Experimental Tropical Cyclone Forecasts Using a Variable-Resolution Global Model

COLIN M. ZARZYCKI
National Center for Atmospheric Research,* Boulder, Colorado

CHRISTIANE JABLONOWSKI
Department of Atmospheric, Oceanic, and Space Sciences, University of Michigan, Ann Arbor, Michigan

(Manuscript received 20 April 2015, in final form 5 July 2015)

ABSTRACT

Tropical cyclone (TC) forecasts at $\Delta x \sim 14$-km horizontal resolution ($0.125^\circ$) are completed using variable-resolution (V-R) grids within the Community Atmosphere Model (CAM). Forecasts are integrated twice daily from 1 August to 31 October for both 2012 and 2013, with a high-resolution nest centered over the North Atlantic and eastern Pacific Ocean basins. Using the CAM version 5 (CAM5) physical parameterization package, regional refinement is shown to significantly increase TC track forecast skill relative to unrefined grids ($\Delta x \sim 55$ km, $0.5^\circ$). For typical TC forecast integration periods (approximately 1 week), V-R forecasts are able to nearly identically reproduce the flow field of a globally uniform high-resolution forecast. Simulated intensity is generally too strong for forecasts beyond 72 h. This intensity bias is robust regardless of whether the forecast is forced with observed or climatological sea surface temperatures and is not significantly mitigated in a suite of sensitivity simulations aimed at investigating the impact of model time step and CAM’s deep convection parameterization. Replacing components of the default physics with Cloud Layers Unified by Binormals (CLUBB) produces a statistically significant improvement in forecast intensity at longer lead times, although significant structural differences in forecasted TCs exist. CAM forecasts the recurvature of Hurricane Sandy into the northeastern United States 60 h earlier than the Global Forecast System (GFS) model using identical initial conditions, demonstrating the sensitivity of TC forecasts to model configuration. Computational costs associated with V-R simulations are dramatically decreased relative to globally uniform high-resolution simulations, demonstrating that variable-resolution techniques are a promising tool for future numerical weather prediction applications.

1. Introduction

Increasing horizontal resolution has been shown to be critically important in the improvement of numerical forecasting over the past few decades. While tropical cyclone (TC) forecasting has correspondingly benefited, current operational grid spacings for global products continue to be less than is required for proper representation of these features (NOAA Science Advisory Board 2006).

To alleviate these issues, the use of limited-area models (LAMs, also referred to as regional or nested models) in the forecasting community has become popular. These models only simulate a small portion of the global domain, allowing for available computational resources to be “targeted” to a specific region of interest. This provides greatly increased horizontal resolution over the cyclone itself, which has been shown to provide a more realistic dynamical representation of storm structure (e.g., Davis et al. 2010; Jin et al. 2014). However, LAMs require lateral boundary conditions to drive the innermost domain. These boundary conditions are typically derived from global forecast model output, which is computed at a coarser resolution and almost always utilizes a different dynamical core and set

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Colin M. Zarzycki, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307.
E-mail: zarzycki@ucar.edu

DOI: 10.1175/MWR-D-15-0159.1

© 2015 American Meteorological Society
of physical parameterizations, potentially introducing parent model biases and interpolation errors into the LAM (Warner et al. 1997; Mesinger and Veljovic 2013). In addition, the application of the boundary conditions to the regional domain is sometimes numerically inconsistent and ill posed (McDonald 2003). For the purposes of TC forecasting, LAMs have historically performed worse than global forecast models in terms of TC track beyond three days, in part because of these issues (Gall et al. 2013).

Variable-resolution general circulation models (VRGCMs) serve to apply this idea of targeted, regional use of computing resources, but can do so without the need for externally forced boundary conditions. Locally high resolution is achievable within a global, unified modeling framework, allowing the global and regional scales to interact in a more physically consistent manner. In addition, the same formulation of subgrid processes, such as turbulence and surface fluxes, can be applied across the global domain. Some examples of VRGCMs in recently published literature include Skamarock et al. (2012), Harris and Lin (2013), and Zarzycki et al. (2014b), who have all demonstrated increased regional model skill when locally refining horizontal resolution.

Since VRGCMs do not require high resolution over the global domain, they come at a greatly reduced computational cost relative to traditional globally uniform models. This may allow simulations to be run at higher horizontal resolutions, for longer durations, or as multimember ensembles given a predetermined computing load. Ensemble forecasting, in particular, has been shown to greatly improve weather and climate forecasts (Krishnamurti et al. 1999), with tropical cyclones being a notable example (Gall et al. 2013).

In section 2, we offer an introduction to the spectral element version of the Community Atmosphere Model, the variable-resolution configuration used in this study, the procedure for initializing the model in a forecast framework, and the techniques utilized to track tropical cyclones and assess forecast errors. Section 3 assesses the results of the base set of simulations while section 4 provides results from a suite of physical parameterization perturbation experiments that investigate the sensitivity of forecasted track and intensity to these modifications. Specific forecasts of Hurricane Sandy are highlighted in section 5, including a discussion of the impact of subgrid physics on the correct forecast of the storm’s recurvature toward the northeastern United States. Section 6 provides parallel scaling results comparing the variable-resolution configuration to forecasts with globally uniform grids. Finally, conclusions and future work are discussed in section 7.

2. Variable resolution in CAM-SE

a. CAM-SE

The Community Atmosphere Model (CAM), version 5.3.1, was used for this study. CAM is jointly developed at the National Center for Atmospheric Research (NCAR) and various Department of Energy (DOE) laboratories. All simulations are completed within the Community Earth System Model (CESM) framework (version 1.2), which allows for coupling of land, ocean, and ice models to CAM (Hurrell et al. 2013). Simulations utilize the spectral element (SE) dynamical core in CAM (Taylor et al. 1997; Taylor 2011; Dennis et al. 2012). The SE scheme’s mathematical compatibility allows for exact local conservation of mass, energy, and 2D potential vorticity (Taylor and Fournier 2010). The model’s horizontal discretization is built upon a cubed-sphere grid (Dennis et al. 2012). Cubed-sphere grids provide for quasi-uniform mesh spacing over the entire surface of the globe. This eliminates issues such as numerical instabilities that result from the convergence of meridians in polar regions on standard latitude–longitude grids. The default polynomial degree on each cell element is three (cubic polynomials), which leads to fourth-order spatial accuracy (Taylor and Fournier 2010). CAM-SE uses a Runge–Kutta time discretization with a floating Lagrangian coordinate on hybrid sigma-pressure surfaces in the vertical. Since spectral element techniques are highly localized numerical discretizations, they require minimal communication between processors on massively parallel computer systems. CAM-SE has been shown to scale nearly linearly up to one element per processor (Dennis et al. 2012), making it a potentially attractive choice for high-resolution runs on large computing architectures.

Because CAM-SE is designed for use with unstructured grids and the hydrostatic primitive equations are solved locally on individual elements, it is trivial to perform model integrations on nonuniform grids. All cell elements must be quadrilaterals and conforming (every edge is shared by exactly two elements) (Dennis et al. 2012). Any variable-resolution grid fulfilling these two criteria is, therefore, acceptable for use in CAM-SE (Zarzycki et al. 2014b).

The variable-resolution (V-R) grid used in this study is shown in Fig. 1. The grid is generated using the procedure outlined in Guba et al. (2014). The base resolution is 0.5° (55 km). This resolution is selected to ensure that synoptic systems and other large-scale atmospheric features are adequately resolved (Holton 2004). The refinement over the North Atlantic and eastern Pacific basins is of a factor of 4, making the resolution in the target areas approximately 0.125° (14 km). This resolution...
is finer than most operational global numerical weather prediction currently in use, although it is coarser than the innermost, moving nests of regional models used specifically for forecasting tropical cyclones (Gall et al. 2013).

The default CAM, version 5 (CAM5), physical parameterizations suite is used (Neale et al. 2012). Subgrid physics of particular interest to TC modeling include the model’s planetary boundary layer and moist turbulence schemes (Bretherton and Park 2009), prognostic double-moment microphysics with ice supersaturation (Morrison and Gettelman 2008), and CAM5’s deep convective (Zhang and McFarlane 1995) and shallow convective (Park and Bretherton 2009) parameterizations. Further details about the CAM5 configuration can be found in Neale et al. (2012).

Table 1 contains information about grid configurations used in this study. The V-R grid comprises 52,611 elements per model level. The standard CAM5 30 vertical levels are used. A globally uniform CAM-SE grid at the same 14-km resolution as the fine nest contains 345,600 elements in the horizontal direction. CAM-SE has been shown to scale nearly linearly with both the number of processors (Dennis et al. 2012) as well as with the number of elements in a mesh (Zarzycki et al. 2014a). This suggests that the V-R simulations should require only ~15% of the computational resources of a globally uniform 14-km grid.

CAM-SE is coupled to both an ocean/ice and land model through the CPL7 trigridd coupler within the CESM framework (Craig et al. 2012). The coupler utilizes conservative remapping weights to allow the atmosphere to run in conjunction with more standard ocean and land grids. SSTs and sea ice concentrations are provided on a gx1v6 tripole grid (approximately 1°).

The land model is the Community Land Model (CLM), version 4.0, run in satellite phenology (SP) mode on a 0.23 by 0.31° latitude–longitude grid (Oleson et al. 2010).

b. Model initialization

The model is initialized with analysis from the National Centers for Environmental Prediction’s (NCEP) Global Data Assimilation System (GDAS), which is the same product used in initialization of NCEP’s Global Forecast System (GFS) model. The GDAS/GFS analysis is regridded from a 0.5° × 0.5° latitude–longitude grid (the highest-resolution product available) to the V-R CAM-SE mesh using high-order remaps generated with
the Earth System Modeling Framework (ESMF; Collins et al. 2005). No additional vortex bogusing, relocation, or other TC-specific initialization procedure is done following this remap. The model is initialized at 0000 and 1200 UTC every day between 1 August and 31 October for 2012 and 2013. We were unable to attain atmospheric analyses for five cycles during the forecast period.

Initial conditions that are strictly mapped from one grid to another generally suffer from an unbalanced initial state of the atmosphere. This unbalanced state is a result of vertical and horizontal interpolation, as well as errors in the pressure gradient force near sharp topography, which arise from differences in the model orography. This is particularly problematic for variable-resolution models, which require different degrees of topographic smoothing at different grid spacings to maintain numerical stability (Zarzycki et al. 2015). To minimize these issues, we apply an offline digital filter initialization (DFI).

The technique used in this framework is a forward filter, which is described in Fillion et al. (1995). This is slightly different than the forward-and-back technique commonly used in numerical weather prediction (NWP) applications (e.g., Lynch and Huang 1992). A drawback of forward filter techniques is that the solution remains noisy for the first few hours of the forecast period, an effect eliminated by the use of forward-and-back methods. However, a forward filter was preferred to eliminate the need to adiabatically integrate backward in time and, therefore, to provide a more consistent treatment of diabatic processes.

To filter, each forecast is initially integrated forward for 6 h. Model state variables are output every 450 s (7.5 min), providing 49 data points (with +0 and +6 h inclusive). A Heaviside (unit) step function is applied in Fourier space at each state location to remove high-frequency noise in the model state variables. Using short-term test simulations, we found that using a cutoff period \([T_c \text{ in } \text{Fillion et al. (1995)}] \) equal to 6 h provided reasonable results. The model is then reinitialized from this filtered forecast at +3 h (the central point of the filtering time period) and integrated for the duration of the forecast length. We defer to Fillion et al. (1995) for a full derivation and further discussion of the technique.

An example of sea level pressure at +3 h (from initialization) with and without the digital filtering technique is shown in Fig. 2. Note the large-amplitude gravity waves that have emanated from mountainous regions (e.g., the Andes and Rocky Mountains) in Fig. 2a. These waves result from small imbalances in the surface mass field due to remapping between model states with differing topographical roughness. These waves are filtered out through the DFI procedure, providing a much more homogeneous, numerically stable, and realistic surface pressure field in Fig. 2b.

Sea surface temperatures (SSTs) and sea ice concentrations (SICs) are taken from the NOAA High-Resolution Blended Analysis (Reynolds et al. 2007). Daily average SSTs and SICs for the model initialization date are used and assumed to persist during the entire forecast period. While more accurate simulations could have used temporally varying quantities, this would make comparison with forecast models (which do not have access to future ocean conditions) less robust.

In addition, SSTs are fixed and do not interact with the atmosphere in a coupled sense. It is well known that strong, slow-moving TCs can generate cold wakes through turbulent upwelling and heat extraction via surface fluxes (Price 1981). This effect provides a negative feedback on intensity. However, the vast majority of global operational NWP models use fixed SSTs (e.g., NCEP Environmental Modeling Center 2003; Donlon et al. 2012). The addition of a slab, mixed-layer, or fully dynamic ocean model within global forecasts is a future research target.

The land is initialized by “nudging” an already spunup CLM model state toward one in balance with the atmospheric initial conditions. This is done by first iteratively cycling the model in late July (prior to the actual forecast period), where the model is integrated for a short period of time (24 h) and the corresponding 24-h land forecast is reingested as initial conditions. After this cycling has been done between 15 and 30 times, full forecasts are started. To ensure a balanced land condition, each successive forecast is initialized
with the previous cycle’s 12-h land forecast (e.g., the CLM initial conditions for the 0000 UTC 21 August run are the 12-h forecasted CLM state from the 1200 UTC 20 August simulation).

c. Tracking tropical cyclones and measuring forecast error

The method used to track the tropical cyclones in model forecast output is described in Marchok (2002). The tracker utilizes six primary parameters (850-hPa, 700-hPa, and 10-m relative vorticity; 850- and 700-hPa geopotential height; and mean sea level pressure) to determine a center fix (geographic location) for a storm. An iterative Barnes analysis (Barnes 1964) over progressively finer grids is completed for each of the six primary parameters to produce an initial fix for a specific forecast hour and cyclone. A mean fix is determined from the six parameters, although parameters with position fixes not within an allowable distance from the initial guess (275 km) are not used in the mean fix calculation.

The initial location of the cyclone is defined by the National Hurricane Center’s (NHC’s) Tropical Cyclone Vitals (TCVitals) at forecast initialization. The product includes estimates of cyclone location and intensity at a specific analysis time (Trahan and Sparling 2012). We only track storms observed in the North Atlantic and eastern Pacific as they are the two basins under the NHC’s purview, as well as the regions where the 14-km portion of the V-R nest is situated.

We discuss two ways to quantify forecast performance in this manuscript. The first, and simplest, is forecast
error \((e_f)\). The variable \(e_f\) is measured as the absolute difference between a forecast quantity \((\Psi_f)\) and its corresponding observation \((\Psi_{\text{obs}})\):

\[
e_f = |\Psi_f - \Psi_{\text{obs}}|.
\]  

(1)

Note, that we can also calculate forecast bias with the same formulation as in Eq. (1), but using the relative difference rather than the absolute difference.

Using \(e_f\), forecast skill \((S_f)\) can also be calculated as

\[
S_{f,m} = 100 \left(\frac{e_{f,b} - e_{f,m}}{e_{f,b}}\right),
\]

(2)

where \(S_{f,m}\) is the skill of a particular forecast \((f)\) from a particular model \((m)\), \(e_{f,m}\) is the error of the model, and \(e_{f,b}\) is the error of some baseline predictor, generally a statistically based method. The variable \(e_f\) for storm track is measured as the great circle distance between a forecast storm center and that published in the NHC’s best track database,\(^1\) while \(e_f\) for storm intensity is the absolute difference between a forecast 10-m wind speed and one that was observed, also in the best track database. CAM forecast 10-m winds are derived by taking the wind speed at the lowest model level (approximately 60 m) and correcting it to 10 m using a logarithmic profile with an open sea roughness coefficient (Garratt 1992; Wieringa 1992).

For the remainder of this paper, we normalize skill using CLIPER5 (track) (Aberson 1998) and SHIFOR5 (intensity) (Knaff et al. 2003) as baseline \((e_{f,b})\) forecasts. These products are readily available within the NHC’s a-decks (files containing operational forecasts for a variety of dynamical and statistical models) and are available out to 120 h for each forecast cycle.\(^2\) Note that we use 10-m wind rather than minimum sea level pressure (MSLP) as a measure of TC intensity. We do so because no forecast of MSLP was available for SHIFOR5 in the a-deck files used for the skill calculations (i.e., no \(e_{f,b}\)). Previous work has indicated that small differences may exist between surface wind and pressure-based metrics, due to the fact that maximum wind is generally defined by a localized value whereas MSLP is a system-integrated quantity (e.g., Zhu and Zhang 2006). However, absolute 10-m wind errors and MSLP error (not shown) qualitatively demonstrated highly similar behavior, implying that using MSLP as an intensity metric in this analysis would result in analogous conclusions.

3. Forecast performance of variable-resolution global simulations

a. Variable-resolution CAM performance

Figures 3a and 3b show the average track and 10-m wind skill for all named cyclones (wind speed \(\geq 17.5\) m s\(^{-1}\) at initialization) normalized by CLIPER5 (CLP5) and SHIFOR5 (SHF5) as a function of forecast hour. Included are the V-R (14 km) CAM forecasts (CAM, solid black), as well as a selection of other operational model forecasts for the months of August, September, and October in 2012 and 2013. Storms in both the North Atlantic and eastern Pacific are included. The names and dates of all included storms are listed in Table 2.

As with CLP5 and SHF5, these model forecasts are obtained from the NHC’s a-decks. Models included are the NCEP GFS (GFS; global model, \(\Delta x \approx 28\) km, red), the Canadian Meteorological Center’s (CMC’s) Global Environmental Model (GEM; global, \(\Delta x \approx 33\) km, purple), the Hurricane Weather Research and Forecasting (WRF) Model (HWRF; regional model, \(\Delta x \approx 9\) km inner nest, dark red), and the Geophysical Fluid Dynamics Laboratory’s (GFDL’s) regional forecast model (regional, \(\Delta x \approx 5\) km inner nest, orange). Note, that only track forecasts are reported for the CMC model. Sample sizes are color coded at each lead time. Note that the sample sizes may not be identical because of factors such as missed initialization cycles or the tracker not being able to detect a coherent vortex in the model output. Since these missing cycles exist, the fact that the sample size of the data is large, and various models are run with different integration lengths, we have chosen not to restrict these results to a homogenous sample.

All dynamical models show positive track forecast skill out to the end of CLP5’s 120-h forecasts (Fig. 3a). The evolution of this skill is also similar across all models, with the highest values occurring between 48 and 72 h. In terms of track, CAM falls within the envelope of performance of the other models shown here, and model error at long lead times (beyond the 120-h normalization offered by CLP5) appears reasonable (Fig. 3c) by either assuming natural extrapolation of error or a comparison to the GFS.

With respect to wind forecast skill, CAM performs similar to the global GFS model through approximately 48 h (Fig. 3b). Note that the short-term wind errors are quite large for both CAM and GFS relative to the regional HWRF and GFDL models, highlighting the “spinup” issues when simulating TCs initialized in global models with coarser grid spacings. Beyond 48 h, however, CAM skill begins to decrease. This runs counter to other models that all show a stabilization of skill bounded approximately by 20% and 40% (relative

---

\(^1\) See [http://www.nhc.noaa.gov/data/](http://www.nhc.noaa.gov/data/).

\(^2\) See [http://www.ral.ucar.edu/hurricanes/repository/](http://www.ral.ucar.edu/hurricanes/repository/).
between days 2 and 5. Since Fig. 3b shows the absolute value of the error (counting storms that are forecast too strong and storms forecast too weak equally), we show wind bias in Fig. 3d. This is measured as the average of the difference of the forecast wind speed minus that which was observed. It is clear that the majority of the problems with the forecast skill in CAM occur because of an overprediction of intensity. At day 5, CAM forecasts, on average, storms that are \( 6 \text{ m s}^{-1} \) stronger than those observed, while both regional models (HWRF and GFDL) forecast storms that are approximately \( 2-3 \text{ m s}^{-1} \) too strong. The global GFS model underforecasts 120-h intensity by approximately \( 1 \text{ m s}^{-1} \).

This bias can also be seen in forecast pressure–wind curves shown in Fig. 4. For every forecast, both the model predicted and observed surface pressures and 10-m winds are paired at a specified lead time. Blue, filled dots indicate observed pressure–wind pairs, while red, crossed dots indicate particular forecasts from CAM at the specified lead time. In Fig. 4a both spreads generally overlay one another, meaning 48-h CAM forecasts were very similar to observations. This corroborates the results in Fig. 3b showing the model is producing skillful forecasts with little overall bias in intensity at this lead time. Figures 4b and 4c show the same curves for 120- and 192-h forecasts. Very few storms during the 2012–13 hurricane seasons in the North Atlantic and east Pacific reached surface pressures below 930 hPa, whereas CAM forecasted numerous storms below this threshold at these longer lead times (red markers in the upper-left portion of the panels), highlighting CAM’s tendency to shift the intensity curve toward overly intense storms with the default model configuration.

To isolate storms that were well represented at initialization, metrics for all TCs which were initialized at hurricane strength (\( \geq 33 \text{ m s}^{-1} \)) at model initialization. Normalized forecast skill is plotted for (a) track and (b) 10-m wind for forecasts out to 120 h. (c) Absolute track error out to the full model forecast length is plotted. (d) Wind bias is shown. The numbers of model forecasts are noted as color-coded text aligned with a particular lead time.
Fig. 5. Figure 5a shows that track forecast skill is better at all lead times when simulating hurricanes relative to all cyclones (Fig. 3a), implying that stronger storms are better represented in models and generally moved more predictably than weaker cyclones during the study period. Figures 5b and 5d also show that the “runaway” intensity errors associated with CAM in Figs. 3b and 3d are not as prevalent with stronger storms. This implies that CAM’s intensity prediction struggles are associated with unrealistic strengthening of weaker storms during the forecast period.

b. Variable-resolution versus unrefined grids

A complimentary set of globally uniform low-resolution 55-km (0.5°) simulations without the refined patch was completed. All forcings for these forecasts were the same (identical SSTs, ice fraction, and land initial conditions). This is the same resolution as the
background resolution in the V-R simulations. As shown in Table 1, this setup contains significantly fewer grid cells. While the physical parameterization time step was equal to the control V-R run ($dt_{phys} = 1800$ s), the dynamics were run with a longer time step due to a less restrictive Courant–Friedrichs–Lewy (CFL) condition.

All other aspects of the CAM configuration remain identical.

Previous work with CAM has demonstrated that storms that are more similar physically to observed TCs are generated as horizontal resolution is increased (e.g., Reed and Jablonowski 2011; Reed et al. 2012; Zarzycki

---

**FIG. 4.** Pressure–wind curves for CAM forecasts at (a) +48, (b) +120, and (c) +192 h from initialization time. Forecasted surface pressure–10-m wind pairs are shown in crossed red circles, observed pairs are indicated by filled blue circles.

**FIG. 5.** As in Fig. 3, but only including storms that were at least hurricane strength ($\geq 33$ m s$^{-1}$) at initialization.
and Jablonowski 2014). An example of the structural differences arising from resolution differences in the forecast configurations is shown in Fig. 6. As resolution is increased because of refinement, the vorticity core of Hurricane Leslie (+120-h forecast, initialized at 0000 UTC 2 September 2012) becomes much stronger and possesses more horizontal structure (Fig. 6e) when compared to the unrefined 55-km forecast (Fig. 6a). Similar results are seen for both 850-hPa horizontal wind and precipitation, with the 14-km simulation generating stronger winds, a much tighter storm core, and both a low-precipitation eye as well as spiral rainbands branching off the central dense overcast region at the core of the cyclone (Figs. 6f,g). The 55-km grid shows a weaker, broader TC that still retains a central region of intense precipitation, but lacks the aforementioned features seen in the V-R simulation (Figs. 6b,c). Last, the higher resolution of the V-R simulation allows for more intense updrafts at the 500-hPa level, complete with a corresponding downdraft in the eye region, as well as surrounding branches of vertical motion associated with the spiral rainbands seen in the radar reflectivity panel (Fig. 6h). The unrefined simulations show a larger, weaker peak in vertical motion with only one band of ascent present around the storm core (Fig. 6d). While the vast majority of work regarding TC structural sensitivity to horizontal resolution has been done primarily with limited-area models at or below 10-km grid spacings, the dynamical response seen here, while at coarser resolution, qualitatively agrees with those studies (e.g., Fierro et al. 2009; Gentry and Lackmann 2010; Manganello et al. 2012; Sun et al. 2013; Jin et al. 2014).

Figure 7 shows the same metrics as Fig. 3. The V-R simulation (CAM, $\Delta x \sim 14$ km) is shown in black while the unrefined simulation (CAM-UNI55, $\Delta x \sim 55$ km) is marked by the dashed green line. Since the initial conditions are identical, we have removed all intensity restrictions on storm selection, and prenamed TCs in the NHC a-decks, such as tropical depressions, are included. Additionally, the samples for the remainder of this manuscript are homogenous because of the ability to control which model cycles were completed using CAM.

Investigation of the track skill indicates that both configurations follow roughly the same pattern with peak skill occurring at day 3. In Figs. 7a and 7c, the skill (error) of the V-R track forecast is higher (lower) than the unrefined case at all lead times, indicating that the
increased resolution directly leads to a more accurate forecast. Since both models are initialized with identical states and use identical subgrid physical parameterizations, this implies that the grid spacing is the key driver in this behavior.

In the early portion of the forecast period (0–96 h), the V-R simulation also outperforms the unrefined grid with respect to forecasted intensity. However, beyond 96 h, both the skill and bias are better for the unrefined grid. As previously seen, Fig. 7d indicates that this arises because of the overintensification of storms in the V-R forecasts at long lead times. The unrefined simulations likely possess more skill in this case because the lack of dynamic resolution (coarser grid spacing) helps offset the high bias seen in the V-R wind speeds.

Table 3 contains two measures further comparing the two simulations. The frequency of superior performance (FSP) is shown as a fractional value for both track and intensity forecasts for all lead times during the 8-day forecasts. Forecasts are first paired by initialization time and storm. Track and wind errors are compared at each lead time for each storm forecast to determine which model had a lower error (superior performance). If the difference in error between the two models was less than 10% of the average error of the control V-R simulation for that lead time, neither model was deemed superior for that forecast. Additionally, black triangles indicate whether the difference in means seen in Fig. 7 were statistically significant. Statistical significance is assessed using a standard bootstrap technique (Efron and Gong 1983). Random samples (allowing for replacement) are generated from an array of the differences in error between two models for all forecasts. Here, we resample 10 000 times and deem a result statistically significant if both the low and high values of the 80% confidence interval lie on either side of zero (zero representing no difference in error). For example, for the track forecast at 120-h lead time, the V-R simulation (CAM) performed best 62% of the time, while the unrefined simulation (CAM-UNI55) only performed best 26% of the time. In 12% of the cases, the differences between the forecasts at 120-h lead time were within 10% of the average forecast error for CAM at that lead time and were not counted. In addition, the mean track forecast...
for CAM was better than CAM-UNI55, determined to be significant by the aforementioned bootstrapping technique, as denoted by the black triangle in the CAM row at 120 h.

At all lead times, the refined grid (CAM) produced a higher percentage of more accurate track forecasts than the unrefined grid (CAM-UNI55). The mean difference in this performance was significant at all lead times except 12 h (likely due to the models being initialized with identical conditions) and at both 168 and 192 h (possibly due to the small number of matching forecasts at these longer lead times; 99 and 76, respectively). In terms of intensity, the same behavior observed in Fig. 7 is seen in Table 3. CAM performs better at short lead times, but beyond 120 h, CAM-UNI55 exhibits more intensity skill both in terms of the mean forecast (black triangles) and FSP.

c. Variable-resolution versus equivalent uniform high-resolution grids

An obvious motivating factor for the use of multi-resolution model grids is decreased computational cost when compared to globally uniform high-resolution simulations. Because a globally uniform 14-km simulation took approximately 6 times longer to run than the V-R simulation, only a few (12) simulations were completed to compare the forecast behavior of the refined simulations to that with globally uniform high resolution.

Figure 8 shows the day 3 (+72 h) forecast of 500-hPa relative vorticity for Hurricane Sandy, initialized at 0000 UTC 25 October 2012. The left panel depicts the V-R simulation. The middle and right panels are the forecasts using the globally uniform 14- and 55-km simulations, respectively. It is apparent that the V-R simulation most closely matches the globally uniform 14-km simulation. The vorticity maximum associated with Hurricane Sandy is the same order of magnitude and size. In addition, the higher-resolution grids provide significantly more structure within the simulations. For example, a strong band radiating outward from the southern edge of the central vorticity core is evident in both the left and middle panels. A secondary vorticity band stretches from the south of the vortex, around the western periphery of the circulation, and then eastward into the Atlantic. These structures are nearly identical in the V-R and 14-km simulation, underscoring the matching dynamical performance in the high-resolution nest. Many features seen in the left two panels are not evident in the right panel (uniform 55-km grid), and those that lack the structure and definition seen with the finer grid spacing.

Figure 9 shows a similar analysis for a 5-day forecast of 500-hPa relative vorticity over a wider geographic area. In this case, the primary storm forecasted is Hurricane Isaac, with the model initialized at 0000 UTC 23 August 2012. As in Fig. 8, we see that, even at 120 h, stark similarities exist between the two top panels, implying the model representation of flow within the V-R high-resolution nest is well matched by a traditional uniform simulation. Vorticity structures, such as bands associated with Isaac in the Gulf of Mexico, a frontal boundary draped across the continental United States, and other features over the Atlantic Ocean, are well matched. This illustrates the striking similarity between a uniform high-resolution and refined V-R grid at typical forecast lead times.

While the sample of matching forecasts was too small for a formal statistical comparison, a cursory comparison of mean forecast performance as a function of lead time for both the uniform high-resolution and refined V-R grids also indicates nearly identical performance between the two configurations, especially at shorter lead times. Absolute track errors (10-m wind biases) for forecasted TCs in matching cycles at +24, +48, +72, +120, and +168 h were 156, 277, 345, 260, and 393 km

### Table 3. Frequency of superior performance (FSP) for CAM (14-km V-R control) and unrefined CAM (55-km globally uniform grid) forecasts during the 2012–13 Atlantic tropical season. Statistically significant performance as measured by bootstrapping the mean error at 80% confidence is denoted by a triangle that points toward the model with the highest mean skill. The threshold for superior performance is a difference between forecasts of 10% of the average error of the control simulation (CAM) for that metric and lead time. Forecasts where this threshold is not met are not included in FSP calculations and are why the fractional sums may not equal 1 at a given lead time.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Model</th>
<th>12</th>
<th>24</th>
<th>36</th>
<th>48</th>
<th>72</th>
<th>96</th>
<th>120</th>
<th>144</th>
<th>168</th>
<th>192</th>
</tr>
</thead>
<tbody>
<tr>
<td>Track</td>
<td>CAM</td>
<td>0.54</td>
<td>0.57</td>
<td>0.61</td>
<td>0.64</td>
<td>0.61</td>
<td>0.59</td>
<td>0.62</td>
<td>0.67</td>
<td>0.53</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>CAM-UNI55</td>
<td>0.20</td>
<td>0.19</td>
<td>0.21</td>
<td>0.23</td>
<td>0.20</td>
<td>0.26</td>
<td>0.26</td>
<td>0.22</td>
<td>0.35</td>
<td>0.34</td>
</tr>
<tr>
<td>Wind</td>
<td>CAM</td>
<td>0.70</td>
<td>0.72</td>
<td>0.67</td>
<td>0.66</td>
<td>0.62</td>
<td>0.54</td>
<td>0.45</td>
<td>0.42</td>
<td>0.38</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>CAM-UNI55</td>
<td>0.21</td>
<td>0.22</td>
<td>0.27</td>
<td>0.30</td>
<td>0.30</td>
<td>0.36</td>
<td>0.46</td>
<td>0.51</td>
<td>0.53</td>
<td>0.60</td>
</tr>
</tbody>
</table>
4. Sensitivity experiments

a. Observed SSTs versus climatological SSTs

To investigate the impact of initializing with observed SSTs, we ran a suite of CAM forecasts where ocean temperatures were initialized with climatological means from 1982 to 2001 (CAM-CLIMO). These amount to forecasts where one has knowledge of the atmospheric initial state but no knowledge of the ocean state. Results are shown in Fig. 10 and Table 4.

The results show that model track forecasts initialized with observed SSTs (CAM) show small improvement over those with climatological SSTs (CAM-CLIMO) (Figs. 10a,c). Moreover, intensity error is also relatively similar between the two models (Figs. 10b,d). Short-term intensity forecasts (less than or equal to 72 h) are slightly better with observed SSTs, while the climatological SSTs outperform at longer lead times (greater than 72 h).

Interestingly, Fig. 10d shows that the wind bias is actually reduced at longer lead times in the climatological simulations. This is likely due to the fact that the observed tropical North Atlantic index (TNAI; Enfield et al. 1999), a measure of the average SST between 5.5°–23.5°N and 15°–57.5°W, is higher in the 2012–13 observations than the climatological dataset. The TNAI difference (calculated as the TNAI of the observed SSTs minus the TNAI of the climatological SSTs) for August, September, and October was +0.27°C, +0.51°C, and +0.38°C (2012) and +0.30°C, +0.31°C, and +0.40°C (2013). Therefore, the climatological data forced the simulation with cooler SSTs and acted as a somewhat artificial “brake” on the overintensification of cyclones in the model.
Init: 2012082300, valid: +120h

FIG. 9. As in Fig. 8, but for +120-h (day 5) forecast of Hurricane Isaac. All simulations initialized at 0000 UTC 23 Aug 2012.
Table 4 shows that the observed SSTs produce more skillful track forecasts with a higher frequency, although the FSP is more mixed when it comes to predicting storm intensity. These results would imply that the observed SSTs lead to better track forecasts by impacting the large-scale steering flow rather than more accurately forecasting storm intensity (and, therefore, internal dynamics). However, it is worth noting that the means were not statistically different at any lead time for either of the metrics. This, in conjunction with the results from the previous section, suggests that any impact of SST forcing appears to be much smaller than the impact of changing model resolution.

b. Sensitivity to parameterized convection

To attempt to discern potential mechanisms for the systematic overintensification of TCs in CAM, we conduct a suite of physical parameterization sensitivity simulations. In these simulations, all forecasted TCs were first sorted by their 120-h 10-m wind bias (difference from observations).

### Table 4

**Comparison of the control V-R simulation forced by observed SSTs (CAM) and an alternative V-R simulation forced by climatological mean SSTs (CAM-CLIMO). The metrics are the same as in Table 3.**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Model</th>
<th>12</th>
<th>24</th>
<th>36</th>
<th>48</th>
<th>72</th>
<th>96</th>
<th>120</th>
<th>144</th>
<th>168</th>
<th>192</th>
</tr>
</thead>
<tbody>
<tr>
<td>Track</td>
<td>CAM</td>
<td>0.15</td>
<td>0.30</td>
<td>0.43</td>
<td>0.54</td>
<td>0.53</td>
<td>0.51</td>
<td>0.50</td>
<td>0.59</td>
<td>0.60</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>CAM-CLIMO</td>
<td>0.08</td>
<td>0.05</td>
<td>0.13</td>
<td>0.17</td>
<td>0.23</td>
<td>0.31</td>
<td>0.30</td>
<td>0.24</td>
<td>0.25</td>
<td>0.30</td>
</tr>
<tr>
<td>Wind</td>
<td>CAM</td>
<td>0.24</td>
<td>0.39</td>
<td>0.40</td>
<td>0.45</td>
<td>0.46</td>
<td>0.41</td>
<td>0.47</td>
<td>0.47</td>
<td>0.39</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>CAM-CLIMO</td>
<td>0.23</td>
<td>0.33</td>
<td>0.39</td>
<td>0.39</td>
<td>0.35</td>
<td>0.45</td>
<td>0.42</td>
<td>0.44</td>
<td>0.52</td>
<td>0.49</td>
</tr>
</tbody>
</table>
The 21 worst forecasts where CAM produced a TC significantly stronger than observed were then rerun with four separate modifications to the model configuration.

First, we conduct two sets of simulations testing TC forecast sensitivity to the Zhang–McFarlane (ZM) deep convective scheme in CAM. In the first, the scheme is turned off, forcing the large-scale microphysics and shallow convection routines to represent all rainfall processes within the TC core (CAM-NODEEP). Conversely, we decrease the convective relaxation time scale $\tau$ from its default value of $\tau = 3600$ to $\tau = 900$ s while holding the physics time step ($dt_{\text{phys}}$) fixed at 1800 s (CAM-NEWTAU). This setup has been shown to increase the fraction of total precipitation from the ZM scheme (Williamson 2013). Next is a reduction of the physics time step ($dt_{\text{phys}}$) from 1800 s (30 min) to 450 s (7.5 min) to more closely resemble timings used in operational NWP models (e.g., Jung et al. 2012) (CAM-DT450). Note that the default $\tau$ is also decreased by a factor of 4 ($\tau = 900$ s) such that the $dt_{\text{phys}}/\tau$ ratio (deep convective forcing) remains constant. Last, we run CAM with Cloud Layers Unified by Binormals (CLUBB) as a component of the physical parameterization package (CAM-CLUBB). CLUBB is a high-order turbulence closure scheme that provides a unified treatment of boundary layer turbulence, stratiform cloud microphysics, and shallow convection (Golaz et al. 2002). Therefore, using CLUBB replaces these three sets of parameterizations in the default CAM configuration (Bogenschutz et al. 2012, 2013). The version of CLUBB packaged with CAM, version 5.3.1, and released in April 2013 was used here. All configurations and their notations in this manuscript are summarized in Table 5.

Figure 11 shows the forecast track and wind performance for the control simulation in addition to these four sensitivity configurations. Note that the skill (error) for the control simulation is smaller (larger) than those shown in Fig. 3 because we have selected a subset of the most poorly performing forecasts for this analysis.

The track skill for all sensitivity runs is generally the same. While some configurations produce slightly more skillful forecasts, none were statistically different given the relatively small sample sizes. It is clear that either increasing the frequency with which the physical parameterization package is called or completely turning off the deep convection parameterization in CAM does little to change the behavior of the control simulation. All three setups (CAM, CAM-NODEEP, and CAM-DT450) produce storms that rapidly become too intense relative to observations, leading to large values of negative skill when normalized by SHF5 (Figs. 11b,d).

Conversely, tuning the model to force the deep convection scheme to become more active (CAM-NEWTAU) appears to somewhat mitigate the high-intensity bias, although this configuration still does not provide skillful intensity forecasts (relative to SHF5). Additionally, this improvement is maximized out through 72 h, with forecasted cyclones intensifying at approximately the same rate as the control simulation beyond day 3 until the end of the forecast period.

Reed et al. (2012) showed that the intensity of idealized tropical cyclones in CAM-SE is sensitive to the physics time step. In those simulations, $\tau$ was held constant. Therefore, shorter physics time steps correspondingly resulted in smaller $dt_{\text{phys}}/\tau$ ratios, thereby decreasing the fraction of parameterized deep convective precipitation. This dependence is also shown in Williamson (2013). Results from both studies imply that increasing the strength of the deep convective scheme

<table>
<thead>
<tr>
<th>Name</th>
<th>Grid</th>
<th>$\Delta x$ (km)</th>
<th>Physics</th>
<th>Deep convection scheme</th>
<th>$dt_{\text{phys}}$ (s)</th>
<th>$dt_{\text{phys}}/\tau$</th>
<th>SST</th>
<th>Runs (No.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAM</td>
<td>V-R</td>
<td>14</td>
<td>CAM5</td>
<td>ZM</td>
<td>1800</td>
<td>0.5</td>
<td>Observations</td>
<td>359</td>
</tr>
<tr>
<td>CAM-UNI55</td>
<td>UNI</td>
<td>55</td>
<td>CAM5</td>
<td>ZM</td>
<td>1800</td>
<td>0.5</td>
<td>Observations</td>
<td>355</td>
</tr>
<tr>
<td>CAM-UNI14</td>
<td>UNI</td>
<td>14</td>
<td>CAM5</td>
<td>ZM</td>
<td>1800</td>
<td>0.5</td>
<td>Observations</td>
<td>15</td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAM-CLIMO</td>
<td>V-R</td>
<td>14</td>
<td>CAM5</td>
<td>ZM</td>
<td>1800</td>
<td>0.5</td>
<td>Climatology</td>
<td>361</td>
</tr>
<tr>
<td>CAM-NEWTAU</td>
<td>V-R</td>
<td>14</td>
<td>CAM5</td>
<td>ZM</td>
<td>1800</td>
<td>2</td>
<td>Observations</td>
<td>21</td>
</tr>
<tr>
<td>CAM-NODEEP</td>
<td>V-R</td>
<td>14</td>
<td>CAM5</td>
<td>None</td>
<td>1800</td>
<td>0.5</td>
<td>Observations</td>
<td>34</td>
</tr>
<tr>
<td>CAM-DT450</td>
<td>V-R</td>
<td>14</td>
<td>CAM5</td>
<td>ZM</td>
<td>450</td>
<td>0.5</td>
<td>Observations</td>
<td>34</td>
</tr>
<tr>
<td>CAM-CLUBB</td>
<td>V-R</td>
<td>14</td>
<td>CLUBB</td>
<td>ZM</td>
<td>1800</td>
<td>0.5</td>
<td>Observations</td>
<td>327</td>
</tr>
</tbody>
</table>
(larger $dt_{\text{phys}}/\tau$) allows deep convection to more adequately remove instabilities and supersaturation in column grid boxes. This, in turn, decreases the strength of the adjustment made by the large-scale microphysics. Williamson (2013) showed that situations where the microphysics dominates this adjustment yields strong local condensation, which is not redistributed in the column. Correspondingly, strong updrafts and moisture convergence result. This implies smaller $dt_{\text{phys}}/\tau$ ratios are expected to produce more thermodynamically efficient, and, therefore, intense, TCs, which is in agreement with the results seen here.

The replacement of the planetary boundary layer, macrophysics, and shallow convection schemes in CAM with CLUBB appears to have an even more significant impact on storm errors. In particular, forecasts of TC intensity within this subset of storms are greatly improved. CAM-CLUBB forecasts show a small negative bias in wind speed for the majority of the forecast period, with positive biases occurring at 120 h (Fig. 11d). However, these biases, on average, are 20 m s$^{-1}$ less than those seen using the default physics package for this subset of poorly forecast cyclones.

Given these results, a more rigorous suite of forecasts was completed with the CLUBB configuration to compare model performance. The vast majority of initialization times with the control V-R (CAM) forecasts were also completed with CAM-CLUBB, producing 494 matching cyclones at initialization (473 when normalized by CLP5 and SHF5). Figure 12 and Table 6 show the results of these forecasts compared to the control run. At all lead times, the mean track forecast error for CAM-CLUBB is superior to the default CAM simulations (Figs. 12a,c). However, although the mean performance of CAM-CLUBB was superior with respect to track, in many cases, Table 6 shows that the FSP was slightly greater for CAM. This implies that CAM performed slightly better than CAM-CLUBB by frequency, but in some instances where CAM-CLUBB was superior the difference in error between the two models was large. This is shown by the fact that the standard deviation of track error at specific lead times was higher for...
CAM (not shown). The reasons why the control configuration shows larger variability than CAM-CLUBB is unknown and is a topic of future research. It is worth noting that these means are not significant at the 80% level using the bootstrapping technique, so differences in the means between CAM and CAM-CLUBB track errors remain much smaller than the differences in the means between CAM and CAM-UNI55, for example.

CAM-CLUBB intensity forecasts behave less skillfully than CAM out to 72 h, but surpass it for longer lead times (Fig. 12b). These differences in skill are quite pronounced, as evidenced by the statistically significant differences in the mean (triangles in Table 6). Figure 12d shows that the main reason for CAM-CLUBB’s short-term poor performance is under-intensification (or overly slow intensification) of storms at short lead times. However, this bias improves from 72 h out to the end of the forecasts (192 h), with CAM-CLUBB producing quite skillful intensity forecasts at these lead times. This stands in contrast to the control

![Figure 12](image_url)

**Table 6.** As in Tables 3 and 4, but comparing the V-R control simulation (CAM) with V-R simulations utilizing CLUBB (CAM-CLUBB).

<table>
<thead>
<tr>
<th>Metric</th>
<th>Lead times (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12 24 36 48 72 96 120 144 168 192</td>
</tr>
<tr>
<td>Track</td>
<td></td>
</tr>
<tr>
<td>CAM</td>
<td>0.29 0.35 0.37 0.43 0.47 0.42 0.50 0.52 0.53 0.53</td>
</tr>
<tr>
<td>CAM-CLUBB</td>
<td>0.26 0.35 0.36 0.56 0.36 0.44 0.40 0.38 0.38 0.32</td>
</tr>
<tr>
<td>Wind</td>
<td></td>
</tr>
<tr>
<td>CAM</td>
<td>0.46 0.57 0.59 0.55 0.50 0.47 0.44 0.38 0.30 0.32</td>
</tr>
<tr>
<td>CAM-CLUBB</td>
<td>0.26 0.29 0.33 0.35 0.35 0.42 0.50 0.60 0.58 0.62</td>
</tr>
</tbody>
</table>

OCTOBER 2015 ZARZYCKI AND JABLONOWSKI 4029
CAM simulations, which exhibit an increasing bias out to 192 h.

Figure 13 shows a snapshot of the tangentially averaged radial cross section of radial (inflow/outflow) wind, tangential wind, and vertical pressure velocity of Hurricane Leslie for the V-R control (CAM5, top) and V-R CAM-CLUBB (bottom) forecasts. Shown is the +120-h forecast initialized at 0000 UTC 3 September 2012. The center of the cross section is defined by the storm’s sea level pressure minimum. These provide insight into some of the structural differences resulting from the use of CLUBB versus the default CAM5 physical parameterization suite. Both surface radial inflow (positive values) and upper-level outflow (negative values) are significantly stronger in CAM5 (Fig. 13a) when compared to CAM-CLUBB (Fig. 13d), indicative of a much stronger circulation. The CAM-CLUBB forecast shows a more vertically diffuse inflow core, indicating that near-surface momentum may not be as concentrated in the model’s lowest levels, thereby decreasing the efficiency of sensible and latent heat fluxes from the ocean surface. This is also shown in the tangential wind (Figs. 13b,e), which indicates a higher maximum wind location with CAM-CLUBB. In addition, CAM5 exhibits tangential wind structure that is more similar to traditional, strong, TCs (calm eye, sloped eyewall, near-surface maximum wind), whereas the structure of the CAM-CLUBB TC is less organized, with a radius of maximum wind approximately 3 times greater than that of the control CAM5 simulation. The vertical pressure velocity (Figs. 13c,f) also highlights the intensity bifurcation between the two simulations, with CAM5 showing a deep, penetrating updraft core in the eyewall. Alternatively, CAM-CLUBB shows a much weaker, more tilted updraft region that does not reach into the upper troposphere.

Only one particular case is shown as an example here, but analysis of other storms indicates that these general structural differences are robust across the majority of simulations. While CAM-CLUBB provides more skillful forecasts at longer lead times, it does not exhibit as many hallmark characteristics of TC structure as the CAM5 simulations, making it unclear whether or not the improved forecast skill is a function of CLUBB better representing atmospheric processes or merely suppressing facets underlying the high bias in CAM5 simulations. These differences are quite striking and further analysis is necessary to understand the mechanisms driving this behavior. However, it is clear that replacing only three parameterizations (boundary layer, shallow convection, macrophysics) has led to a dramatic change in storm behavior within the model, and the dynamical
reasons behind this response are a target of ongoing research.

5. Hurricane Sandy

The most prolific tropical cyclone in either the North Atlantic or eastern Pacific during the 2012–13 forecast period was Hurricane Sandy. Sandy was a rare storm that made landfall along the mid-Atlantic coast of the United States just before 0000 UTC 30 October 2012. One factor behind Sandy’s destructiveness was the abnormal trajectory the cyclone took as it traversed the western North Atlantic basin. While storms in the mid-latitudes generally turn eastward due to westerly forcing, Sandy turned back toward the west, moving nearly perpendicular to the New Jersey coast at landfall. This motion was rather unusual, being estimated to occur once every seven centuries (Hall and Sobel 2013).

This unusual motion posed an interesting challenge for forecasters. In particular, numerical model guidance at longer lead times (greater than 72 h) was highly scattered. Additionally, the lead time required for individual models to trend toward a realistic landfall case were widely varied. Some models forecasted recurvature more than a week in advance while others continued to show “out to sea” solutions as late as a few days before landfall, with the two most frequently scrutinized NWP models being the GFS [and corresponding GFS-based ensemble (GEFS)] and the global model from the European Centre for Medium-Range Weather Forecasts (ECMWF; Blake et al. 2013).

Figure 14 shows track forecasts for Hurricane Sandy for the daily 0000 and 1200 UTC cycles between 22 October (8 days from landfall) and 25 October (5 days from landfall). The black solid curve is Sandy’s observed track. The thin, pink curves are individual GEFS ensemble members, with the thicker red curve being the GEFS ensemble mean. The darker red curve is the deterministic GFS forecast. The thin, light blue forecasts are CAM ensemble members. These include globally uniform 55-km and globally uniform 14-km forecasts as well as multiple V-R forecasts. These V-R forecasts included the default control configuration, the configuration using climatological SSTs, the configuration with the 450-s physics time step, and the configuration where the deep convective scheme is omitted. The CAM-CLUBB forecast is shown in green as the subgrid parameterizations were sufficiently different such that the forecast was not considered a physics-perturbed member of the default CAM suite.

The CAM ensemble mean forecast is shown in blue. It is readily apparent that CAM forecasts Sandy’s recurvature into the eastern seaboard of the United States significantly earlier than either the deterministic GFS or GEFS mean. The forecast initialized at 0000 UTC 22 October shows a mean landfall location in coastal Maine 8 days prior to landfall, although the sharpness of the left turn of the storm is not well represented until the following cycle at 1200 UTC 22 October. For all following forecast cycles, CAM is consistent with bringing Hurricane Sandy back toward the coastline as opposed to predicting eastward movement out to the open Atlantic. The skill shown by CAM is similar to that seen in the ECMWF ensemble that also predicted a northeastern United States landfall beginning on 22 October (Magnusson et al. 2014). The GFS/GEFS do not forecast this recurvature until 1200 UTC 24 October and do not cluster on a landfall location from Long Island southward until 1200 UTC 25 October.

We note here that the results demonstrate that using the same initial conditions as the GFS model, CAM is able to forecast this recurvature more than a week before the storm made landfall. This result is robust across an ensemble of simulations utilizing various grids and perturbations to the physics time step and deep convective scheme, as shown by the multimember mean for CAM. The fact that CAM uses the same initial conditions as the deterministic GFS forecast (the DFI results in a slightly different state at +6 h, with measured anomaly correlations between GFS and CAM forecasts being approximately 3%–5%, not shown) strongly implies that model initialization was not the key reason why the GFS was unable to forecast this recurvature during the early stages of Sandy’s evolution.

These results agree with those in Bassill (2014) who used WRF ensembles to demonstrate that differences between the GFS and ECMWF forecast models were dominated by cumulus parameterization. In that study, forecasts initialized with similar initial conditions showed a clear track forecast bifurcation that was delineated by the choice of either the simplified Arakawa–Schubert (GFS) or Tiedtke (ECMWF) cumulus parameterizations. Further work indicates that this difference may be constrained by the deep convective entrainment rate in the simplified Arakawa-Schubert scheme (Bassill 2015). Torn et al. (2015) suggest additional atmospheric sampling to the north of Sandy might have improved forecast skill as well, although their reasoning behind this can also be rooted in model physics. They showed that, within a multimember GFS ensemble, moist parameterizations produced less vertical moisture flux convergence and weaker upward motion in members that pushed Sandy eastward into the Atlantic Ocean. These effects fed back into the large-scale dynamics, eventually producing a less amplified subtropical ridge over the Atlantic than what was observed. These results, as well as those contained in
FIG. 14. Track forecasts for Hurricane Sandy for each 12-h forecast cycle between 0000 UTC 22 Oct and 1200 UTC 25 Oct 2012 for various configurations of CAM as well as the GFS/GEFS models. Light, thin lines are ensemble members [CAM (light blue), GEFS (pink)], dark, bold lines are ensemble means [CAM (blue), GEFS (red)], or single, deterministic runs [GFS (dark red), CAM-CLUBB (green)]. The black solid line is the observed storm trajectory.
this paper, highlight the potential for significant sensitivity in tropical cyclone track forecasts to arise from physical parameterization selection.

6. Computational performance

The primary goal of this study was to investigate the performance of V-R CAM-SE from a forecasting standpoint. Therefore, there was no rigorous effort to control for aspects of the study that may have influenced the computational performance of the model. Nevertheless, some meaningful data can be ascertained from an informal analysis of the timing results. Table 7 shows the median timing statistics for individual forecasts using each model grid as well as the various model configurations for the V-R simulations. Note that the control forecasts utilizing the V-R grid were completed before a significant update to the cluster’s InfiniBand communication links. This resulted in improved performance for all other forecast simulations. Additionally, timing data were not collected for the simulations using climatological SSTs.

We show model forecast days per actual wall-clock day normalized to 1000 central processing units (CPUs) in an attempt to normalize the simulations since different numbers of processors were used for different configurations. The uniform 55-km forecasts easily produced the highest model throughput of all simulations. This is to be expected since the grid contains many less elements (e.g., 16 times fewer than the uniform 14-km grid), but is also integrated with a longer dynamical time step due to a less restrictive CFL constraint than either the V-R or uniform 14-km simulations.

Only considering the forecasts that were completed following the InfiniBand upgrade, the V-R simulations produced approximately 80–90 simulated days per wall-clock day, assuming 1000 CPUs. These numbers differ because of the different physical parameterizations that were used, run-to-run variability driven by node selection, as well as the relative load on the cluster due to other users. Note that these simulations produce approximately 6–6.5 times more throughput than the uniform 14-km simulations. As shown in Table 1, the uniform 14-km grid has approximately 6.5 more grid cells than the V-R simulations. Given the fact that the V-R simulation is constrained to the same time step as the uniform 14-km simulations, these results corroborate previous work that has shown that CAM-SE scales nearly linearly with the number of elements in the mesh (e.g., Dennis et al. 2012; Zarzycki et al. 2014a) and validates that variable-resolution simulations can produce regionally refined forecasts at a cheaper computational cost than models with globally uniform high resolution.

7. Conclusions and future work

In this paper we have developed a framework that allows the Community Atmosphere Model to be run as a short-term forecast model using variable-resolution grids with a digital filter initialization scheme. We demonstrate that increased resolution via regional refinement improves both TC track and intensity forecasts when compared to identically initialized unrefined simulations using CAM-SE. This skill increase likely arises from a combination of improved representation of the large-scale flow, more realistic interactions between TCs and their environment, and more accurate representations of small-scale features, which may provide upscale effects to the synoptic environment. These results demonstrate the potential viability of variable-resolution models as alternatives to either globally uniform or high-resolution limited-area simulations in future NWP applications. Furthermore, simulations show that, for lead times on the order of less than a week, a variable-resolution simulation can produce nearly identical forecast behavior inside the refined nest as to that produced by a globally uniform simulation of the same high resolution. This can be achieved with a much lower computational cost, potentially enabling further horizontal refinement, additional ensemble members, or longer model integrations.

Hurricane track skill in the control V-R runs with the default CAM5 physics is comparable to operational models. This occurs even though, in practice, CAM is primarily used as a multidecadal climate model at significantly coarser resolutions. There has been very little validation or subsequent tuning of subgrid physical parameterizations at or below 0.25° (28 km) grid spacing. These skillful forecasts highlight the importance of the
large-scale flow on TC track, which is likely well represented in CAM. The performance of CAM-SE with respect to TC intensity is worse, however. In the model’s default configuration, the error of 10-m wind at long lead times is approximately the same as the statistical SHF5 model. Examination of the 10-m wind biases show that this results from the model producing, on average, storms that are stronger than observed. This over-intensification was strongest among storms that were weaker at initialization, but appears systematic across the majority of forecasts.

We note that while overall activity was at or slightly above average for the 2012–13 hurricane seasons for the Atlantic and eastern Pacific, only one major hurricane (10-m wind $\geq 50\text{ m s}^{-1}$) occurred in 2013 between the two basins (cf. a 1981–2010 average of seven) (Stewart 2014; Blake 2014; Pasch 2014; Kimberlain 2014). In particular, this implies that the ratio of weak to strong storms was slightly higher during the study period when compared to climatology. It is conceivable that the relative skill of V-R CAM-SE, particularly with respect to intensity, might be different in years with more numerous intense storms. On one hand, stronger storms may highlight deficiencies arising from the inability of the relatively “coarse” grid spacings in this study to resolve processes in the cyclone core, such as eyewall asymmetries and replacement cycles. On the other hand, it is possible that V-R CAM-SE would perform better with additional strong storms given the model’s capability to produce very intense TCs with the control configuration. Additional work is needed to determine if this sensitivity exists beyond what was discussed in section 3a.

Simulations where climatological SSTs are used instead of observed SSTs show little impact on model forecast. Track skill is slightly degraded with climatological SSTs, implying that accurate SST representation helps control cyclone motion through either local dynamics or large-scale steering flow. Intensity skill is slightly improved at longer lead times, although this is likely due to the fact that the climatological SSTs are cooler, which systematically reduces intensity through decreased latent and sensible heat fluxes into the model’s lowest level. It is possible that global models are pushing toward horizontal resolutions that can dynamically support TCs intense enough that the assumptions allowing for the use of prescribed SSTs as surface forcing break down. Work is ongoing to test CAM-SE coupled to slab, mixed layer, and fully dynamic ocean models to investigate whether ocean–atmosphere interactions are playing a major role at horizontal resolutions $O(\Delta 10)\text{ km}$. Parameterization sensitivity studies were completed to attempt to discern potential mechanisms for this intensity bias. Increasing deep convective activity reduced intensity somewhat, implying that the partitioning between large-scale microphysics and convective parameterization in the moist physics may be playing a significant role in controlling TC intensity, a result also seen in Reed et al. (2012). Of all of the sensitivity studies shown here, the use of CLUBB in lieu of CAM5’s default boundary layer, macrophysics, and shallow convection schemes played the largest role in modulating the V-R control setup’s tendency to produce TCs stronger than those seen in observations. Interestingly, forecasts using CLUBB tend to intensify TCs more slowly and only become more skillful than the control CAM5 simulations at lead times beyond 72 h. It is also worth noting that the CLUBB configuration still utilizes the default ZM deep convective scheme with the default convective relaxation time scale ($\tau = 3600\text{ s}$). Sensitivity studies also showed some degree of sensitivity to this convective relaxation time scale $\tau$ when the default CAM5 physics were left in place. This strongly implies nonlinearities in the moist physics are important in driving forecasted TC intensity in CAM. Future work will attempt to explore these interactions more thoroughly, although they highlight just how crucial the choices of these parameterizations are in forecasting TCs in any numerical setup.

Particular focus was paid to forecasts of Hurricane Sandy. Using the same initial conditions as the GFS model, CAM-SE robustly forecasts recurvature of Sandy at lead times of $+8\text{ days}$, similar to the ECMWF model (Magnusson et al. 2014). This stands in contrast to the performance of the operational GFS model. The fact that CAM-SE was initialized with the same analysis used in initializing the deterministic GFS and ensemble GEFS models provides further evidence that model physical parameterizations played the most critical role in forecasted track spread in the week before the storm’s landfall (e.g., Bassill 2014). The fact that many NWP models generally show similar average skill [e.g., see Figs. 3 and 5 as well as Cangialosi and Franklin (2013)] when averaged over multiple seasons, but can produce widely different forecasts for individual storms (e.g., Fig. 14) highlights the sensitivities of the forecast of individual storms to model configurations.

All simulations shown in this paper used the publicly available GDAS analysis for model initialization. Work is ongoing to allow variable-resolution CAM-SE to be used in conjunction with the Data Assimilation Research Testbed (DART; Anderson et al. 2009) for native atmospheric and land data assimilation on variable-resolution grids. This will provide a more physically consistent initial state of the atmosphere and eliminate the need for digitally filtering the initial conditions.
Future work will explore this framework as well as compare model performance with the configuration used here to configurations that ingest different data as part of the initialization process.

While the simulations in this study were not extended below 14-km horizontal grid spacing because of available computing resources and the hydrostatic nature of the dynamical core, the only impediment to completing global, variable-resolution simulations at single-kilometer resolution in a nonhydrostatic framework is the behavior of physical parameterizations that span the various resolutions of the simulations. This manuscript highlights a high bias in forecast intensity with the default CAM5 physical parameterization set; however, this package has not been tuned for NWP-type simulations at high resolutions. In addition, future developments of scale-aware parameterizations are likely to lead to significantly improved performance of models at multiple grid spacings, which should dramatically reduce some of the biases seen in these simulations and make V-R models a more feasible tool for next-generation weather forecasting.

Acknowledgments. The authors thank Mark A. Taylor for assistance in running CAM-SE as well as fruitful discussions regarding model behavior and configurations, Paul A. Ullrich for assisting in grid generation and providing computational support, and Timothy Marchok for providing the operational TC tracker. Some of the forecast statistics were computed with the help of the Model Evaluation Tools (MET) package developed by the Developmental Testbed Center (http://www.dtcenter.org/met/users/index.php). Diana Thatcher, Alan Rhoades, Hilary Weller, as well as three anonymous reviewers provided constructive comments that greatly improved this manuscript. Simulations were primarily completed using Yellowstone (ark:/85065/d7wd3xhc), operated by NCAR's Computational and Information Systems Laboratory and sponsored by the National Science Foundation. The remainder of the simulations were completed using resources maintained by the College of Agriculture and Environmental Sciences at the University of California, Davis. This research was partially supported by the Office of Science, U.S. Department of Energy (DOE), Awards DE-SC0003990 and DE-SC0006684.

REFERENCES


Reed, K. A., and C. Jablonowski, 2011: An analytic vortex initialization technique for idealized tropical cyclone studies in


