 2013; Tao and Zhang 2014). Individual case studies suggest that wind and moisture asymmetries can lead to rapid growth of forecast errors associated with specific aspects of the TC, thus limiting intensity predictability (e.g., Sippel and Zhang 2008; Zhang and Sippel 2009; Munsell et al. 2013; Rios-Berrios et al. 2016a,b).

Given the inherent uncertainties in both model initial conditions, and model formulation (i.e., parameterizations), which can hence translate into TC intensity forecast errors, it is prudent to adopt a probabilistic forecast approach. Ideally, these probabilistic forecasts would be based on ensemble forecasts, where multiple realizations of a particular forecast are generated, either via a single or multi-model approach. Ideally, these ensemble prediction systems should take into
account both initial condition errors and errors in the formulation of the model that are directly applicable to TC dynamics. While most real-world TCs are not axisymmetric or steady-state, Potential Intensity (PI) theory (Emanuel 1986) can provide a starting point for determining where to introduce uncertainties into a TC ensemble prediction system. In this framework, TC intensity is primarily a function of the drag ($C_d$) and enthalpy ($C_k$) exchange coefficients, thermal disequilibrium between the atmosphere and ocean, and upper-tropospheric outflow temperature. As a consequence, this paradigm suggests that a TC ensemble prediction system should account for uncertainty in the atmosphere initial conditions (both associated with the vortex and environment), ocean initial conditions (primarily via sea-surface temperature), and in the exchange coefficients, which are not particularly well known at TC-force wind speeds (e.g., Drennan et al. 2007; French et al. 2007; Zhang et al. 2009; Andreas et al. 2012; Bell et al. 2012).

Although TC intensity forecasting could benefit from ensemble prediction, most of the work on TC ensemble forecasting has focused around probabilistic prediction of TC track (e.g., Majumdar and Finocchio 2010; Reynolds et al. 2011a; Dupont et al. 2011; Hamill et al. 2011; Lang et al. 2012), or TC genesis (e.g., Elsberry et al. 2010; Snyder et al. 2010; Majumdar and Torn 2014; Komaromi and Majumdar 2015). Most of these studies employ relatively coarse global ensemble prediction systems that do not have sufficient resolution to simulate internal processes within the TC. In the deep tropics, initial condition errors tend to grow relatively slowly or not at all (e.g., McLay et al. 2008; Yamaguchi and Majumdar 2010); therefore, skillful ensemble large-scale and TC track forecasts require model error representations, either by perturbing model parameters (e.g., Reynolds et al. 2011b), or adding stochastic tendencies to the equations (e.g., Lang et al. 2012). In addition, Kunii and Miyoshi (2012) obtained improved TC track errors when sea-surface temperature perturbations were added to regional model ensemble forecasts of Typhoons Sinlaku and Jangmi (2008), which they attribute to having physically meaningful uncertainty.
While there are numerous studies looking at TC track, relatively fewer studies have investigated how to generate ensemble forecasts that address TC intensity. A majority of these studies have focused mainly on the role of atmospheric initial conditions, often taken from mesoscale data assimilation systems (e.g., Zhang and Sippel 2009; Poterjoy and Zhang 2014; Wu et al. 2014; Qian et al. 2013). By contrast, there have been fewer studies with uncertainty in the model formulation, with most of those dedicated to estimating uncertain model parameters via data assimilation for a single case study. Green and Zhang (2014) found that multiplying the drag drag coefficient ($C_d$) by a single number produced greater variability in forecasts of Hurricane Katrina’s (2005) intensity compared to multiplying the enthalpy exchange coefficient ($C_k$) because $C_d$ has a large impact on structure. Moreover, Godinez et al. (2012) obtained lower forecast errors of Hurricane Guillermo (1997) when they estimated various surface-based parameters via data assimilation. Zhang et al. (2014) generated ensemble forecasts for many initialization times using initial conditions taken from a global ensemble and employing a stochastic trigger for the cumulus scheme. While the ensemble-mean intensity track and intensity errors were smaller than the deterministic forecast, the ensemble standard deviation was smaller than the ensemble-mean errors.

The goal of this paper is to quantify how current estimates of the uncertainty in atmosphere, ocean, and the exchange coefficients lead to variability in TC intensity forecasts over a variety of cases. This is accomplished by generating ensemble forecasts, which are characterized by various combinations of uncertainties (atmosphere, ocean, or exchange coefficients), over a variety of cases and comparing the resulting ensemble forecast experiments against one another. In addition, these experiments test the hypothesis that combining various initial condition and model uncertainty sources will lead to the most skillful TC ensemble intensity forecasts. The cases are taken from the Atlantic Basin during 2008-2011 and are characterized by TCs with variety of locations, intensities, and points in the TC lifecycle. The deterministic forecasts used here were originally
produced as part of the 2012 Hurricane Forecast Improvement Project (HFIP; Gall et al. 2013)

Stream 1.5 Evaluation. Although other studies have explored the role of atmosphere, ocean, and exchange coefficient variability on TC intensity forecasts for a small number of cases, this work represents the first time all three uncertainty sources are compared in a single framework over many cases.

The paper proceeds as follows. Section 2 describes the model and design of the ensemble experiments. Section 3 compares the ensemble standard deviation and error in various intensity metrics for the ensemble experiments. In addition, section 4 analyzes the processes that lead to intensity variability, while section 5 describes two sensitivity experiments. A summary and conclusions are given in section 6.

2. Experiment Description

The growth of TC intensity forecast variability is evaluated using a set of Atlantic Basin ensemble TC forecasts selected from the 2008-2011 seasons. Table 1 provides the list of cases and initialization times used in this study. The 35 initialization times represent a fairly broad cross section of TCs in the Atlantic Basin, including 16 forecasts where rapid intensification (30 knots 24 h$^{-1}$; Kaplan and DeMaria 2003) takes place, 13 major hurricanes (maximum wind speed $> 95$ knots), as well as 2 storms that never reached hurricane status (maximum wind speed $< 64$ knots). For the longer-lasting TCs, multiple initialization times are employed to capture different points in the TC’s lifecycle. In those instances, the initialization times are separated by at least 2 d, so that the forecast errors at various initialization times should not be correlated. All forecasts are initialized at 0000 UTC, so that the observation density, which partially determines analysis quality, is not significantly different from one case to another.
All forecasts were generated using version 3.3.1 of the Advanced Hurricane Research version (AHW) of the Weather Research and Forecasting (WRF) model (Skamarock et al. 2005), including the modifications for TCs described in Davis et al. (2010). The model is integrated for 120 h over a 36 km grid-spacing domain that spans the Atlantic Basin and upstream areas (see Fig. 1 for domain extent), as well as 12 and 4 km TC-following two-way nests of size 1600 km × 1600 km and 800 km × 800 km, respectively, over 36 vertical levels with a top at 20 hPa. The physical parameterizations include WRF 6-class microphysics scheme (Hong et al. 2004), Rapid Radiative Transfer Model (RRTMG) longwave and shortwave radiation (Mlawer et al. 1997; Iacono et al. 2008), Yonsei University (YSU) boundary-layer scheme (Hong et al. 2006), NOAH land surface model (Ek et al. 2003), modified Tiedtke cumulus scheme (Tiedtke 1989; Zhang et al. 2011) on the 36 and 12 km domains (the 4 km domain did not employ a cumulus parameterization), and positive-definite moisture advection (Skamarock and Weisman 2009). In addition, this version of the model uses the isftcflx option 1 from version 3.4.1\(^1\), which is a blend of the Powell et al. (2003) and Donelan et al. (2004) formulations. The interested reader is directed to Green and Zhang (2013) for an extensive discussion on how the exchange coefficients are formulated in the model. Finally, ocean mixing is parameterized using the Pollard et al. (1973) single-column mixed layer model, which cools the ocean based on the surface wind stress and mixed-layer depth. The interested reader is directed to Davis et al. (2008) for a complete description of how this parameterization is implemented in AHW.

Lateral Boundary conditions are taken from the corresponding-time National Centers for Environmental Prediction Global Forecast System (GFS) forecast. For simplicity, every member uses the same set of lateral boundary conditions; the lack of variance coming from the lateral bound-

\(^1\)This version was originally developed using this version and released in version 3.4.
aries likely does not impact the results given the lateral boundaries are well-removed from all TCs in this study because of the large 36 km domain.

Each ensemble forecast experiment consists of a combination of atmosphere, ocean, and model error sources, which are detailed in Table 2\(^2\). In each experiment, the variability is obtained using an ensemble of either initial conditions or model parameters, while all other components are deterministic. For example, the experiment denoted ATM is comprised of a 30-member ensemble where each member is initialized with different atmosphere initial conditions, but the same ocean initial condition and unperturbed drag and enthalpy exchange coefficients (elaborated below). By contrast, the experiment denoted OC is comprised of a 30-member ensemble where each member is initialized with different ocean initial conditions, but the same atmosphere initial condition and unperturbed drag and enthalpy exchange coefficient. Finally, other experiments will use combinations of variability sources, such as ATM+OC, which is a 30-member ensemble where each member is initialized with a different atmosphere and ocean initial condition, but unperturbed drag and enthalpy exchange coefficients. These experiments can be subsequently compared to one another to evaluate the impact of different sources of variability on ensemble TC intensity forecasts. For example, the difference between ATM and ATM+OC describes the impact of adding ocean initial condition uncertainty to atmospheric initial-condition uncertainty. Deterministic and ensemble initial conditions and model formulations are obtained using the following procedures.

\(\text{a. Atmospheric Initial Conditions}\)

Atmospheric ensemble initial conditions are obtained from a cycling ensemble Kalman filter system, similar to the one used in Cavallo et al. (2013); Torn and Davis (2012). A short summary

\(^2\)Each experiment consists of 1050 individual AHW model simulations (35 cases × 30 members)
of the system is provide below; the interested reader is directed to Torn (2010); Cavallo et al. (2013); Torn and Davis (2012) for additional detail and justification for the choices made.

Observations are assimilated every 6 h from the sources listed in Table 3 using the Data Assimilation Research Testbed (DART; Anderson et al. 2009), which is an implementation of the Ensemble Adjustment Kalman Filter (Anderson 2001), on the 36-km domain and 12-km TC-following nests. The location and motion of the nests in the data assimilation is described in Torn and Davis (2012). At each analysis time, the data assimilation system generates an ensemble of 96 equally-likely analyses, where the ensemble perturbations are consistent with the expected analysis errors. It is worth pointing out that the perturbations include uncertainty in both the TC position, intensity, structure, and large-scale environment. For brevity, four different observation cycling periods are used in this study, one for each year (given in Table 4). The initial ensemble for each cycling period is produced from the fixed-covariance perturbation (FCP) technique described in Torn et al. (2006). Specifically, the ensemble perturbations are obtained by drawing random perturbations from the NCEP error covariances contained in the WRF VAR system (Barker et al. 2004) and adding those perturbations to the 6-h GFS forecast valid at the cycling initialization time. For each period, there is at least 3 d between the beginning of the cycling period and the genesis time of each storm; this choice ensures that the ensemble has little memory of the initial perturbations by the time of genesis.

All ensemble data assimilation systems are characterized by sampling errors due to using a finite-sized ensemble to compute the forecast error covariances. In order to address this issue, the covariances are localized using eqn. 4.10 of Gaspari and Cohn (1999) where the localization factor reduces to zero 2000 km in the horizontal and 2 scale heights in the vertical from the observation location. For state variables that have more than 1600 observations within the ellipsoid defined by these distances, the horizontal and vertical covariance length scales are reduced using the method
described in Torn (2010). In addition, the ensemble perturbations are inflated at each analysis time using the adaptive technique of Anderson (2009), where the inflation factor is damped by 10% at each assimilation time and the inflation standard deviation is fixed at 0.6.

For those experiments that use ensemble initial conditions for the atmosphere, the first 30 members of the analysis ensemble are employed, whereas a single member of the analysis ensemble is selected for those experiments with identical (i.e., deterministic) atmospheric initial conditions. This ensemble member is selected by finding the analysis ensemble member that minimizes the following cost function that measures how “close” each ensemble member’s 0-h TC position and minimum SLP is to the ensemble-mean value,

$$J(n) = \left( \frac{\text{Lat}_n - \overline{\text{Lat}}}{\sigma_{\text{Lat}}} \right)^2 + \left( \frac{\text{Lon}_n - \overline{\text{Lon}}}{\sigma_{\text{Lon}}} \right)^2 + 2 \left( \frac{\text{MSLP}_n - \overline{\text{MSLP}}}{\sigma_{\text{MSLP}}} \right)^2$$

(1)

where, \(\text{Lat}_n\), \(\text{Lon}_n\), and \(\text{MSLP}_n\) are ensemble member \(n\)’s estimates of the TC latitude, longitude and minimum SLP, respectively, \(\sigma\) is a climatological normalizing constant (0.15° for latitude and longitude, 4 hPa for minimum SLP) and overbar indicates the ensemble mean of that quantity. An alternative to this approach is to use the ensemble-mean analysis; however, variability in the TC position within the ensemble can reduce the amplitude of the TC mass and wind fields, which subsequently leads to a weak bias in the first 24 h of the forecast relative to using a single ensemble member.

b. Oceanic Initial Conditions

Deterministic sea-surface temperature initial conditions (SST) are obtained from the 1/12° daily real-time global sea-surface temperature analysis generated by the NCEP Marine Modeling and Analysis Branch (NCEP/MMAB; Gemmill et al. 2007), while the mixed-layer depth (MLD) is de-
terminated from operational Hybrid Coordinate Ocean Model (HYCOM; Bleck 2002) temperature
analysis and computed using the method outlined in Davis et al. (2010).

As of this writing, a 30-member ensemble of ocean initial conditions is not available for this
period; therefore, it is necessary to develop an alternative strategy of generating ocean initial
condition perturbations that can be added to the deterministic initial condition. One potential
candidate for creating an ensemble of initial conditions is to add spatially-uncorrelated white noise
to each horizontal grid point that is consistent with SST analysis uncertainty. While this approach
is relatively simple, it ignores the spatial correlations in SST analysis errors. As an alternative, one
may employ an approach similar to Kunii and Miyoshi (2012) and generate ocean perturbations
by randomly choosing ocean analyses from climatology and scaling the perturbations to match
the analysis uncertainty. While this approach is admittedly primitive, it does allow for spatial
correlations in ocean analysis errors that are at least consistent with climatology.

In this study, ocean ensemble initial conditions are produced using the following approach. For
each case, 30 ocean initial conditions from 2006-2011 are randomly selected from the same month
from which the forecast is initialized (i.e., for a forecast initialized in August, 30 analyses would
be randomly chosen from all August dates from 2006-2011) and the mean value is removed.
The resulting SST and MLD perturbations are then scaled by a spatially-varying factor, so that the
standard deviation of the SST is 0.5 K at every horizontal grid point; the same scaling factor is used
on the corresponding-location MLD perturbations. The SST standard deviation is consistent with
the RMS difference between the SST analyses and buoy observations in the tropics (e.g., Gemmill
et al. 2007; Reynolds and Chelton 2010, Robert Grumbine, MMAB, Personal Communication),
but is larger than what was used in Kunii and Miyoshi (2012) (0.2 K). Fig. 1 shows a randomly-
selected SST and MLD perturbation for September. The SST perturbations have reasonable spatial
coherence, with larger-scale perturbations in the middle of the basin, but small-scale perturbations
near the Gulf Stream, which is characterized by mesoscale eddies. Ensemble initial conditions for each member are then obtained by adding the resulting perturbations to the deterministic initial condition.

c. Parameter Variability

MPI theory indicates that the maximum intensity is proportional to the drag coefficient ($C_d$) and enthalpy exchange coefficients ($C_k$); however, these two quantities are not well known, particularly at wind speeds characteristic of TCs. As a consequence, these quantities probably should not be treated as a deterministic value for a given wind speed. Instead, these quantities should include a stochastic element that takes into account the uncertainty in these values. A perturbed version of $C_d$ or $C_k$ is obtained via:

$$X' = (1 + \varepsilon)X,$$

where $X$ is either $C_d$ or $C_k$ and $\varepsilon$ is a random number drawn from some distribution, which could be a function of space and time; however, for this study it will only a function of space. This approach differs from Green and Zhang (2014), who employed a spatially-constant $C_d'$ and $C_k'$ for each ensemble member. At each horizontal location ($x,y$), $\varepsilon$ is determined by sampling from a Gaussian distribution with mean zero and standard deviation that is consistent with the measurement uncertainty ($0.5 \times 10^{-3}$ for $C_d$, $0.2 \times 10^{-3}$ for $C_k$; e.g., Andreas et al. 2012; Bell et al. 2012; Drennan et al. 2007; French et al. 2007; Zhang et al. 2009).

In this version of AHW, both $C_k$ and $C_d$ are diagnosed quantities and are not used directly in the calculation of momentum and enthalpy exchange\(^3\). While it is possible to multiply the heat and moisture fluxes by a stochastic parameter, which is equivalent to perturbing $C_k$, the same cannot

---

\(^3\)See Green and Zhang (2013) for exact formulation
be done for $C_d$. As a consequence, $C_d$ variability is introduced via eqn. (2) then transformed into a perturbed roughness length ($z_o'$) via

$$z_o' = z_{ref} e^{-\kappa(C_d')^{-\frac{1}{2}} + \psi_m}$$

(3)

where $z_{ref}$ is the reference height (here 10 m), $\kappa$ is the von Karman constant, and $\psi_m$ is the Monin-Obukhov stability correction function for momentum. This equation is obtained by solving for $z_o$ in eqn. A9 of Green and Zhang (2013). In order to maintain consistency in $C_d'$ for all three domains, $C_d'$ on the 12 and 4 km domains is interpolated from the 36 km domain values.

3. Results

Prior to comparing the various ensemble experiments against each other, it is worthwhile to document the skill of the deterministic AHW forecast (i.e., forecast with deterministic atmosphere, ocean, and exchange coefficient formulations) against the corresponding-time forecasts from other forecasts. For TC position, the AHW track forecasts are compared with the National Hurricane Center Official Forecast (OFCL) and NOAA Global Forecast System (GFS; denoted AVNO). For maximum wind speed, the AHW forecasts are compared with NHC Official forecasts, as well as two statistical models, the Logistical Growth Equation Model (LGEM; DeMaria 2009) and the Decay-SHIFOR Model (Knaff et al. 2003); the latter is used by NHC as a benchmark skill assessment from one season to another (e.g., Beven II and Blake 2015).

For both track and intensity, the AHW deterministic forecasts are characterized by forecast errors that are comparable to other NWP systems (Fig. 2). With the exception of 120 h, the AHW position errors are larger than GFS at all lead times; however, the differences are not statistically significant (determined via a bootstrap resampling method similar to Davis et al. (2010)). By contrast, the AHW position forecasts are characterized by forecast errors that are roughly one day behind the
NHC Official (i.e., 72 h NHC position errors are similar to 48 h AHW position errors), which is statistically significant at the 95% level.

AHW intensity forecasts show similar intensity errors to other operational guidance, with the exception at 0 h (Fig. 2b). The larger error at 0 h is related to AHW not being initialized with the advisory intensity (the intensity is updated through data assimilation; see Torn (2010)) and the positive maximum wind speed bias at this time. This bias is related to noise introduced by the ensemble-derived covariances in the data assimilation system and the nature of the maximum wind speed metric (see Torn (2010) for a detailed explanation of this). Beyond that time, the AHW forecast errors are similar to or smaller than LGEM and OFCL forecasts until 72 h, then become larger, while the bias is roughly zero. Moreover, the AHW forecasts have skill relative to Decay-SHIFOR forecasts, with the exception of 0 and 120 h. Nevertheless, these results suggest that AHW errors are at least comparable to operational systems and thus have sufficient skill to carry out this type of study.

a. Standard Deviation Comparisons

With the deterministic performance established, the remainder of the paper focuses on comparing the ensemble standard deviation and ensemble-mean forecast errors for various metrics related to TC intensity and structure that result from each uncertainty source. Given that TC intensity errors are increasing with forecast lead-time, the ensemble standard deviation in various metrics should also increase with it. Moreover, one would also expect that experiments with a combination of various uncertainty sources would produce a higher standard deviation than experiments with fewer uncertainty sources.

Fig. 3a shows the ensemble standard deviation in maximum wind speed as a function of lead time averaged over all cases. Not surprisingly, ATM has the largest standard deviation among the
single-uncertainty-source experiments (OC, C_d, C_k). This result is expected because ATM has uncertainty in the TC structure at 0 h, while all members of OC, C_d and C_k are initialized with the same atmospheric state, thus intensity variability can only result from how uncertainty associated with various processes feedback onto TC intensity change. Despite this, ATM is characterized by a small decrease in standard deviation during the first 3 h due to the noise introduced by data assimilation (described in Torn (2010)) and remains relatively flat over the next 24 h, before increasing at roughly 2 kts d^{-1} between 24-48 h. For C_d, the standard deviation is 3.5 kts at 24 h, which is similar to OC, but larger than C_k. Whereas C_d and C_k level off at roughly 4 and 5 kts, respectively, starting around 48 h, the standard deviation for OC continues to grow at a rate of roughly 1.7 kts d^{-1}, so that by 96 h the OC standard deviation is within 1 knot of ATM. The shape of the standard deviation curves for C_k and C_d (growth during first 48 h, leveling off thereafter) suggests that perturbing these parameters has a bounded ability to introduce variability in the maximum wind speed; once this value is reached, there is no additional error growth. It is unclear why the standard deviation levels off after 48 h, though it is at least consistent with the intrinsic predictability timescales identified by Hakim (2013) and Brown and Hakim (2013). By contrast, the continuous increase in standard deviation within OC suggests that initial condition uncertainty in the ocean is associated with sustained error growth in maximum wind speed.

Experiments with multiple sources of uncertainty suggest that combining ATM uncertainty with other uncertainty sources leads to larger standard deviation in maximum wind speed compared to ATM alone. Adding C_d variability to ATM (ATM+C_d) is associated with on average a 0.2-0.4 kts increase in standard deviation at most lead times (statistically significant after 48 h). Adding ocean uncertainty to ATM (ATM+OC) does not produce a statistically significant difference in standard deviation until 48 h; however, the differences between ATM and ATM+OC increase with increasing lead time, suggesting that ocean uncertainty is providing an extra source of uncertainty.
in processes that dictate the standard deviation in maximum wind speed. By contrast, adding $C_k$ variability to ATM (ATM+$C_k$) does not result in a statistically significant difference in standard deviation at any lead time relative to ATM. Given the lack of difference between ATM and ATM+$C_k$ for any intensity or structure metric (others shown below), ATM+$C_k$ is not discussed. While combining $C_d$ or OC with ATM uncertainty results in statistically larger standard deviation relative to ATM, combining all three uncertainty sources together (ATM+OC+$C_d$) does not result in a statistically larger standard deviation relative to ATM+OC.

Although the maximum wind speed standard deviation suggest that $C_d$ perturbations are effective at creating variability in TC intensity, other intensity metrics suggest that this result is particular to the definition of maximum wind speed. While the difference in maximum wind speed standard deviation between $C_d$ and $C_k$ is statistically significant at all lead times, the standard deviation difference in minimum SLP is not significant at any lead time (Fig. 3b), suggesting that $C_d$ variability has an outsized impact on maximum wind speed, but not other measures of TC intensity. In order to demonstrate how $C_d$ variability impacts the maximum speed, Fig. 4 shows the 48 h forecast of wind speed and $C_d$ perturbations for two members of the $C_d$ ensemble during the Bill forecast initialized 16 August 2009. Given that this storm is moving westward with time, the highest wind speeds should be on the northern side of the storm where the cyclonic winds of the TC constructively add with the easterly steering flow. For one member, there is a negative $C_d$ perturbation that aligns with the northern side of the storm, implying less drag in that quadrant (Fig. 4a). As a result, the maximum wind speed for this member is 82 kts. By contrast, a second member has a positive drag perturbation in the northern quadrant and a negative perturbation in the southwestern side of the storm (Fig. 4b). As a consequence, there is a second area of higher wind that aligns with the negative drag perturbation and the maximum wind speed and is relatively
smaller (71 kts). These figures suggest that drag perturbations tend to shift the maximum wind speed depending on how the drag perturbations align with the wind field.

Unlike $C_d$ and $C_k$ uncertainty, for which the minimum SLP standard deviation begins to level off after 48 h, the standard deviation for OC increases at a roughly constant rate ($1.8 \text{ hPa d}^{-1}$) until 96 h, when it becomes comparable to ATM. Combinations of various uncertainty sources also suggest limited value from adding $C_d$ uncertainty. While the difference in minimum SLP standard deviation between ATM+OC and ATM increases from 0.2 hPa at 24 h to 2.8 hPa at 96 h, suggesting that the addition of OC uncertainty increases in value with increasing lead time, the minimum SLP standard deviation in ATM+$C_d$ is less than 0.2 hPa larger than ATM (not statistically significant at any time). Moreover, adding $C_d$ uncertainty to ATM+OC results in a small, statistically insignificant decrease in standard deviation.

The previous results suggest that $C_d$ uncertainty adds variability to one measure of TC intensity (maximum wind speed) relative to ATM, but has little impact in another intensity metric (minimum SLP), thus it is of interest to assess the standard deviation in other intensity metrics. One alternative is the maximum axisymmetric (i.e., wavenumber-0) tangential (i.e., azimuthal) 10 m wind speed, which is related to maximum wind speed\(^4\), but does not suffer from the single grid-point issues of maximum wind speed. Fig. 3c shows that the standard deviation in this metric is similar to minimum SLP in that $C_d$ is associated with a statistically larger standard deviation during the first 24 h, but is similar thereafter. Moreover, adding $C_d$ to ATM or ATM+OC does not lead to a statistically larger standard deviation, while adding OC to ATM does.

The different sources of variability also project onto TC size, which can be measured by the radius of the 34-knot wind speed in each quadrant. During the first 96 h, the standard deviation in this quantity saturates between 3 and 6 nmile for $C_d$ and $C_k$, respectively, while OC continues

\(^4\)Vukicevic et al. (2014) showed that the sum of wavenumber 0 and 1 explain a significant amount of variability in maximum wind speed
to increase with time. By contrast, ATM has decreasing standard deviation through 96 h, then increases beyond that. It is not completely clear why the ATM result is obtained, though it is likely related to having generally weak storms during early forecast lead times that subsequently develop into stronger TCs at longer lead times and the nature of the 34 knot radius metric. As a TC develops from a weak to strong tropical storm, the 34 knot radius increase from 0 to 50 nmile in 6 h as the far-field wind increases from just under 34 knots to above. Any variability in the rate at which this occurs within the ensemble means that some members have a 34 knot radius greater than zero, while others have a value of zero. Once the storm intensifies, the 34 knot radius becomes better defined, leading to less variability in the radius itself. By 120 h, many of these cases represent TCs that have moved significantly poleward of their 0 h position, whereby the wind field expands and the wind speed gradient decreases, leading to more uncertainty in the 34 knot radius. Combining different uncertainty is not associated with a statistically meaningful difference relative to ATM.

**b. Forecast Error Comparisons**

While it is desirable to construct an ensemble that maximizes variability, the ensemble-mean forecast errors should also be less than the deterministic forecast. One way to evaluate this is to compute the percent improvement in the mean-absolute error for the ensemble-mean relative to the deterministic forecast averaged over all cases. This metric is computed via:

\[
\text{Improvement} = \frac{\text{MAE}_{det} - \text{MAE}_e}{\text{MAE}_{det}} \times 100\% ,
\]

where \(\text{MAE}_{det}\) (\(\text{MAE}_e\)) is the mean-absolute error in maximum wind speed, minimum SLP, or 34 knot wind radius for the deterministic (ensemble-mean) forecast. In this metric, positive values indicate that the ensemble mean has lower errors.
Ensembles that consist of various uncertainty sources are characterized by various improvements relative to using a deterministic forecast (Fig. 5a). Experiments that include ATM uncertainty are characterized by a 5-20% improvement relative to the deterministic forecast (except at 48 h), with relatively little difference among the experiments with multiple uncertainty sources. Not surprisingly, the experiments with a single uncertainty source (OC, C_d, C_k) show little improvement over the deterministic forecast, which suggests that these sources of uncertainty are merely adding variability about a mean value. Moreover, the bias values are similar at all lead times, suggesting that the uncertainty sources do not add systematically change the model behavior (not shown). Minimum SLP improvements show similar behavior to maximum wind speed in that experiments with ATM uncertainty are characterized by $\geq 10\%$ improvement at most lead times\(^5\) (Fig. 5c). By contrast, the single uncertainty sources without ATM uncertainty (OC, C_d, C_k) are near zero at most lead times. Finally, 34 knot radius improvements are between 5-10% for experiments with ATM uncertainty and slightly negative for those without (Fig. 5e).

In addition to having lower ensemble-mean errors relative to a deterministic forecast, a well-calibrated ensemble is characterized by the root-mean-square errors in the ensemble mean being comparable to the ensemble standard deviation when averaged over many cases (e.g., Murphy 1988; Houtekamer et al. 2005). This condition can be measured by computing the ratio between the ensemble standard deviation and the RMS error in the ensemble mean via:

$$\text{Ratio} = \sqrt{\frac{1}{N_{\text{cases}}} \sum_{i=1}^{N_{\text{cases}}} \left( \sigma^2_{\text{ens}} + \sigma^2_{\text{bt}} \right)}$$

$$\sqrt{\frac{1}{N_{\text{cases}}} \sum_{i=1}^{N_{\text{cases}}} (x_i - x_{\text{bt}})^2}$$

where $x$ denotes the forecast metric to verify (either maximum wind speed, minimum SLP, or 34 knot wind radius in each quadrant), $x_{\text{bt}}$ denotes the ensemble mean, $x_{\text{bt}}$ is the corresponding-

\(^5\)The small improvement at 48 h is related to the correct timing of RI in the deterministic forecast, while most of the ensemble members have a delayed RI.
time best-track value, $N_{\text{cases}}$ is the number of forecasts at that particular lead time, $\sigma_{\text{ens}}$ is the ensemble standard deviation in $x$ and $\sigma_{\text{br}}$ is the uncertainty in the best-track value taken from the method outlined in section 5 of Torn and Snyder (2012). While most ensemble studies compare the ensemble-mean error to the ensemble standard deviation because the uncertainty in the verification value is generally small compared to the ensemble standard deviation, the standard deviation in the best track values is comparable to the ensemble standard deviation at many lead times (e.g., Torn and Snyder 2012; Landsea and Franklin 2013). If the ensemble standard deviation is consistent with the ensemble-mean errors, this ratio should be approximately one at all lead times; values below (above) one indicate that the ensemble standard deviation is too small (large).

While the ensemble standard deviation is generally smaller than the ensemble mean errors for all three metrics, combining uncertainty sources with ATM generally improves the ensemble performance. For all experiments, the highest consistency for maximum wind speed occurs within the first 24 h of the forecast; however, the values decrease thereafter suggesting that the ensemble-mean errors are increasing faster than the ensemble standard deviation (c.f., Fig. 2b, Fig. 5b). Not surprisingly, the experiments with ATM uncertainty have a ratio closer to one compared to the ones that do not; however, there is little difference between ATM and the experiments with ATM combined other uncertainty sources, particularly when $C_d$ uncertainty is added (Fig. 5b). For minimum SLP, the consistency ratios are generally similar to maximum wind speed; however, unlike maximum wind speed, adding OC uncertainty generally leads to a better match between errors and standard deviation over ATM uncertainty alone (statistically significant beyond 48 h; Fig. 5d). Nevertheless, all of the experiments are still deficient in ensemble standard deviation at most lead times, mainly because of the slow growth in the ensemble standard deviation. Finally, the 34 knot wind radius consistency ratio is generally around 0.9 for all experiments with ATM uncertainty, compared to less than 0.8 for OC, $C_d$ and $C_k$; however, there are few times where combinations of
uncertainty are associated with a better consistency ratio that is statistically different from ATM (Fig. 5f).

4. Variability Growth

Given the differences in TC intensity standard deviation shown above, it is of interest to understand how these various sources of uncertainty translate into variability in the processes that modulate TC intensity and the axisymmetric TC structure. Specifically, this section will evaluate the impact of variability in axisymmetric surface latent heat flux, which drives convection in the TC and the associated lower tropospheric convergence, resulting in a spin up of the tangential wind via angular momentum conservation. Moreover, angular momentum is lost via surface friction; therefore, the variability in the axisymmetric momentum flux (measured through the friction velocity, $u^*$) is also evaluated. While axisymmetric averages of these two processes are not the only ways to modulate intensity, particularly given that all storms have asymmetries, this framework provides a relatively straightforward paradigm for evaluating the variety of cases used here.

a. Friction Variability

Tropical cyclones lose momentum to the surface through friction; therefore, one possible way of introducing variability in TC intensity is to have uncertainty in the amount of friction that is experienced by the TC. Friction variability can result from two different sources: differences in the structure of the TC wind field, which can be introduced by atmospheric uncertainty, or through differences in the momentum flux, either by changing the wind speed or perturbing $C_d$. By contrast, OC and $C_k$ uncertainty would be expected to have an indirect impact on friction variability through changing the TC wind field over time. As a consequence, it is of interest to evaluate how each of the sources of uncertainty translate into variability in the area-average $u^*$. For
all experiments and lead times, the standard deviation of $u^*$ is proportional to $u^*$ itself; therefore, it is easier to interpret the ratio of the $u^*$ standard deviation and the corresponding-time deterministic forecast $u^*$.

The presence of individual uncertainty sources is associated with distinctive distributions in the standard deviation ratio of $u^*$. Figure 6a shows histograms of the $u^*$ standard deviation ratio during the first 48 h of each case. This time period is chosen because it represents the duration over which the ensemble standard deviation in the various intensity metrics (maximum wind speed, minimum SLP, maximum tangential wind) are increasing with time and not leveled off. In the $C_d$ experiment, the histogram of the standard deviation ratio is strongly peaked between 0.02-0.04 (i.e., the standard deviation in $u^*$ is between 2-4% of the deterministic value). By contrast, the distribution for ATM is much broader and skewed toward higher values; the median value in ATM is 0.10, compared to 0.03 for $C_d$. Combining ATM and $C_d$ uncertainty together shifts the distribution toward slightly higher values on the low end, while the median value also increases by 0.005. By contrast, the standard deviation ratio for OC and $C_k$ peaks at or below the maximum in the $C_d$ distribution, which is not surprising since any $u^*$ variability would be indirect. The upshot of these results is that stochastic representations of $C_d$ introduce a limited amount of variability to the area-average friction experienced by the TC; however, uncertainty in the TC structure, as occurs with ATM uncertainty, is associated with greater variability in the surface friction.

Given the dynamical link between TC intensity and roughness as described above, it is possible that individual cases that have large uncertainty in the area-average $u^*$ over a prolonged period of time could result in larger TC intensity standard deviation. This hypothesis is evaluated by computing the correlation between the standard deviation in the 0-48 h area-average $u^*$ and the 48 h maximum tangential wind standard deviation over all cases. For all experiments, the correlation between these two quantities is between 0.5-0.6, suggesting that cases with large friction
variability are associated with large intensity variability. This result is not necessarily surprising because the friction velocity is proportional to the wind speed itself; therefore, variability in the TC intensity should be associated with variability in the TC wind field, and hence variability in area-average surface friction. Nevertheless, it will be shown later that the correlation between TC intensity standard deviation and latent heat standard deviation is smaller.

b. Latent Heat Variability

One of the intriguing aspects of these results is the lack of variability in any intensity metric by the introduction of $C_k$ perturbations, which should be expected to modulate the surface latent heat flux at individual grid points. By contrast, if TC intensity only responds to changes to the area-averaged latent heat flux, then this perturbation method might not be as effective. In addition, variability in surface latent heat flux can also result from the air-sea disequilibrium, which is a function of variability in the atmosphere and ocean, and the horizontal distribution of wind. The role of various uncertainty sources on surface latent heat flux variability is evaluated by computing histograms of the standard deviation of area-average surface latent heat flux within 240 km of the TC center in a manner similar to what is done with $u^*$. This radius represents the area over which the latent heat flux is maximized; other radii show similar results (not shown).

In general, experiments with combinations of ATM and OC uncertainty are associated with the largest variability in latent heat flux (Fig. 6b). The $C_k$ ensemble histogram is strongly peaked below 0.02, while ATM peaks between 0.04-0.08 (median of 0.07), but with a tail that extends beyond 0.28. Moreover, the experiment ATM+$C_k$ is very similar to ATM, which provides some explanation for why ATM+$C_k$ had similar intensity variability as ATM. The OC distribution is peaked at an even higher value than ATM (0.12-0.16), with a median value of 0.13. Furthermore, combining ATM and OC uncertainty is associated with greater latent heat flux variability than
either ATM and OC alone (median of 0.15), suggesting that the combination of ATM and OC uncertainty drive the largest variability in latent heat flux.

Given that tropical cyclones derive energy from surface latent heat fluxes, one might expect that cases with larger latent heat flux standard deviation might be associated with greater TC intensity variability; however, that does not appear to be the case. With the exception of $C_k$, the correlation between the standard deviation in the 0-48 h area-average latent heat flux and the standard deviation in the 48 h maximum tangential wind is less than 0.15 (not statistically significant) for each experiment. As a consequence, it appears that variability in the area-average latent heat flux is not a strong determining factor in generating intensity variability.

c. Structure Variability

The results presented to this point suggest that all of these uncertainty sources introduce variability in intensity metrics and surface fluxes; however, it is also important to evaluate how these uncertainty sources impact TC structure variability. Here, the variability in TC structure is evaluated via radius-height cross-sections of axisymmetric dry total energy (DTE) norm averaged over all cases, which has been used in other TC predictability studies (e.g., Sippel and Zhang 2008; Lang et al. 2012). Specifically, this quantity is calculated as:

$$DTE(r,z) = \frac{1}{N_{cases}N_{ens}} \sum_{i=1}^{N_{cases}} \sum_{n=1}^{N_{ens}} \left( u_{i,n}(r,z)^2 + v_{i,n}(r,z)^2 + \frac{C_p}{T_r} T_{i,n}(r,z)^2 \right)$$  \hspace{1cm} (6)

where $v_i(r,z)'$ is the perturbation axisymmetric tangential wind for member $n$ at radius $r$ and height $z$ for case $i$, $u'$ is the perturbation radial wind, $C_p$ is the specific heat of air, $T_r$ is a reference temperature (here 300 K), $T'$ is the perturbation temperature, $N_{cases}$ is the number of forecasts, and $N_{ens}$ is the number of ensemble members. Although this quantity could be calculated at any
arbitrary lead time, the analysis presented here is limited to the 48 h forecasts for brevity; other times showed similar patterns (not shown).

Figure 7a shows the distribution of DTE for ATM relative to the average tangential and radial winds averaged over all cases. Overall, the DTE is maximized radially inward of the tangential wind maximum at every vertical level, which coincides with the largest horizontal gradients in tangential wind. Moreover, the largest DTE values are within the boundary layer in the part of the TC characterized by radial wind convergence. In addition, there is a region of DTE exceeding 5 J kg$^{-1}$ extending radially-outward from the TC between 12-15 km within the TC outflow.

Each of the remaining ensemble experiments exhibit distinctive distributions of dry energy relative to ATM. In order to account for the variability in DTE within the TC, the DTE of the remaining experiments are computed as a percent-difference with respect to the ATM values; positive (negative) values indicate greater (less) total energy relative to ATM. For OC, the dry total energy is less than ATM at most radii and height, though the largest percentage change is within the DSE maximum (Fig. 7b), while the reduction in DSE for $C_d$ and $C_k$ compared to ATM is more uniform (Fig. 7c,d). This result suggests that while the pattern of structure variability for $C_d$ and $C_k$ is similar to ATM (just lower amplitude), OC has somewhat greater variability beyond the radius of maximum winds compared to inside the tangential wind maximum. This result is supported by the ATM+OC DSE differences, which are maximized in between 60-150 km radius and within the outflow layer above 15 km. This result suggests that atmospheric and oceanic uncertainty introduce complementary variability in TC structure. ATM uncertainty leads to TC structure variability concentrated near the TC core, while OC uncertainty tends to create more TC structure variability at larger radii.
5. Sensitivity Experiments

The remainder of this paper evaluates whether the OC uncertainty results are sensitive to having a dynamic ocean and understanding how the spatial structure of the $C_d$ and $C_k$ perturbations translate into TC intensity variability. This is accomplished by repeating the respective single uncertainty source experiments and comparing to the original experiment.

a. Variability from Ocean Model

While having a dynamic ocean is a physically more realistic representation of the processes that occur with a TC, this dynamic ocean may actually limit the impact of SST variability. In isolation, a positive SST perturbation would initially lead to a more intense TC and hence stronger winds over a large area. These stronger winds can subsequently induce greater ocean upwelling, thus reducing the magnitude of the positive SST perturbation and slightly weaken the TC. As a consequence, the sensitivity of TC intensity to SST could be smaller with a dynamic ocean compared to using fixed SST. To evaluate this possibility, the OC experiment is repeated, but with the 1D ocean model turned off (denoted OC-Fixed) and compared to OC; differences between these two experiments are thus solely related to ocean mixing.

Even though many of the TCs used in this study achieve major hurricane status and hence would be expected to generate significant upwelling of cold water, turning off the 1D ocean model has relatively little impact on the standard deviation of various TC intensity metrics (Fig. 8). During the first 48 h, the mean standard deviation in maximum wind and minimum SLP are nearly identical, while beyond that time, the maximum wind speed and maximum tangential wind standard deviation in the OC-Fixed simulations are 3-4% greater with a fixed ocean (difference not statistically significant). As a consequence, it appears that having dynamic SSTs does not significantly
influence variability in TC intensity forecasts; most of the TC intensity variability results from the initial SST uncertainty.

\textit{b. Spatial Variations to }C_d\textit{ and }C_k\textit{ }

In the above experiments, \( C_d \) and \( C_k \) variability is introduced by adding stochastic noise to these parameters, which does not have any spatial correlation on the outer (36 km) grid. As a consequence, the previous section demonstrated that the variability in these parameters does not introduce significant variability in either the area-averaged \( u^* \) or surface latent heat flux. In order to assess the role of spatially-uncorrelated \( C_d \) and \( C_k \) perturbations, two additional experiments are carried out where either \( C_d \) (denoted \( C_d \)-global) or \( C_k \) (denoted \( C_d \)-global) are perturbed using eqn. (2), but where \( \epsilon \) is not a function of space or time. In these experiments, the \( C_d \) and \( C_k \) perturbations are perfectly correlated at all horizontal locations. These experiments are similar to what was carried out in Green and Zhang (2014) and represent the maximum amount of intensity variability that could be introduced by perturbing the exchange coefficients.

As might be expected, spatially-constant perturbations to the exchange coefficients generate much larger variability in TC intensity than spatially-varying perturbations. Figure 9 indicates that the ensemble standard deviation in maximum wind speed for \( C_d \)-global is nearly twice the value of \( C_d \) at all lead times and comparable to OC after 24 h (c.f., Fig. 3a). Similarly, the difference between \( C_k \)-global and \( C_k \) is \( \leq 1 \) knot before 24 h, and increases thereafter, with values that are indistinguishable from \( C_d \)-global. For MSLP, both global experiments introduce significantly more variability, with \( C_k \)-global statistically larger starting 60 h (Fig. 9). As a consequence, these results suggest that part of the reason for the lack of intensity uncertainty in \( C_d \) and \( C_k \) experiments is from the uncorrelated noise that was added to these quantities. More systematic methods of
introducing stochastic noise to these quantities that maintains spatial coherence may provide more beneficial impact from perturbing the exchange coefficients, but is beyond the scope of this study.

6. Summary and Conclusions

This manuscript describes the variability in AHW TC intensity forecasts that results from current estimates of the uncertainty in the atmosphere, ocean, and exchange coefficients. Atmospheric uncertainty is obtained using the analysis perturbations from a cycling ensemble Kalman filter system that includes uncertainty in both the large-scale environment and the vortex. Ocean uncertainty is obtained by adding scaled perturbations sampled from a climatology of ocean analyses, while parameter uncertainty is produced by multiplying the exchange coefficients by a spatially-varying factor that is consistent with the measurement uncertainty in these two quantities. Ensemble experiments are carried out for 35 initialization times that are characterized by different combinations of uncertainty and are subsequently compared to one another to quantify what processes contribute the greatest uncertainty to TC intensity forecasts and suggest an optimal design TC ensemble prediction systems.

On average, the various sources of uncertainty contribute different amounts of variability to TC intensity forecasts and produce different rates of growth in the ensemble standard deviation. When considering a single uncertainty source, ATM uncertainty leads to the largest ensemble standard deviation for all TC intensity metrics. While ocean uncertainty results in consistent growth in the ensemble standard deviation throughout the 5 d forecast, $C_d$ and $C_k$ uncertainty is associated with increasing ensemble standard deviation during the first 48 h, which saturates thereafter, suggesting that variability in these parameters has limited ability to introduce variability in TC intensity forecasts. Whereas $C_d$ uncertainty is associated with a statistically-significant increase in maximum wind speed standard deviation relative to $C_k$, the two values have similar minimum SLP and
maximum tangential wind standard deviation. This result is due to the nature of using maximum wind speed as a metric, which is defined as the highest 10 m wind speed, and thus is quite closely tied to how the $C_d$ perturbations line up with the region of largest wind speed. As a consequence, this result underlines the importance of looking at multiple TC metrics, and not just the maximum wind speed. Combining ATM uncertainty with OC uncertainty is associated with the larger TC intensity standard deviation relative to ATM, though the standard deviations are not additive. By contrast, combining $C_d$ or $C_k$ uncertainty with ATM shows little change in standard deviation relative to ATM alone, suggesting that the forecasts are relatively insensitive to exchange coefficient variability.

For TC maximum wind speed and 34 knot wind radius, ensemble experiments with ATM uncertainty are characterized by an ensemble-mean forecast error that is lower than the deterministic forecast for most forecast lead times; however, those experiments without ATM uncertainty are characterized by little reduction in forecast error in the ensemble-mean, suggesting that these perturbations are just adding variability about the ensemble mean, without providing a more skillful forecast. Combining ATM with OC variability does not further reduce the ensemble-mean forecast errors; however, because the ensemble standard deviation is larger in ATM+OC relative to ATM, the ensemble standard deviation is more consistent with the forecast errors. Adding uncertainty in the exchange coefficients did not appreciably reduce the ensemble-mean errors further, or increase the standard deviation. Nevertheless, all of the ensemble experiments are characterized by a lack of standard deviation relative to the ensemble-mean errors.

The introduction of various sources of uncertainty are associated with different amounts of latent heat and surface friction variability, which in turn results in different variability in TC structure. Despite adding uncertainty in $C_d$, which is proportional to the surface friction, area-average surface-friction in the $C_d$ experiment is roughly 3% of the typical friction value, compared to 10%
for ATM uncertainty. Moreover, the combination of ATM and $C_d$ uncertainty has relatively little impact on the area-average friction variability compared to ATM alone. Similarly, variability in $C_k$ alone also produces limited variability in area-average surface latent heat flux compared to ATM and OC uncertainty, with the greatest latent heat flux variability coming from the combination of ATM and OC uncertainty. On a case-by-case basis, the standard deviation in maximum wind speed is highly correlated with the standard deviation in area-average friction, but not necessarily the latent heat fluxes. This result suggests that cases with larger variability in surface latent heat flux are not necessarily tied to large intensity variability. Moreover, introducing ocean uncertainty to ATM is associated with a relatively larger increase in TC structure variance away from the TC core compared to near the TC core, suggesting that ocean uncertainty is effective in generating variability in the TC rainband region.

One of the surprising aspects of this study is that stochastic representation of the exchange coefficients have little impact on TC intensity variability. This result is likely a consequence of adding spatially-uncorrelated noise to the exchange coefficients, which appears to produce little systematic impact on either the surface friction or latent heat flux on the scale of the TC. This conclusion is supported by the large TC intensity standard deviation for experiments where $\varepsilon$ is not a function of space. This approach is expected to produce maximum systematic impact on the TC; however, spatially-constant exchange coefficient perturbations are probably not physically meaningful because it is unlikely that $C_k$ or $C_d$ in nature is systematically higher or lower than the formulation used in AHW at all wind speeds. Instead, it is more likely that the exchange coefficients are not a single number for a given roughness length. A compromise approach would be to introduce stochastic perturbations to $C_d$ and $C_k$ that have spatial correlation using an approach similar to the Stochastic Perturbed Physics Tendency (SPPT; e.g., Buizza et al. 1999), which would allow the
user to specify a spatial and timescale for the exchange coefficient perturbations; future work will likely employ this method.

Overall, these results suggest that TC ensemble prediction systems could benefit from the inclusion of ocean uncertainty. Even though the ocean model and ocean initial condition methods used in this study are fairly primitive, the TC intensity standard deviation increases with time, which demonstrates the importance of using different SST fields for each ensemble member. Future modeling systems should consider using a fully 3D ocean model with initial condition perturbations coming from an ocean ensemble prediction system that has initial condition perturbations that are consistent with the analysis error magnitude and structure. Furthermore, the lack of consistency between the ensemble standard deviation and error suggests that additional sources of uncertainty are necessary to have a skillful ensemble prediction system. Those are most likely to come from adding uncertainty in the microphysics or sub-gridscale mixing parameterizations. Future work will likely involve evaluating the impact of uncertainty in these processes on TC intensity forecasts.

Acknowledgments. This paper benefited from discussions with Rosimar Rios-Berrios, Chris Davis, Jim Doyle, Jimy Dudhia, Mike Montgomery, and Wei Wang. The author acknowledges the NOAA Research and Development High Performance Computing Program for providing computing and storage resources that have contributed to the research results reported within this paper. This research is supported by HFIP via NOAA awards NA12NWS4680003 and NA14NWS4680027.

References


LIST OF TABLES

Table 1. Forecast initialization times used in this study (all are initialized at 0000 UTC). 42

Table 2. List of ensemble experiments performed in this study. A D denotes deterministic values (i.e., all ensemble members use the same value), while E denotes ensemble values. 43

Table 3. Observations Assimilated by the ensemble data assimilation system. 44

Table 4. Observation cycling beginning and ending dates. 45
**Table 1.** Forecast initialization times used in this study (all are initialized at 0000 UTC).

<table>
<thead>
<tr>
<th>Storm</th>
<th>Dates</th>
<th>Storm</th>
<th>Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gustav</td>
<td>26 August, 28 August 2008</td>
<td>Hanna</td>
<td>29 August, 1 September 2008</td>
</tr>
<tr>
<td>Ike</td>
<td>2 September, 6 September, 9 September 2008</td>
<td>Bill</td>
<td>16 August, 19 August 2009</td>
</tr>
<tr>
<td>Erika</td>
<td>2 September 2009</td>
<td>Fred</td>
<td>8 September 2009</td>
</tr>
<tr>
<td>Danielle</td>
<td>22 August, 24 August, 27 August 2010</td>
<td>Earl</td>
<td>26 August, 29 August, 31 August 2010</td>
</tr>
<tr>
<td>Igor</td>
<td>9 September, 12 September, 16 September 2010</td>
<td>Julia</td>
<td>13 September, 16 September 2010</td>
</tr>
<tr>
<td>Karl</td>
<td>15 September 2010</td>
<td>Otto</td>
<td>7 October 2010</td>
</tr>
<tr>
<td>Richard</td>
<td>21 October 2010</td>
<td>Tomas</td>
<td>31 October, 3 November 2010</td>
</tr>
<tr>
<td>Irene</td>
<td>23 August, 25 August 2011</td>
<td>Katia</td>
<td>30 August, 3 September 2011</td>
</tr>
<tr>
<td>Maria</td>
<td>9 September 2011</td>
<td>Ophelia</td>
<td>28 September 2011</td>
</tr>
<tr>
<td>Philippe</td>
<td>28 September 2011</td>
<td>Rina</td>
<td>24 October 2011</td>
</tr>
</tbody>
</table>
TABLE 2. List of ensemble experiments performed in this study. A D denotes deterministic values (i.e., all
ensemble members use the same value), while E denotes ensemble values.

<table>
<thead>
<tr>
<th>Experiment Name</th>
<th>Atmosphere</th>
<th>Ocean</th>
<th>$C_d$</th>
<th>$C_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHW</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>ATM</td>
<td>E</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>OC</td>
<td>D</td>
<td>E</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>$C_d$</td>
<td>D</td>
<td>D</td>
<td>E</td>
<td>D</td>
</tr>
<tr>
<td>$C_k$</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>ATM+OC</td>
<td>E</td>
<td>E</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>ATM+$C_d$</td>
<td>E</td>
<td>D</td>
<td>E</td>
<td>D</td>
</tr>
<tr>
<td>ATM+$C_k$</td>
<td>E</td>
<td>D</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>ATM+OC+$C_d$</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>D</td>
</tr>
</tbody>
</table>
Table 3. Observations Assimilated by the ensemble data assimilation system.

<table>
<thead>
<tr>
<th>Observation Type</th>
<th>Information Assimilated</th>
<th>Error Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automated (ASOS)</td>
<td>Altimeter</td>
<td>NCEP GFS Statistics</td>
</tr>
<tr>
<td>Rawinsondes</td>
<td>$u, v, T, q_{vapor}$</td>
<td>NCEP GFS Statistics</td>
</tr>
<tr>
<td>Dropwindsondes</td>
<td>$u, v, T, q_{vapor}$</td>
<td>NCEP GFS Statistics</td>
</tr>
<tr>
<td>ACARS</td>
<td>$u, v, T$</td>
<td>NCEP GFS Statistics</td>
</tr>
<tr>
<td>Atmospheric Motion Vectors</td>
<td>$u, v$</td>
<td>NCEP GFS Statistics</td>
</tr>
<tr>
<td>GPS Radio Occultation Data</td>
<td>refractivity</td>
<td>NCEP GFS Statistics</td>
</tr>
<tr>
<td>Tropical Cyclone Vitals</td>
<td>Lat., Lon., Min. SLP</td>
<td>Torn and Snyder (2012)</td>
</tr>
</tbody>
</table>
### TABLE 4. Observation cycling beginning and ending dates.

<table>
<thead>
<tr>
<th>Start Date</th>
<th>End Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000 UTC 12 August 2008</td>
<td>1200 UTC 13 September 2008</td>
</tr>
<tr>
<td>0000 UTC 8 August 2009</td>
<td>1200 UTC 12 September 2009</td>
</tr>
<tr>
<td>0000 UTC 18 August 2010</td>
<td>1200 UTC 7 November 2010</td>
</tr>
<tr>
<td>0000 UTC 29 July 2011</td>
<td>1200 UTC 28 October 2011</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Fig. 1. (a) Sea-surface temperature (units: K) and (b) mixed layer depth (units: m) initial condition perturbation obtained from a randomly-sampled September day.

Fig. 2. Mean-absolute error (solid) in (a) TC position (nmile) and (b) maximum wind speed (knots) as a function of forecast hour for the cases listed in Table 1. AHW4 denotes the AHW deterministic forecast, OFCL is the National Hurricane Center official forecast, AVNO is the GFS forecast, LGEM is the Logistical Growth Equation Model, and OCD5 is the Decay-SHIFOR model. The dashed lines in (b) denote bias (forecast-best track), while the numbers across the bottom denote the number of cases at each lead time.

Fig. 3. Ensemble standard deviation in (a) maximum wind speed, (b) minimum SLP, (c) maximum tangential wind, and (d) 34 knot wind radius in each quadrant for the ensemble experiments averaged over all cases as a function of lead time. See Table 2 for a description of each experiment.

Fig. 4. Drag Coefficient perturbation (shading) and 48 h 10 m wind speed (contours, units: m s\(^{-1}\)) for two members of the \(C_d\) ensemble forecasts of Bill initialized 0000 UTC 16 August 2009. The values along the x and y axis are the distance from the center of the TC (units: km). The + denotes the location of the maximum 10 m wind speed.

Fig. 5. (a) Percent reduction in mean-absolute maximum wind speed error for the ensemble-mean from each ensemble experiment relative to the deterministic AHW forecast as a function of lead time. Each line corresponds to a different ensemble experiment (see Table 2 for a description of each experiment). (b) Ratio of the ensemble standard deviation in maximum wind speed to the RMS error in the ensemble-mean forecast as a function of forecast lead time. (c) and (d), as in (a) and (b), but for minimum SLP. (e) and (f), as in (a) and (b), but for the 34 knot wind radius.

Fig. 6. Histogram of the ratio of the ensemble standard deviation in \(u^*\) averaged within 120 km of the TC center to the corresponding-time AHW deterministic forecast every 1 h during the first 48 h of each initialization time. (b) as in (a), but for the surface latent heat flux averaged within 240 km of the TC center.

Fig. 7. (a) 48 h storm-centered dry total energy averaged over all cases (shading, units: J kg\(^{-1}\)). The thick contours are the tangential wind (units: m s\(^{-1}\)), while the thin contours are the radial wind every 3 m s\(^{-1}\) (dashed negative) averaged over all cases. (b) Percentage difference between the 48 h storm-centered dry total energy averaged over all cases for OC relative to ATM (shading). The contours show the 48 h storm-centered dry total energy averaged over all cases for ATM (units: J kg\(^{-1}\)). (c), (d), and (e), as in (b), but for the \(C_d\), \(C_k\), and ATM+OC, experiments, respectively.

Fig. 8. As in Fig. 3a-b, but for the experiment with the 1D ocean model turned on (OC) and turned off (OC-Fixed).

Fig. 9. As in Fig. 3a-b, but for the experiment with spatially-uncorrelated \(C_d\) and \(C_k\) perturbations (solid) and spatially-fixed \(C_d\) and \(C_k\) perturbations (dashed).
FIG. 1. (a) Sea-surface temperature (units: K) and (b) mixed layer depth (units: m) initial condition perturbation obtained from a randomly-sampled September day.
Fig. 2. Mean-absolute error (solid) in (a) TC position (nmile) and (b) maximum wind speed (knots) as a function of forecast hour for the cases listed in Table 1. AHW4 denotes the AHW deterministic forecast, OFCL is the National Hurricane Center official forecast, AVNO is the GFS forecast, LGEM is the Logistical Growth Equation Model, and OCD5 is the Decay-SHIFOR model. The dashed lines in (b) denote bias (forecast-best track), while the numbers across the bottom denote the number of cases at each lead time.
FIG. 3. Ensemble standard deviation in (a) maximum wind speed, (b) minimum SLP, (c) maximum tangential wind, and (d) 34 knot wind radius in each quadrant for the ensemble experiments averaged over all cases as a function of lead time. See Table 2 for a description of each experiment.
FIG. 4. Drag Coefficient perturbation (shading) and 48 h 10 m wind speed (contours, units: m s$^{-1}$) for two members of the $C_d$ ensemble forecasts of Bill initialized 0000 UTC 16 August 2009. The values along the x and y axis are the distance from the center of the TC (units: km). The + denotes the location of the maximum 10 m wind speed.
FIG. 5. (a) Percent reduction in mean-absolute maximum wind speed error for the ensemble-mean from each ensemble experiment relative to the deterministic AHW forecast as a function of lead time. Each line corresponds to a different ensemble experiment (see Table 2 for a description of each experiment). (b) Ratio of the ensemble standard deviation in maximum wind speed to the RMS error in the ensemble-mean forecast as a function of forecast lead time. (c) and (d), as in (a) and (b), but for minimum SLP. (e) and (f), as in (a) and (b), but for the 34 knot wind radius.
FIG. 6. Histogram of the ratio of the ensemble standard deviation in $u^*$ averaged within 120 km of the TC center to the corresponding-time AHW deterministic forecast every 1 h during the first 48 h of each initialization time. (b) as in (a), but for the surface latent heat flux averaged within 240 km of the TC center.
Fig. 7. (a) 48 h storm-centered dry total energy averaged over all cases (shading, units: J kg$^{-1}$). The thick contours are the tangential wind (units: m s$^{-1}$), while the thin contours are the radial wind every 3 m s$^{-1}$ (dashed negative) averaged over all cases. (b) Percentage difference between the 48 h storm-centered dry total energy averaged over all cases for OC relative to ATM (shading). The contours show the 48 h storm-centered dry total energy averaged over all cases for ATM (units: J kg$^{-1}$). (c), (d), and (e), as in (b), but for the $C_d$, $C_k$, and ATM+OC, experiments, respectively.
Fig. 8. As in Fig. 3a-b, but for the experiment with the 1D ocean model turned on (OC) and turned off (OC-Fixed).
Fig. 9. As in Fig. 3a-b, but for the experiment with spatially-uncorrelated \( C_d \) and \( C_k \) perturbations (solid) and spatially-fixed \( C_d \) and \( C_k \) perturbations (dashed).