

Diagnosing Model Errors from Time-Averaged Tendencies in the Weather Research and Forecasting (WRF) Model

STEVEN M. CAVALLO

University of Oklahoma, Norman, Oklahoma

JUDITH BERNER AND CHRIS SNYDER

National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 27 March 2015, in final form 26 October 2015)

ABSTRACT

Accurate predictions in numerical weather models depend on the ability to accurately represent physical processes across a wide range of scales. This paper evaluates the utility of model time tendencies, averaged over many forecasts at a given lead time, to diagnose systematic forecast biases in the Advanced Research version of the Weather Research and Forecasting (WRF) Model during the 2010 North Atlantic hurricane season using continuously cycled ensemble data assimilation (DA). Erroneously strong low-level heating originates from the planetary boundary layer parameterization as a consequence of using fixed sea surface temperatures, impacting the upward surface sensible heat fluxes. Warm temperature bias is observed with a magnitude $\sim 0.5\text{ K}$ in a deep tropospheric layer centered $\sim 700\text{ hPa}$, originating primarily from the Kain–Fritsch convective parameterization.

This study is the first to diagnose systematic forecast bias in a limited-area mesoscale model using its forecast tendencies. Unlike global models where relatively fewer time steps typically encompass a DA cycling period, averaging all short-term forecast tendencies can require potentially large data. It is shown that 30-min averaging intervals can sufficiently represent the systematic model bias in this modeling configuration when initializing forecasts from an ensemble member that is generated using a DA system with an identical model configuration. However, the number of time steps before model error begins to dominate initial condition (IC) errors may vary between modeling configurations. Model and IC error are indistinguishable in short-term forecasts when initialized from the ensemble mean, a global analysis from a different model, and an ensemble member using a different parameterization.

1. Introduction

The construction of comprehensive weather and climate models is one of the greatest challenges in computational science. Since 2009, the National Center for Atmospheric Research (NCAR) has been performing tropical cyclone (TC) real-time forecasts in the Atlantic basin during hurricane season using the Advanced Hurricane Weather

Research and Forecasting (AHW) modeling system (Davis et al. 2008, 2010; Cavallo et al. 2013). This modeling system uses an ensemble-based data assimilation (DA) method [ensemble Kalman filter (EnKF)] with forecasts using a special version of the Advanced Research (ARW) version of the Weather Research and Forecasting (WRF) Model for TC applications. This ensemble DA approach employs uninterrupted cycling of forecasts using ARW; that is, each DA cycle uses as background a 6-h forecast initialized from the previous analysis. Systematic drifts of the numerical model from the observed atmospheric state hinder the usefulness of such forecasting approaches, and the source of these model drifts remain a persistent and unsolved problem in WRF and in all other numerical weather prediction models.

Each year, the AHW modeling system is frozen during the hurricane forecasting season, and later its performance is evaluated. During the performance evaluations, improvements are devised, tested, and implemented for use in subsequent years. However, over time, numerous upgrades

Denotes Open Access content.

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

Corresponding author address: Steven Cavallo, School of Meteorology, University of Oklahoma, 120 David L. Boren Blvd., Norman, OK 73072-7307.
E-mail: cavallo@ou.edu

occur to both the DA and forecast modeling components that ultimately culminate in complex bias patterns due to multiple physical interactions and feedbacks, making it difficult to identify the sources of bias from standard performance metrics. Here, we adopt a method to diagnose model biases for the AHW modeling system through use of the analysis increment and the model dynamical and physical tendencies from short-term model forecasts.

Data assimilation combines models with observations, and an investigation of the resulting analysis increments can be used for model evaluation. When averaged over a sufficiently large sample of flow patterns, a nonzero mean analysis increment allows one to determine the systematic difference between the model and the observations, and in the case of unbiased observations, points directly to model bias. Hence, for stationary systems, a nonzero mean analysis increment derives from the divergence of the model state from observations through the model forecast tendencies, providing a way to decompose the mean analysis increment into the individual contributions of model bias arising from the dynamical tendencies and physical parameterizations (e.g., Klinker and Sardeshmukh 1992). If an initial atmospheric state is exactly correct, then the short lead-time errors in a forecast initialized from this atmospheric state are attributable to the errors in the forecast model.

The use of short-term model forecast tendencies to diagnose model error is best utilized when the analysis used to initialize a forecast is produced by a DA system using the same model (e.g., Rodwell and Palmer 2007). In these cases, the analysis is referred to as a “native” analysis since the same model is used in both the DA and forecasts, and the use of short-term model forecast tendencies in diagnosing systematic model error is referred to as the “initial tendency” method (e.g., Klocke and Rodwell 2014). The departure of the short-term forecasts from the observed atmospheric state in the early time steps allows for the potential identification of process-level errors, and this growth of errors is referred to as model “spinup” (Rodwell and Palmer 2007). There is often a distinct initial transition where the model parameterizations adjust from initial conditions, and Rodwell and Palmer (2007) found that this adjustment occurred mainly during the first time step.

Centers that perform model simulations that are initialized from their own DA system, such as the European Centre for Medium-Range Weather Forecasts (ECMWF), and the Met Office, routinely use the initial tendency method to diagnose and correct model bias. It has been shown to be particularly effective when implementing model upgrades, such as described by Rodwell and Jung (2008), where the technique demonstrated that a new aerosol climatology introduced a warm low-level temperature bias that enhanced convection. Martin et al. (2010) used this approach to reduce long-term systematic errors in the Met

Office Unified Model tropical circulation and precipitation distribution and the Northern Hemisphere continental surface temperature and moisture biases. The initial tendency method can also be used in climate models, as demonstrated by Kay et al. (2011) with atmospheric analyses generated with the Community Atmosphere Model, version 4 (CAM4), where short-term CAM4 forecasts were used to diagnose the source of unrealistic cloud increases in the model over open Arctic Ocean water.

To better understand the source of model bias in general circulation models (GCMs), Phillips et al. (2004) proposed the idea of running climate models in “weather forecast mode,” meaning that short-term forecasts are initialized from a given analysis or reanalysis based on observations of the atmospheric state. From this perspective, the initial state of the model is considered to be correct, and thus early forecast errors are expected to be associated with errors in the model formulation (e.g., Williams and Brooks 2008; Xie et al. 2012; Williams et al. 2013). Many of the physical processes evolve on fast time scales (such as cloud microphysical processes, radiative processes, or turbulence), such that once these errors are established, the model state becomes a less accurate estimate of the atmosphere. Therefore, from this perspective, the evolution of long-term climate model bias can be studied by examining the model tendencies from short-term “weather” forecasts. However, for climate models, the DA system that provides the initial conditions does not usually use the same model. Thus, in this case, forecasts are initialized from “nonnative” analyses produced with a different forecast model, and the approach to examine the model bias using short-term model tendencies is referred to as the “transpose-AMIP” method since it was adopted in the Transpose Atmospheric Model Intercomparison Project (transpose-AMIP; <http://www.transpose-amip.info>; Williams et al. 2013). However, because forecasts begin from nonnative initial conditions, there is additional error at short lead times that is due to the forecast model’s drift to its own attractor and is referred to as “initial shock” (Klocke and Rodwell 2014). The time scales for which this additional drift occurs is not well understood, thereby yielding considerable uncertainty regarding the appropriate forecast time steps to evaluate short-term forecast tendencies.

The transpose-AMIP method has successfully been applied at centers that do not have DA systems on hand. In these cases, GCM forecasts are initialized from an analysis that is produced from an operational numerical weather prediction model, such that the initial state of the model is derived from a reasonably sophisticated estimate of the true atmospheric state. The Cloud-Associated Parameterizations Testbed (CAPT) has shown that deficiencies in GCM parameterizations often cannot be easily identified by merely analyzing climate

statistics, and that analyzing short-term forecasts in GCMs is a more effective way to identify these weaknesses (e.g., Phillips et al. 2004; Williamson et al. 2005; Williamson and Olson 2007; Hannay et al. 2009; Medeiros et al. 2012). Williamson et al. (2005) compared moisture tendencies averaged over many forecasts from CAM, version 2 (CAM2), with estimates from observations from two intensive observing periods (IOPs) at the Atmospheric Radiation Measurement (ARM) Southern Great Plains site. They found that the largest forecast errors formed from fast processes that occurred within the first 24 h, and that subsequent errors evolved slowly in response to the early errors. While they were able to attribute a substantial contribution of model bias to the convective parameterization during the summer-season IOP, relatively biased initial conditions could not be eliminated as a source leading to the large model tendencies since forecasts were initialized from nonnative analyses. Williamson and Olson (2007) used CAM, version 3 (CAM3), to find that the error patterns were similar to CAM2 with regard to the fast-scale errors evolving within the first 24 h. Decomposition of the individual tendency components led to diagnosing and correcting an issue in the shallow convection scheme, which subsequently alleviated the model bias. Hannay et al. (2009) compared observations from the Eastern Pacific Investigation of Climate Processes in the Coupled Ocean–Atmosphere System (EPIC) IOP to four different GCMs, and found that these models typically produced a planetary boundary layer (PBL) that was too shallow. Short-term model tendencies from five different global GCMs exposed an accumulation of model error from an erroneous positive feedback between cloud layers that were often too thin and boundary layers that were too shallow (Williams et al. 2013).

While the initial tendency approach has been commonly used in global GCMs, it has yet to be applied to limited-area mesoscale numerical weather prediction models. Here, we utilize the initial tendency approach to diagnose sources of model bias in the Advanced Hurricane WRF system. We test this approach during the 2010 North Atlantic hurricane season during a 15-day period of continuously cycled ensemble DA. Section 2 describes the method we employ with this technique. We discuss the evaluation of this approach in section 3, with a summary of the relevant results given in section 4.

2. Methods

a. AHW and the AHW–DART system overview

The AHW Model (Davis et al. 2008) and ensemble DA system using the implementation of the ensemble adjustment Kalman filter (EAKF) in the Data Assimilation Research Testbed (DART) (Anderson et al.

2009) is used in this study. This system is described by Torn (2010) and Cavallo et al. (2013), and the interested reader is directed to those manuscripts for more details. We subsequently refer to this combined modeling and DA system as AHW–DART. Model biases are examined using this system during the 2010 Atlantic hurricane season. Changes implemented for the 2010 simulations used in the present study are provided below.

In the 2010 implementation of this system, real-time ensemble analyses are generated every 6 h from 28 July to 7 November 2010 by cycling a 96-member EAKF system over a limited-area domain with 36-km horizontal grid spacing and $289x \times 174y \times 36z$ grid points, respectively (Fig. 1a). The ensemble is initially generated by drawing random perturbations to the National Centers for Environmental Prediction (NCEP) error covariances from the WRF VAR system (Barker et al. 2004) to create deviations from the ensemble mean for each ensemble member (Torn et al. 2006). The resulting perturbations are then scaled and added to the 36-h NCEP Global Forecast System (GFS) forecast valid at the initial starting time. Automated Surface Observing System (ASOS) stations, ships, buoys, rawindsondes, aircraft dropsondes, cloud motion vectors (Velden et al. 2005), aircraft data from the Aircraft Communication Addressing and Reporting System (ACARS), real-time estimates of TC position and minimum sea level pressure (SLP) produced by the National Hurricane Center (NHC; i.e., working best track or TCVitals), and Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) global positioning system (GPS) refractivity profiles (Anthes et al. 2008) are assimilated. After an analysis is completed on the 36-km grid, a moving nest with 12-km grid spacing ($1000 \text{ km} \times 1000 \text{ km}$) centered on the TC position is generated for each system that NHC designates as an “invest” area or greater in real time by interpolating from the 36-km grid. Each ensemble member is integrated forward to the next assimilation time on this 12-km two-way interactive nest, which follows the TC based on its movement from the NHC advisory position over the previous 6 h for the remainder of its life cycle or until NHC declares that it is no longer tropical. Observations are also assimilated on this domain every 6 h when observations are assimilated on the 36-km grid. In the subsequent analysis, all calculations are performed on the 36-km grid unless otherwise noted. The ensemble members are advanced in time using a third-order Runge–Kutta method with fifth-order horizontal advection within ARW version 3.1.1 (Skamarock et al. 2008) with the following physics choices: Rapid Radiative Transfer Model (RRTM) longwave radiation (Mlawer et al. 1997) employing an upper-boundary longwave radiation correction (Cavallo et al. 2011), the National Aeronautics and Space Administration

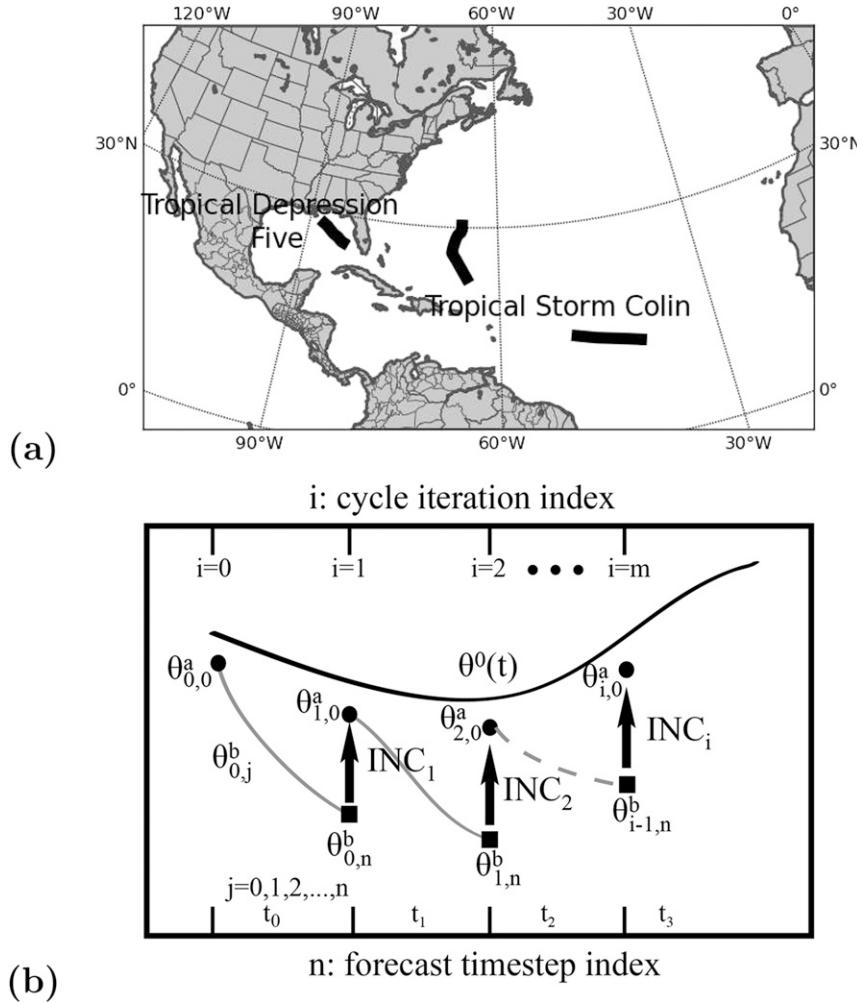


FIG. 1. (a) Parent 36-km domain and NHC tropical cyclone advisory tracks from 1 to 15 Aug 2010. (b) Cartoon of the analysis-forecast cycle for potential temperature θ . The data assimilation cycle is indicated by subscript index i , ranging from cycle number 0 to m , while model time step is denoted by index j , ranging from 0 to n . Observations are indicated by the thick, black line labeled $\theta^0(t)$. The gray line denotes an ensemble mean forecast evolution in potential temperature, where the initial state is equal to the ensemble mean values. Superscripts a and b refer to values valid at the analysis (after data assimilation) and background forecast times (before data assimilation), respectively. Potential temperature ensemble mean analysis increment at data assimilation cycle i is indicated by INC_i . [Adapted from Rodwell and Palmer (2007).]

(NASA) Goddard shortwave radiation (Chou and Suarez 1994), WRF single-moment 5-class microphysics scheme (WSMS; Hong et al. 2004), Kain–Fritsch cumulus convection (Kain and Fritsch 1990), Yonsei University (YSU) planetary boundary layer (Hong and Pan 1996), thermal diffusion surface physics (Dudhia 1996), and Monin–Obukhov surface layer physics (Paulson 1970; Dyer and Hicks 1970; Webb 1970).

Model bias is apparent in the 2010 real-time AHW system when comparing radiosonde, dropsonde, and ACARS observations to 6-h AHW forecasts (Fig. 2). Model forecast temperatures are warmer than radiosonde

observations throughout the troposphere, except near the surface (Fig. 2a). A similar warm tropospheric temperature bias is present with respect to ACARS and dropsondes below ~ 400 hPa (Figs. 2c,e) and has a peak magnitude of ~ 0.5 K near the 700-hPa level. Above 400 hPa, the forecast biases relative to ACARS and radiosondes have opposite signs. This points to observational errors either in ACARS or in the radiosonde measurements [with bias in ACARS most likely; see Ballish and Kumar (2008)], whatever the actual AHW forecast bias may be. The following section outlines a method to diagnose the sources of model

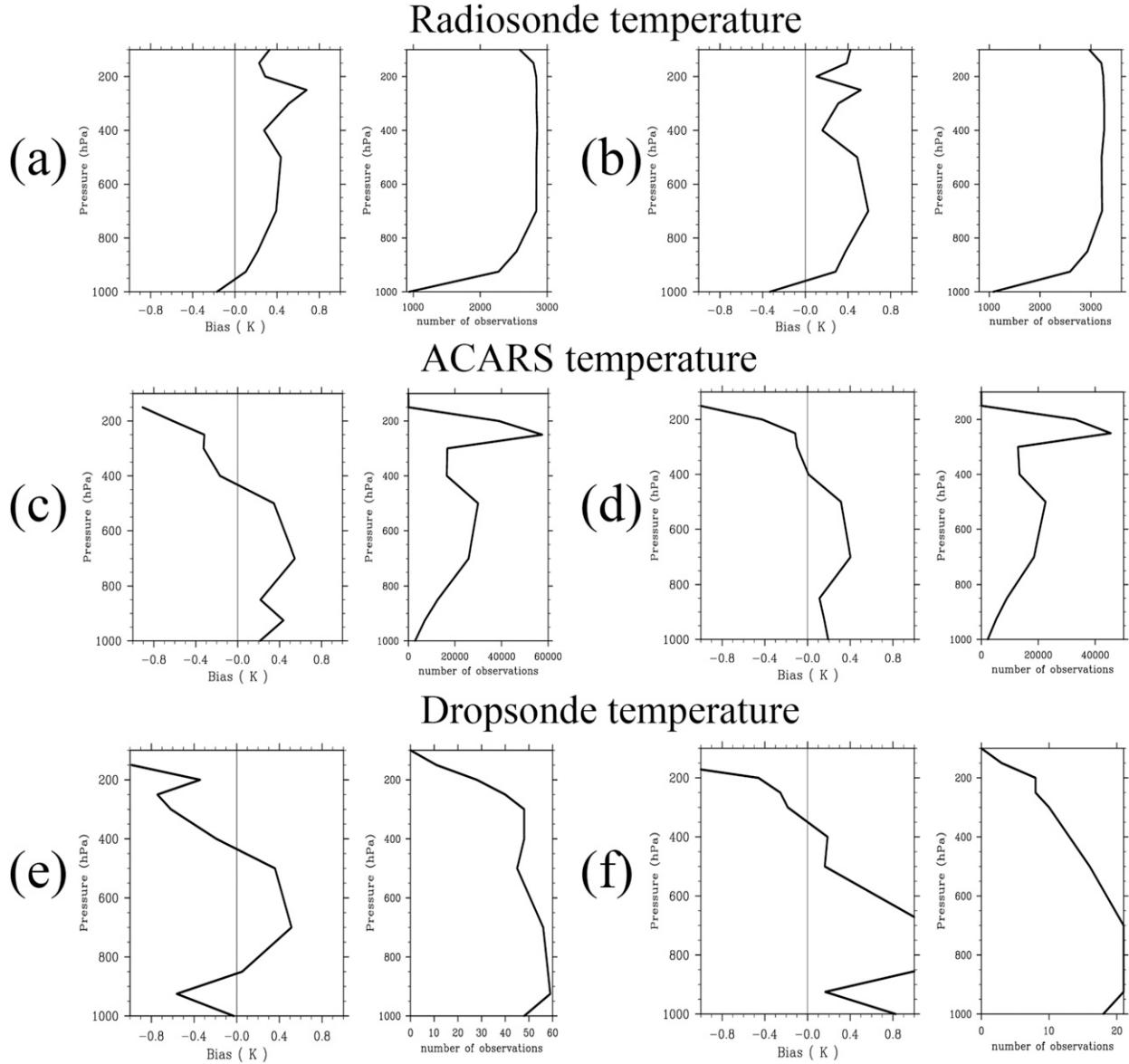


FIG. 2. Average AHW 6-h forecast bias (forecast – observation) with respect to (a),(b) radiosonde; (c),(d) ACARS; and (e),(f) dropsonde observations averaged over the (a),(c),(e) entire season and (b),(d),(f) period 0000 UTC 1 Aug–1800 UTC 15 Aug.

biases in the absence of systematic observation biases. Given that there are possible observation errors around 250 hPa, we focus on diagnosing the broad tropospheric warm temperature biases described above at pressure levels greater than 400 hPa.

b. Time-averaged tendencies

A schematic diagram of a cycled DA system following Rodwell and Palmer (2007) is shown in Fig. 1b for the potential temperature (θ) state variable. The schematic represents a situation for which there is a model bias in potential temperature, illustrated in the divergence of model forecasts ($\theta_{i,j}^b$) from the observed state $\theta^o(t)$

during model time steps $j = 0$ to $j = n$ beginning at each i th assimilation time. At each i th assimilation time, new observations with sufficiently small observation error shift model forecast values at model time step n ($\theta_{i-1,n}^b$) toward the observed state [$\theta^o(t)$] to form a new model analysis ($\theta_{i,0}^a$). The difference between the model analysis and background model forecast values valid at the same time ($\theta_{i,0}^a - \theta_{i-1,n}^b$) is the analysis increment (INC $_i$). If both the observations and model are unbiased, then when averaged over many forecasts, INC $_i$ will be zero, and the observed and modeled values are identical. If observation error is random, then over many samples, the mean observation error is zero. In

this case, a nonzero analysis increment measures the errors in the model due to the model's representation of the atmosphere's physical or dynamical processes. From Fig. 1b, the analysis increment and mean model tendencies at the i th DA cycle are as follows:

$$\text{INC}_i = \theta_{i,0}^a - \theta_{i-1,n}^b \quad \text{and} \quad (1)$$

$$\overline{\dot{\theta}_i^b} = \frac{\theta_{i,n}^b - \theta_{i,0}^a}{\Delta t_i} = \frac{1}{n} \sum_{j=0}^n \dot{\theta}_{i,j}^b, \quad (2)$$

where θ_i^a and θ_i^b are the model's analysis and background potential temperature of the i th DA cycle, respectively; $\overline{\dot{\theta}_i^b}$ is the mean potential temperature model forecast tendency averaged over n model time steps on the i th DA cycle; and Δt_i is the length of DA cycles. Since $\Delta t_i = \text{constant}$, we define $\Delta t_i = \Delta t_{\text{da}}$. Summing over m DA cycles and separating the endpoints, (1) becomes

$$\sum_{i=1}^m \text{INC}_i = \sum_{i=1}^{m-1} (\theta_{i,0}^a - \theta_{i-1,n}^b) + \theta_{m,0}^a - \theta_{m-1,n}^b. \quad (3)$$

Adding and subtracting $\sum_{i=1}^{m-1} \theta_{i,n}^b$ into (3), solving for $\theta_{i,n}^b$ in (2), and substituting its summation into (3) expands the analysis increment to

$$\sum_{i=1}^m \text{INC}_i = -\Delta t_{\text{da}} \sum_{i=1}^{m-1} \overline{\dot{\theta}_i^b} + \sum_{i=1}^{m-1} (\theta_{i,n}^b - \theta_{i-1,n}^b) + \theta_{m,0}^a - \theta_{m-1,n}^b. \quad (4)$$

Noting that

$$\begin{aligned} \sum_{i=1}^{m-1} (\theta_{i,n}^b - \theta_{i-1,n}^b) &= (\theta_{1,n}^b - \theta_{0,n}^b) + (\theta_{2,n}^b - \theta_{1,n}^b) \\ &+ (\theta_{3,n}^b - \theta_{2,n}^b) + \cdots + (\theta_{m-1,n}^b - \theta_{m-2,n}^b) \\ &= -\theta_{0,n}^b + \theta_{m-1,n}^b, \end{aligned} \quad (5)$$

then substitution of the above into (4) leaves us with

$$\sum_{i=1}^m \text{INC}_i = -\Delta t_{\text{da}} \sum_{i=1}^{m-1} \overline{\dot{\theta}_i^b} + \theta_{m,0}^a - \theta_{0,n}^b. \quad (6)$$

Noting from (2) that $\theta_{0,n}^b = \Delta t_{\text{da}} \overline{\dot{\theta}_0^b} + \theta_{0,0}^a$, then substitution into the above yields the form of the analysis increment used subsequently:

$$\begin{aligned} \sum_{i=1}^m \text{INC}_i &= -\Delta t_{\text{da}} \sum_{i=0}^{m-1} \overline{\dot{\theta}_i^b} + \theta_{m,0}^a - \theta_{0,0}^a \\ &= -\Delta t_{\text{da}} \sum_{i=0}^{m-1} \overline{\dot{\theta}_i^b} + \text{analysis drift}, \end{aligned} \quad (7)$$

where analysis drift $= \theta_{m,0}^a - \theta_{0,0}^a$ is the difference between analysis values of θ . The last term on the right-hand-side of (7) is associated with the natural evolution in θ that occurs over the duration of the analysis period. Furthermore, Klinker and Sardeshmukh (1992) and Rodwell and Palmer (2007) showed that the systematic model bias can be established by using only the mean of the *initial* forecast tendencies from the first few model time steps, meaning that only short forecast segments are needed to determine the systematic model bias. Given the above relationship between the mean initial tendencies and the analysis increment, we subsequently refer to this method as the mean initial tendency and analysis (MITA) increment method for brevity.

In the WRF Model, the mean potential temperature forecast tendency over a short term forecast of time Δt_i is the mean of the sum of the potential temperature forecast tendencies from the model's dynamical and physical numerical schemes:

$$\overline{\dot{\theta}_i^b} = \frac{1}{n} \sum_{j=0}^n (\dot{\theta}_{\text{dynamics},j} + \dot{\theta}_{\text{physics},j}) \quad \text{and} \quad (8)$$

$$\overline{\dot{\theta}_i^b} = \frac{1}{n} \sum_{j=0}^n (\dot{\theta}_{\text{dynamics},j} + \dot{\theta}_{\text{radiation},j} + \dot{\theta}_{\text{latent heating},j} + \dot{\theta}_{\text{cumulus},j} + \dot{\theta}_{\text{pbl},j}), \quad (9)$$

where $\dot{\theta}_{\text{dynamics},j}$, $\dot{\theta}_{\text{radiation},j}$, $\dot{\theta}_{\text{latent heating},j}$, $\dot{\theta}_{\text{cumulus},j}$, and $\dot{\theta}_{\text{pbl},j}$ are the potential temperature forecast tendencies at forecast step j from the dynamics, radiation scheme, grid-scale microphysical scheme, subgrid-scale convective heating parameterization scheme, and the PBL scheme (including model mixing and diffusion processes), respectively; and $\dot{\theta}_{\text{physics},j} = \dot{\theta}_{\text{radiation},j} + \dot{\theta}_{\text{latent heating},j} + \dot{\theta}_{\text{cumulus},j} + \dot{\theta}_{\text{pbl},j}$ is the net physics forecast tendency.

In this study, we initialize a series of short-term forecasts from AHW initial conditions (ICs) and use the same model to compute the mean initial forecast tendencies, $\overline{\dot{\theta}_i^b}$, and average them over m DA cycles. Short-term forecasts in this study are considered forecasts with lead times that are shorter than the DA cycling period of 6 h, since the analysis increment is the difference between 6-h forecasts and analyses. We also retain the individual tendency components given by (9) in order to diagnose the component of the model bias in the AHW system. Model θ tendencies are output at every model time step of $\Delta t = 200$ s. Because of the large amount of data generated by this frequent output interval, a single ensemble member is arbitrarily chosen as the ICs for all short-term forecasts and a 15-day subset of the entire

AHW cycling period is examined. We choose to perform these 60 short-term AHW forecasts from 0000 UTC 1 August to 1800 UTC 15 August 2010. Seasonal drift is included in our computations in the analysis drift term in (7). A starting time early in the DA cycling period, while allowing several days for the system to have equilibrated away from the GFS initial conditions on 28 July 2010, is chosen here. To summarize results for the entire domain, we spatially average quantities over every grid point on each vertical level, including all land and water points, and excluding the four nearest grid points to the model boundary. A comparison of the all terms computed from (7) during this period shows that the independent computation of the terms in the right-hand side of (7) for the 60–90-min forecast segments over all DA cycles closely matches the analysis increment, thus verifying our implementation of this method (Fig. 3a).

We average short-term forecast tendencies in 30-min intervals, with the starting time of each interval evenly spaced by exactly 1 hour. Two different stages in model adjustment can be highlighted when comparing the initial 0–30-min forecast tendencies with the 60–90-min forecast tendencies; these results will be discussed in greater detail in section 3. This sensitivity of a temporal averaging period of the short-term forecasts was evident in Bretherton et al. (1995), where bias patterns differed between 1- and 5-h forecasts. While appropriate averaging intervals have not yet been determined, and very likely depend on the model, routine use of this technique at ECMWF generally takes averages of the tendencies over the entire DA period (e.g., Rodwell and Jung 2008). Here, we also investigate how the tendencies vary over the assimilation interval, for two reasons. First, as will be seen below, there are nontrivial variations of the time-averaged tendencies within the assimilation window that are important to understand. Second, our implementation of the MITA increment method involves saving model output at each time step before averaging, which requires considerable disk space and led us to consider shorter averaging intervals.¹

We establish statistical significance of $\dot{\theta}$ as follows. Forecast tendencies of the chosen 30-min interval described above are averaged for each DA cycle to obtain 6-h average tendencies over the 15-day subset period of evaluation so that we have exactly 60 evenly spaced time series data points at each vertical level. Variations that are due to the diurnal cycle are then removed by computing running 24-h averages over the tendencies. The tendencies are resampled 10 000 times with replacement

¹ After submission of this paper, we changed our implementation to accumulating the temporal averages during the model integration.

and accounting for serial correlation using a moving-blocks bootstrap resampling method where each time series is divided evenly into 4 blocks (Wilks 2011, section 5.3). We use 70% confidence intervals following Rodwell and Palmer (2007). All results discussed are at levels for which the mean forecast tendencies are statistically distinguishable from zero following this procedure and are shown in Fig. 3a.

3. Evaluation over the Atlantic hurricane basin during the 2010 season

We choose the period 0000 UTC 1 August–1800 UTC 15 August 2010 to implement the MITA increment method because of the relative inactivity in TCs and, thus, the lack of anomalous TC effects. During this period, Tropical Storm Colin forms out of an African easterly wave at 1500 UTC 2 August, and is downgraded below tropical depression status by NHC at 2100 UTC 3 August (Fig. 1a). Colin reforms on 5 August and quickly reaches its maximum wind of 60 kt ($\sim 31 \text{ m s}^{-1}$) and minimum SLP of 1005 hPa. Tropical Depression Five forms and decays in the Gulf of Mexico at 2300 UTC 10 August–2100 UTC 11 August. Model bias within this period is similar to the bias of the full season (Fig. 2).

a. Application to forecasts initialized from an individual ensemble member

Time-averaged forecast tendencies for forecasts initialized from ensemble member 31 during the first 30 min of all simulations show little similarity to the model bias with respect to observations (cf. Figs. 2 and 3b). In particular, there is a broad local maximum in θ bias $\sim 700 \text{ hPa}$ with respect to observations. In the forecast tendencies, both the dynamics and physics components exhibit local maxima or minima around 450 hPa, which corresponds primarily to the large time-step dynamics² tendencies and to the tendencies from the microphysics and convective heating parameterizations (Fig. 4).

Given the short time steps in our mesoscale limited-area model (LAM) system in comparison to typical

² The WRF Model employs a split-explicit third-order Runge-Kutta (RK3) integration scheme (Wicker and Skamarock 2002). This consists of dividing the model physics and dynamics into slow and fast components. For example, advection, Coriolis, and diffusive tendencies are accumulated into a tendency referred to as “large time-step dynamics” and integrated using a time step Δt . Using these tendencies, fast components such as the acoustic and gravity wave modes are integrated using a small time step where $\Delta t_s = \Delta t/N$, where N is the number of small time steps; we refer to these accumulated fast tendencies as “small time-step dynamics.”

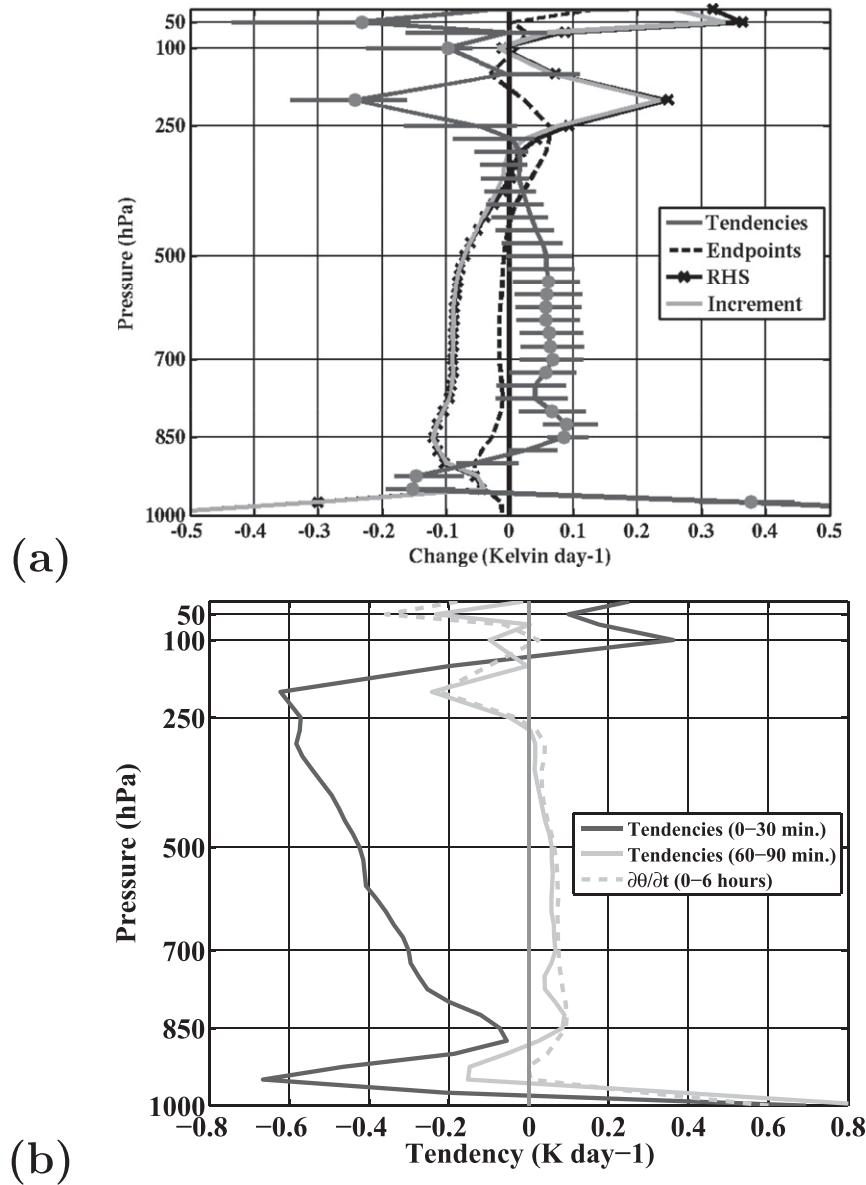


FIG. 3. Time-averaged vertical profiles of the 60–90-min forecast tendencies (“tendencies”; dark gray line), analysis drift (“endpoints”; black dashed line), right-hand side of (7) (“RHS”; black “ \times ”), and analysis increment (light gray line) for forecasts initialized with ensemble member 31. Error bars corresponding to the forecast tendencies represent the 70% confidence intervals as described in the text. Levels at which the tendencies are statistically different from zero are denoted by gray circles. (b) Time-averaged vertical profiles of the net tendencies averaged from 0 to 30 min (black) and from 60 to 90 min (solid light gray). The change in theta between the initial time and hour 6 is shown for comparison with a light dashed gray line. Values are temporal averages over the period 0000 UTC 1 Aug–1800 UTC 15 Aug 2010 and are spatially averaged over the entire domain, but not including points within 4 grid points of the domain boundary.

global GCMs, it is not immediately obvious which time steps within our 6-h cycling period best encapsulates the net model bias. To examine this aspect, we extend our analysis to the 60–90-min forecast tendencies. A comparison of the 60–90-min tendencies with the 0–30-min

tendencies reveals that the later forecast tendencies more closely resemble those from the net 6-h bias and the mean model tendency averaged over the entire 6-h DA cycling period (cf. Figs. 3 and 2). The 60–90-min forecast tendencies exhibit a pronounced maximum

near the surface, and a vertically broad net positive tendency in the middle troposphere from ~ 875 hPa upward to ~ 275 hPa. For the subsequent discussion, we refer to the tendency patterns evident in Fig. 3 by locations where local maxima or minima on vertical isobaric levels are apparent: the near surface (1000 hPa) or the middle troposphere (275–875 hPa).

Although the dynamics and physics tendencies largely cancel each other, a net tendency imbalance remains with positive tendencies in the middle troposphere, strongly positive near-surface tendencies, and negative near-tropopause tendencies (Figs. 3 and 5a). The mean dynamics tendencies are negative at all levels below 250 hPa, with large contributions from both the large and small time-step dynamics tendencies (Fig. 5c). The positive near-surface tendencies are dominated by a mean positive tendency of over 4 K day^{-1} in the PBL scheme (Fig. 5e). In the middle troposphere, the convection scheme contains the largest positive tendencies, with mean heating rates exceeding 2 K day^{-1} . The maximum physics tendencies are located around 400 hPa, where there are substantial contributions from both the convection and microphysics schemes. The convective heating parameterization appears to frequently trigger deep convection, followed by evaporative cooling derived from the grid-scale microphysical heating parameterization as precipitation falls below the cloud level. Radiative cooling rates are around -1.2 K day^{-1} with peak cooling near the tropopause around 225 hPa. Weaker radiative cooling rates above the tropopause are a reflection of lower concentrations of water vapor since water vapor is a strong absorber in longwave bands (e.g., Cavallo et al. 2011). The peak in near-tropopause diabatic cooling is likely a combination of maximum latent heating remaining below the tropopause, and enhanced longwave radiative cooling rates above clouds and near the tropopause where water vapor concentrations decrease rapidly with height.

The greatest changes in forecast tendencies between the 0–30- and 60–90-min time steps are in the physics tendencies, with a mean increase (decrease) at all vertical levels below (above) 125 hPa (Fig. 5b). Differences in large and small time-step tendencies have opposing signs, yielding an almost complete cancellation in the net dynamics changes (Fig. 5d). Increases in upper-tropospheric physics tendencies derive from increases in both the microphysical and convective heating tendencies (Fig. 5f). In the lower troposphere, there is a large increase in PBL heating of 0.6 K day^{-1} with the maximum increase at the surface; there is a corresponding increase in convective heating of the same magnitude ~ 925 hPa. The large increase in the physics tendencies relative to dynamics tendencies is consistent with a

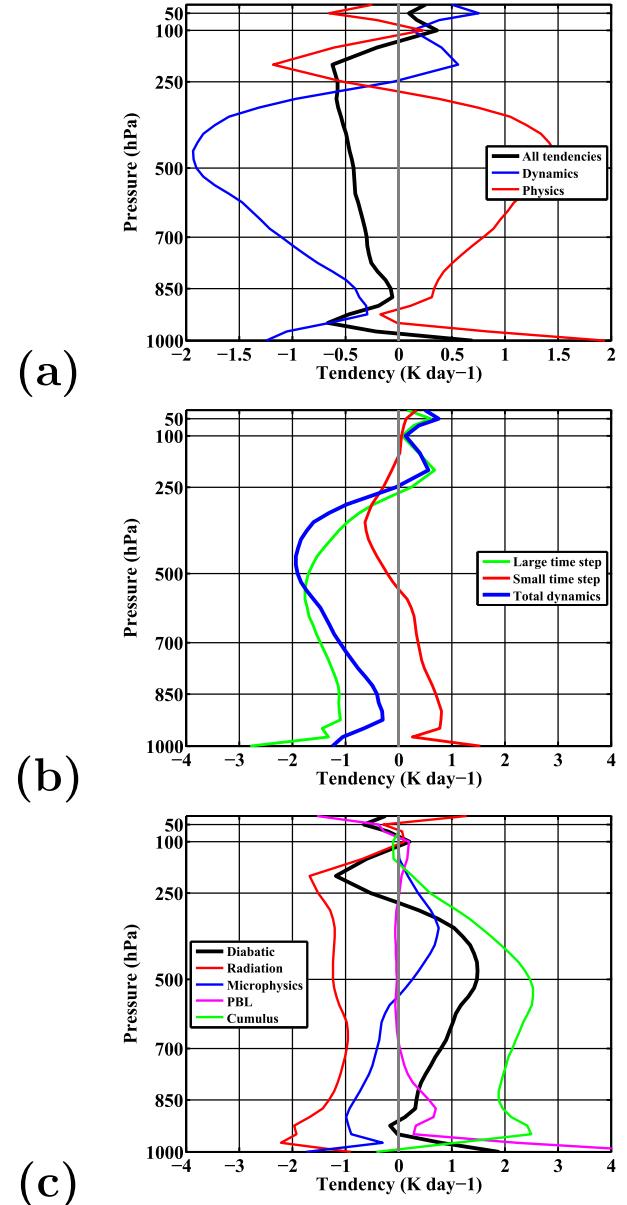


FIG. 4. Comparison of time-averaged vertical profiles of (a) net (black line), dynamics (blue line), and physics (red line) tendencies; (b) net dynamics (blue line), large time-step dynamics (green line), and small time-step dynamics (red line) tendencies; and (c) net diabatic (black line), radiation (red line), microphysics (blue line), planetary boundary layer (magenta line), and cumulus (green line) tendencies from forecasts initialized from the ensemble member 31. Tendencies are averaged over the 0–30-min period of each simulation over the period 1–15 Aug 2010 and over the entire domain, but not including points within 4 grid points of the domain boundary.

conclusion of Rodwell and Palmer (2007) that the behavior of the first time step is different from those in subsequent time steps due to an initial adjustment of the model's physical parameterizations. Given their use of a

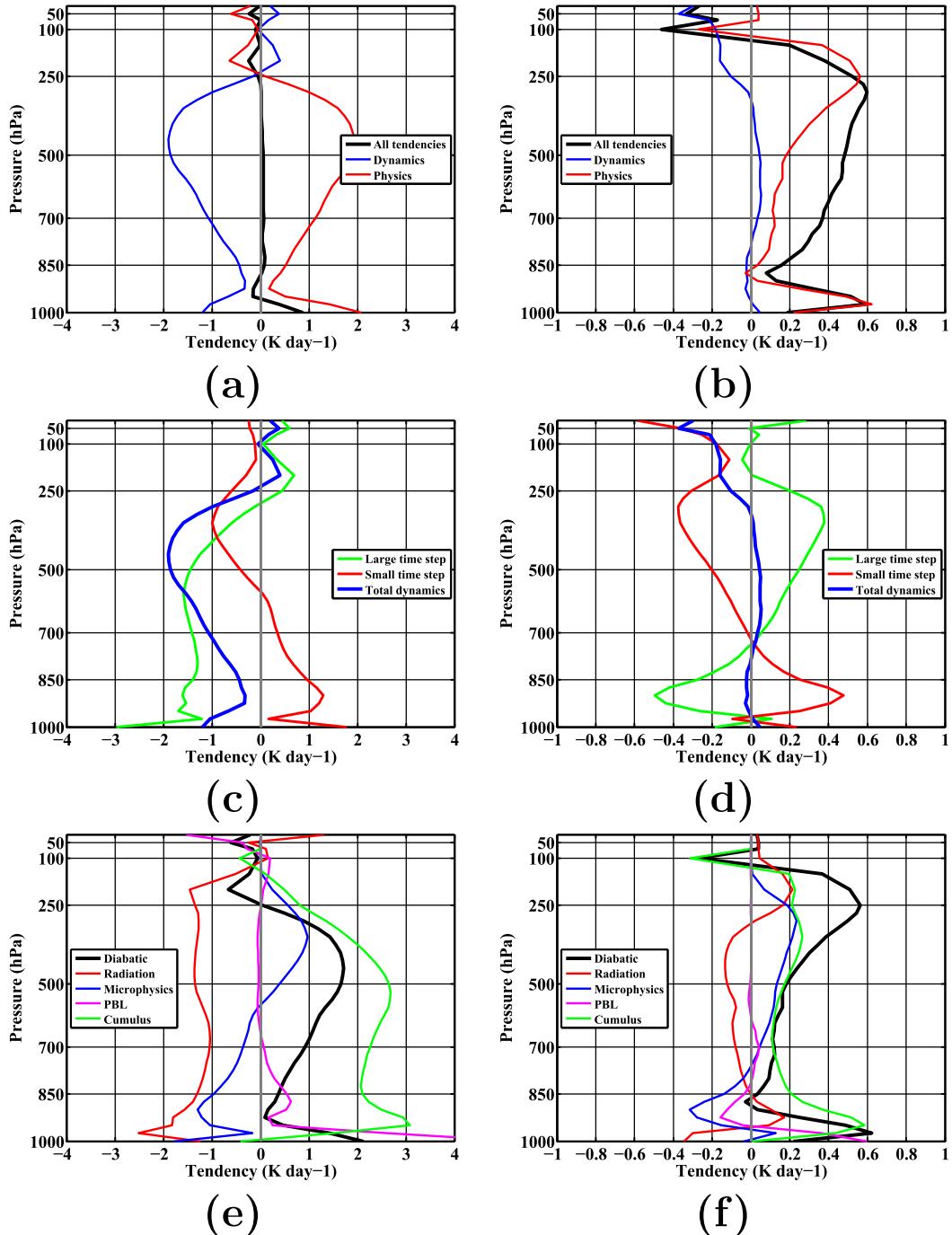


FIG. 5. (a),(c),(e) Comparison of time-averaged vertical profiles of (a) net (black line), dynamics (blue line), and physics (red line) tendencies; (c) net dynamics (blue line), large time-step dynamics (green line), and small time-step dynamics (red line) tendencies; and (e) net diabatic (black line), radiation (red line), microphysics (blue line), planetary boundary layer (magenta line), and cumulus (green line) tendencies from forecasts initialized from the ensemble member 31 averaged over the 60–90-min period of each simulation over the period 1–15 Aug 2010 and over the entire domain, but not including points within 4 grid points of the domain boundary. (b),(d),(f) As in (a),(c),(e), but for the time-averaged tendencies over the 60–90-min period of each simulation minus the time-averaged tendencies over the 0–30-min period of each simulation.

global GCM with a model time step of 30 min in comparison to 3.33 min used here, it is plausible that this initial adjustment may occur over many time steps; we will refer to this as a “fast” component of model adjustment and examine this behavior in more detail later in this section.

The tendencies near the surface are strongly indicative of an erroneous surface heat flux, potentially enhancing low-level instability and convective cloud formation. The 1000-hPa analysis increment makes it immediately apparent that there is a large area over which negative analysis increments are present over the northern Atlantic Ocean north of 30°N and over the Gulf of Mexico (Fig. 6a). Note from (7) that the mean analysis increment should be approximately equal in magnitude but of opposite sign to the mean model tendencies (Fig. 6b). The analysis increment and model tendency patterns show that there is a consistent contrast between observations and model forecasts; that is, the assimilation of observations is consistently counteracting the low-level heating produced in the model forecast simulations. Differences with GFS analyzed sea surface temperatures (SSTs) during the same period confirm that SSTs are considerably warmer in AHW–DART (Fig. 6c). This diagnosis led to the discovery that SSTs were not being updated in the AHW–DART system in real time. This example demonstrates use of the MITA increment method and how it can be used to identify and alleviate issues in complex modeling systems such as AHW–DART. Upon this discovery, SSTs in the AHW–DART system were corrected during real-time operations so that subsequent problems did not persist through the remainder of the TC season.

We now evaluate the source of bias in the middle troposphere at 700 hPa, where there is a relatively strong and statistically significant departure of $\bar{\theta}$ from zero. The net physics tendencies have an oscillatory pattern, varying around 1 K day^{-1} , and this oscillatory pattern derives from the convective heating tendency (Fig. 7a). Examination of tendencies at every model time step elucidates the reason for the discrepancy discussed earlier between the 0–30- and 60–90-min mean tendencies. In particular, there is an apparent fast component of model adjustment, where the net physics tendencies begin at -1.2 K day^{-1} on the first time step, followed by a rapid increase to $+0.6 \text{ K day}^{-1}$ on the second time step. This fast component of model adjustment in the first two time steps occurs over a time of 6.67 min. Subsequently, the tendencies rather quickly equilibrate to heating rates around $+1.25 \text{ K day}^{-1}$ for the duration of the DA cycling period. There remain slow increases in convective heating rates for ~ 16 time steps (approximately 50 min) followed by brief and “sudden”

decreases approximately every hour. The sudden changes that occur every hour in the convective parameterization occur at the convective adjustment time scale that is set within the parameterization. The convective adjustment time scale is approximately the time scale for clouds to advect over one grid element, determines how often convective initiation is checked for on each grid point, and is restricted to be within 1800–3600 s (e.g., Kain and Fritsch 1993; Bullock et al. 2015). Here, average wind speeds at common cloud levels between 850 and 500 hPa range from 4.9 to 7.7 m s^{-1} , respectively, so that with a grid spacing of 36 km, the advective time scale is exactly 1 h. Once a grid point is flagged as “convective,” then convective heating rates will be computed on that grid point for the duration of the convective adjustment time period. The convective heating rate can subsequently change at each model time step due to feedbacks with the model’s other parameterized heating rates.

In addition to the oscillatory pattern in convective heating rates at 700 hPa discussed above, there is a slow increase in the magnitude of convective heating until the end of the convective adjustment time period. The PBL tendencies are nearly zero at 700 hPa, while both the microphysics and radiation tendencies are negative, but much smaller in magnitude than the convective heating tendencies. Since all other tendency components are steady in comparison to the convective heating term, it is quite evident that the oscillatory pattern in the total physics tendencies is derived from the convective heating parameterization, and it is evident that the net positive imbalance in tendencies is derived from the convective heating parameterization. Furthermore, the MITA increment method cannot be applied over at least the first two time steps in order to correctly diagnose the systematic bias in our modeling configuration, as the systematic bias is representative of the equilibrium tendencies after the fast component of model adjustment has occurred (cf. Figs. 3 and 7a). Moreover, it is also apparent that there are time steps throughout the 6-h forecast integration that also may not be representative of the systematic bias due to the convective adjustments computed within the Kain–Fritsch convective parameterization.

The character of the model spinup from initial conditions seen here is somewhat unexpected, in that the tendencies from the first time steps are not representative of the systematic model bias, as might be expected from the results of Klinker and Sardeshmukh (1992) and Rodwell and Palmer (2007). Furthermore, given that the same forecast model is used between the AHW–DART DA system and forecast tendencies discussed here so far, and assuming the DA is adequate, we

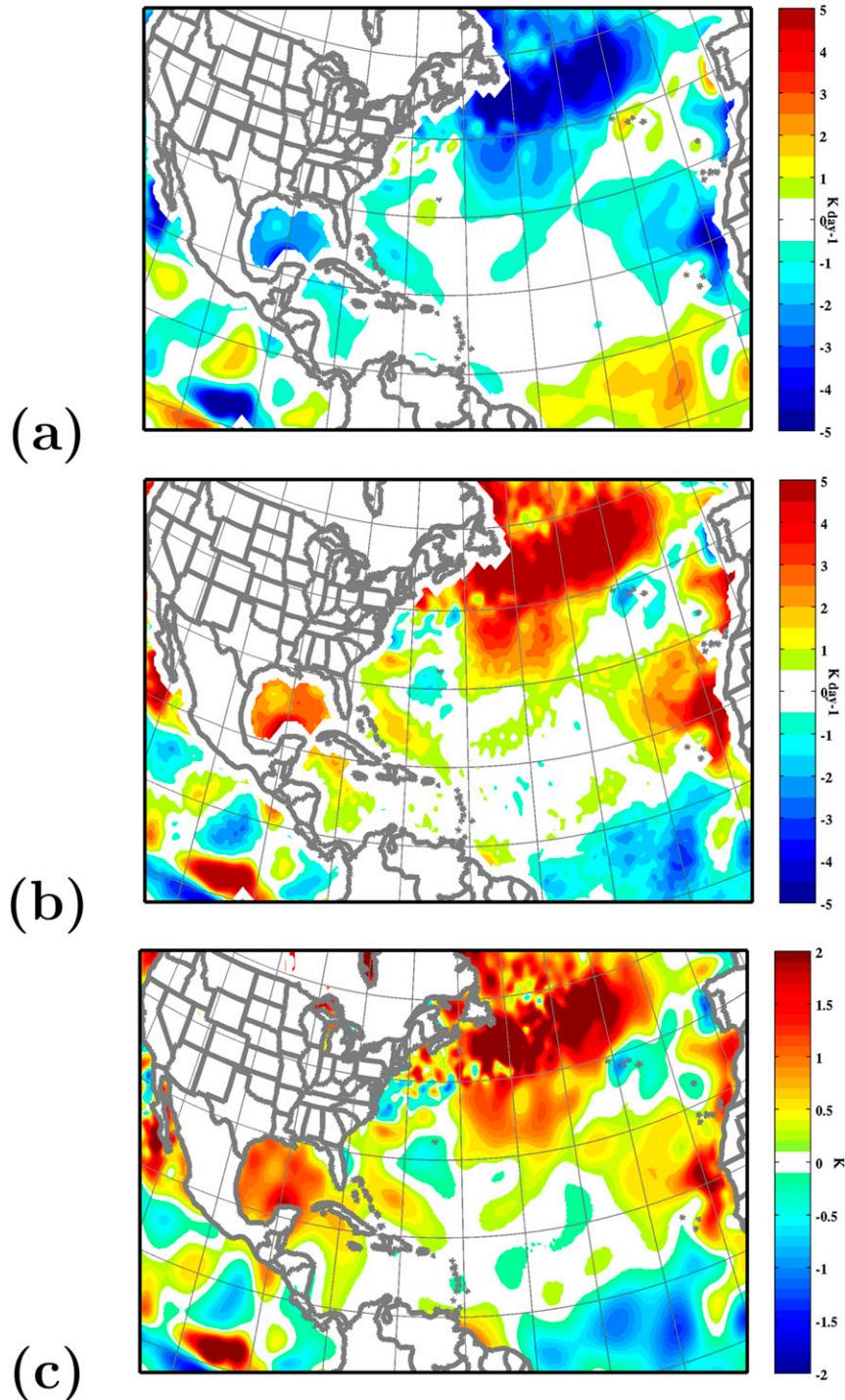


FIG. 6. Ensemble member 31 1000-hPa potential temperature mean (a) analysis increment, (b) net model tendency, and (c) difference between AHW-DART and GFS (AHW-DART – GFS) sea surface temperature field. Averages are computed over the period 0000 UTC 1 Aug–1800 UTC 15 Aug 2010.

expect a relatively fast adjustment from initial conditions. We now explore the spinup process in more detail by performing transpose-AMIP-type experiments where the model configuration and initial conditions are varied to

determine whether short-term forecasts can be utilized with the MITA increment method in our mesoscale modeling configuration; experiments are summarized in Table 1. In the first experiment (E1), model physics are

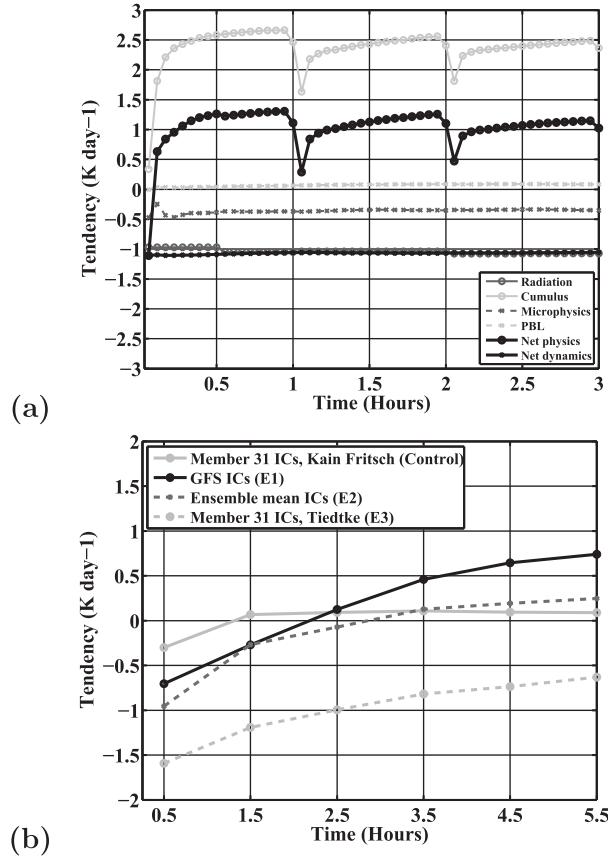


FIG. 7. Ensemble member 31 (control) 700-hPa potential temperature model (a) tendencies at each time step for each physics component and (b) total tendencies averaged over 30-min intervals ending at the time denoted with a filled circle for forecasts initialized using GFS (solid black line; E1), AHW-DART ensemble mean (dashed line, dark gray line; E2), and AHW-DART ensemble member 31 with Tiedtke convection (dashed, light gray line; E3). Tendencies in both (a) and (b) are composites over the period 0000 UTC 1 Aug–1800 UTC 15 Aug 2010. The tendencies in (a) include the net physics tendency (solid, black circles), radiation (solid, dark gray circles), cumulus (solid, light gray circles), microphysics (dashed, dark gray circles), PBL (dashed, light gray circles), and net dynamics tendencies (solid, black triangles).

identical to the AHW-DART control, but forecasts are initialized with the operational version of GFS analyses. In the second experiment (E2), model physics are identical to AHW-DART, but forecasts are initialized with the AHW-DART ensemble-mean analyses instead of ensemble member 31. In the third experiment (E3), initial conditions are the same as in the control, but forecast simulations are performed using the Tiedtke convective parameterization (Zhang et al. 2011), since it has been previously shown to exhibit lower convective heating rates over tropical regions (Torn and Davis 2012).

Results show substantial differences in the mean forecast tendencies between experiments (Fig. 8), with

statistically significant differences at all levels between 350 and 850 hPa. To further evaluate these differences, we examine the forecast tendencies at 700 hPa. If we would like to use the forecast tendencies to diagnose systematic model bias, then it is desirable that the tendencies equilibrate to a nearly constant value. That is, the analysis increment diagnoses the mean drift of the model state from observations over the duration of a DA cycling period. Net forecast tendencies that are consistently close to the negative value of this mean analysis increment over many time steps lends confidence that those forecast tendencies in particular represent the model trajectories that evolve the state toward the biased state present at the end of the DA cycling period. We would also expect that if model error dominates the influence of initial conditions, then the forecast tendencies in the experiments would become indistinguishable from the control, implying that the forecasts are on the same “trajectory” as those in the control (Klocke and Rodwell 2014). A comparison of the 700-hPa forecast tendencies averaged over each of the 30-min forecast intervals for all experiments is shown in Fig. 7b. In the control, the 60–90-min forecast tendencies are nearly identical to all subsequent averaging intervals. In all other experiments, forecast tendencies are still changing through hour five, reducing confidence that those tendencies represent the systematic model bias. Moreover, at forecast hour five, the tendency magnitudes are significantly different among the experiments: +0.75 K day⁻¹ in E1, -0.25 K day⁻¹ in E2, and -0.6 K day⁻¹ in E3, compared to +0.1 K day⁻¹ in the control. The large tendency differences in all experiments relative to the control, even by the end of our short-term forecast period, implies that none of the forecasts have reached a time window where local model error dominates and the MITA increment method can be successfully applied. This suggests that we can only use the control to diagnose the model bias with much confidence. All experiments show an initial fast model adjustment component, where the net tendency is initially negative. In all experiments except for the control, this stage is followed by a slower model adjustment component where it takes hours to days before tendencies reach a nearly steady value. Regarding the overall model adjustment process, we conclude from these experiments that it is minimized when ICs are an individual ensemble member, and that even using the first few time steps to diagnose model bias in a mesoscale model will provide misleading results about the source of bias. For example, if using ensemble mean or GFS ICs, applying the MITA increment method to the first 50 time steps may not be sufficient to correctly diagnose the systematic model bias. The number of time steps that it takes the model tendencies to achieve a quasi-constant state likely depends

TABLE 1. Summary of numerical modeling experiments.

Expt name	Model physics	Initial conditions
Control	Standard*	AHW-DART ensemble member 31
E1	Standard	GFS operational analyses
E2	Standard	AHW-DART ensemble mean
E3	Standard except with Tiedtke cumulus parameterization	AHW-DART ensemble member 31

* Details of the standard AHW-DART model physics are described in section 2a.

on several factors, including horizontal grid spacing, model time step, and the specific choice of physics parameterizations.

b. Experiment E1: Initializing forecasts from a nonnative analysis

The forecast tendencies in E1 do not equilibrate by forecast hour five, and in fact, do not do so until around forecast day 3 (not shown). However, toward the end of the DA cycle when the magnitude of the net tendencies are changing the least in time (recall Fig. 7b), the sign of the mean tendency is the same as in the control, and the time-averaged vertical profiles exhibit a similar pattern to the control in the middle troposphere (recall Fig. 8). It is worth exploring whether the MITA increment method could lead to the same conclusion as with the control.

To evaluate whether the E1 tendencies are representative of the systematic model bias, we compare mean tendencies ending at hour five of experiment E1 with the mean 60–90-min forecast tendencies during the control case (Fig. 9). The net tendencies are ~ 7 times stronger throughout the middle troposphere from ~ 500 to 850 hPa than in the control case, with substantially larger dynamics and large time-step dynamics tendencies, particularly around 400 hPa (Figs. 9a–d). The shape of the vertical physics tendency profile is similar to the control case except with slightly larger (smaller) tropospheric values below (above) 500 hPa, primarily associated with additional microphysical and convective heating (Figs. 9e,f). The increase in convective heating, alone, in the middle troposphere around 700 hPa is responsible for almost the entire increase in heating from the physics parameterizations, and is larger than the total heating in the control. There is less heating in the PBL scheme near the surface because SSTs are from GFS ICs and, thus, are not erroneously fixed as in the control. Given that the erroneous PBL heating is not because of a parameterized process error in the PBL scheme itself, it appears that this experiment would correctly indicate the possibility that

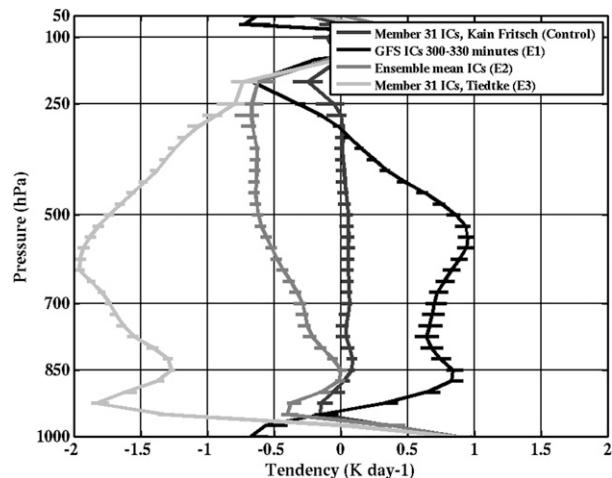


FIG. 8. Comparison of time-averaged vertical profiles of the net potential temperature tendencies for forecasts initialized using AHW-DART ensemble member 31 (dark gray line), GFS (black line; E1), AHW-DART ensemble mean (medium gray line; E2), and AHW-DART ensemble member 31 with Tiedtke convection (light gray line; E3) averaged over the period 1–15 Aug 2010 and over the entire domain, but not including points within 4 grid points of the domain boundary. All tendencies shown are averages over the 60–90-min forecast time steps except for E1, which is shown for averages over the 300–330-min forecast time steps. Error bars represent the 70% confidence intervals of the forecast tendencies as described in the text.

there is a physical process error in the convective heating while correctly providing no indication of erroneous heating from the PBL parameterization. However, the fact still remains that the average behaviors of the two forecast trajectories are drastically different (recall Fig. 7b), and, therefore, the tendencies are dominated by systematic differences in the initial conditions. This perhaps is an indication that on individual grid points, the tendencies move θ forward in time correctly, but that the different scales of motion from GFS ICs leads to convective initiation over more grid points, and therefore leads to a higher mean tendency.

c. Experiment E2: Initializing forecasts from the ensemble mean

Though it is somewhat expected that forecasts initialized from ICs of another model may take longer to equilibrate to a balanced model state, the reason is not immediately apparent why forecasts initialized from an ensemble mean of the same ensemble exhibit similar imbalance characteristics. We first note that the net forecast tendencies do not balance the analysis increment for forecasts initialized from the ensemble mean (Fig. 10a). After some consideration, it is important to note that the ensemble mean of the 6-h forecasts, $\theta_{i,n}^b$, does not equal the 6-h deterministic forecast initialized from the ensemble mean, $\theta_{i,n}^b$ on the i th DA

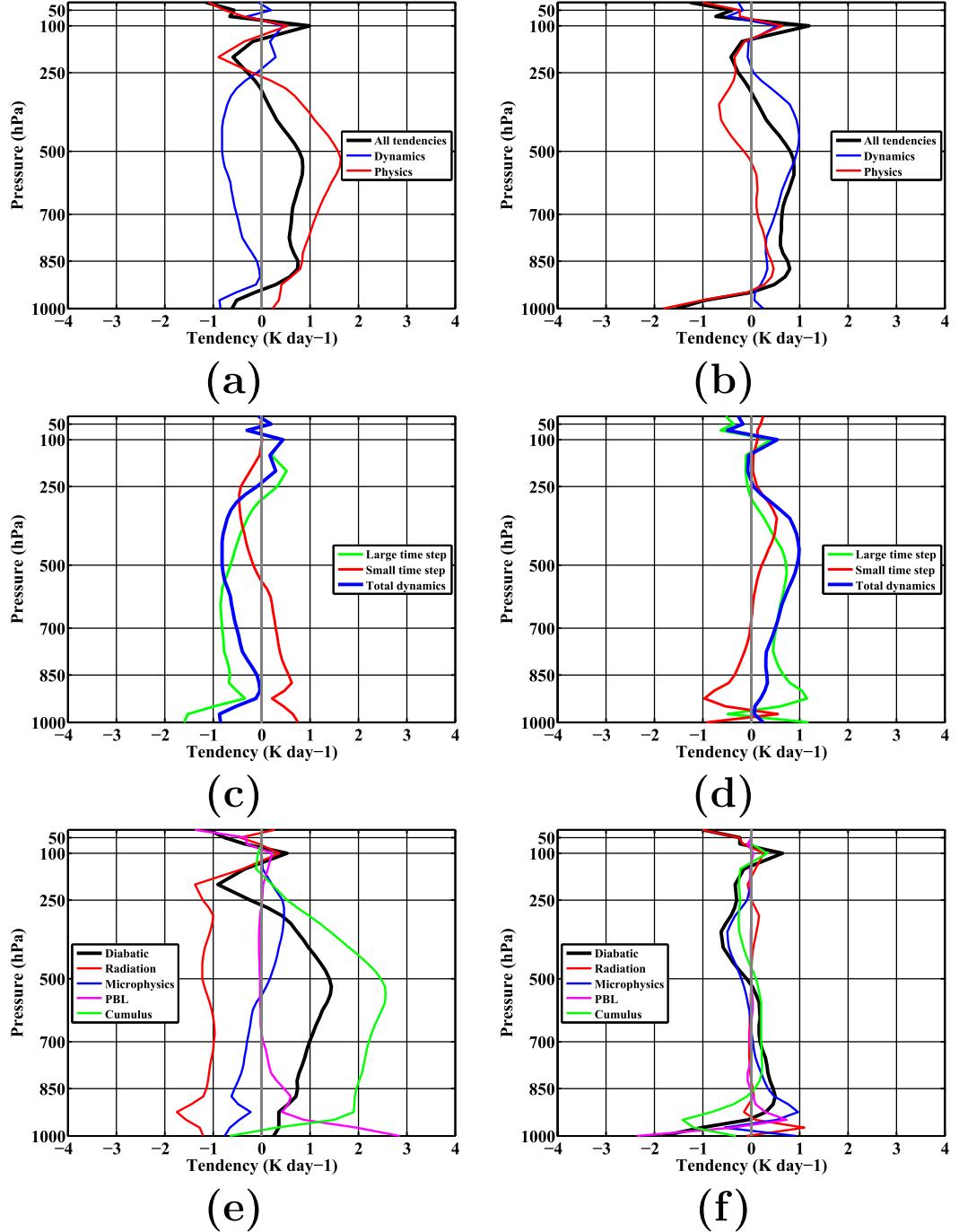


FIG. 9. (a),(c),(e) Comparison of time-averaged vertical profiles of (a) net (black line), dynamics (blue line), and physics (red line) tendencies; (c) net dynamics (blue line), large time-step dynamics (green line), and small time-step dynamics (red line) tendencies; and (e) net diabatic (black line), radiation (red line), microphysics (blue line), planetary boundary layer (magenta line), and cumulus (green line) tendencies from forecasts initialized from GFS over the 150–180-min period (hour 5) of each simulation over the period 1–15 Aug 2010 and over the entire domain, but not including points within 4 grid points of the domain boundary. (b),(d),(f) As in (a),(c),(e), but for the time-averaged tendencies over the 150–180-min period of each simulation from the GFS IC case minus the time-averaged tendencies over the 60–90-min period of each simulation from ensemble member 31.

cycle. That is, instead of (1) and (2), the mean increment and tendencies when initializing forecasts from the ensemble mean are

$$\text{INC}_i = \widetilde{\theta_{i,0}^a} - \widetilde{\theta_{i-1,n}^b} \quad \text{and} \quad (10)$$

$$\overline{\dot{\theta}_i^b} = \frac{\theta_{i,n}^b - \widetilde{\theta_{i,0}^a}}{\Delta t_{\text{da}}}, \quad (11)$$

where $\widetilde{\theta_{i,0}^a}$ is the ensemble mean analysis at the i th DA cycle. Replacing (1) and (2) with (10) and (11) and following the procedure in section 2b, the equivalent analysis increment in this case is

$$\sum_{i=1}^m \text{RHS}_i = -\Delta t_{\text{da}} \sum_{i=1}^{m-1} \overline{\dot{\theta}_i^b} + \sum_{i=1}^{m-1} (\theta_{i,n}^b - \widetilde{\theta_{i,n}^b}) + \text{endpoints}, \quad (12)$$

where $\text{endpoints} = \widetilde{\theta_{m,0}^a} - \widetilde{\theta_{0,n}^b}$, and where we refer to the equivalent analysis increment as the right-hand side of the i th DA cycle (RHS_i) to avoid ambiguity with the analysis increment of the ensemble given by (7). We note that the second term on the right-hand side of (12) is a new term that does not appear in (7). This new term is the sum of the differences between the 6-h forecasts of the individual ensemble members and the mean of the prior, which we subsequently refer to as the “prior difference” term. Figure 10a shows that the analysis increment is in close agreement to the sum of each term comprising the right-hand side of (12) when including an estimate of the prior difference term. The difference between the analysis increment term and the right-hand side of (12) arises because the right-hand side no longer explicitly contains an analysis increment term as defined by (10). Note that the prior difference term is quite large, and in fact, it is the largest term in (12) at 700 hPa, and is larger in magnitude than the analysis increment.

To elucidate the differences between the evolution of the forecasts initialized from an individual ensemble member (control) and the forecasts initialized from the ensemble mean (experiment E2), a cartoon is provided to illustrate forecasts within an individual DA cycle at three different atmospheric levels: lower (850 hPa), middle (500 hPa), and upper (200 hPa) (Fig. 11). In the control case at 200 hPa, the forecast tendencies are negative at around -0.2 K day^{-1} (Fig. 10b). Thus, the forecast tendencies generally decrease θ after the analysis time, and the ensemble envelope reflects the overall decrease in θ of each of the ensemble members over time (Fig. 11a). Nonzero mean tendencies imply that the model θ values are drifting away from observations. Thus, at the end of the DA cycle, the analysis increment reflects the “pull” that the observations exert on the model forecast.

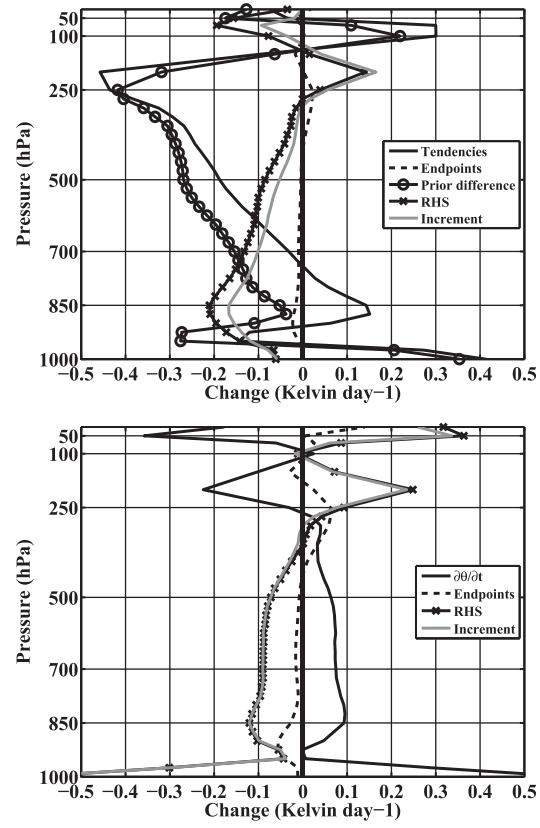


FIG. 10. Vertical profiles of 6-h potential temperature change ($d\theta/dt$; solid black line), analysis drift (“endpoints”; black dashed line), 6-h forecast minus ensemble mean potential temperature (prior difference; black “O”), right-hand side (RHS; black “ \times ”), and analysis increment (gray line) for the (a) forecasts initialized from the ensemble mean and (b) forecasts initialized with ensemble member 31. Values are averages over the period 0000 UTC 1 Aug–1800 UTC 31 Aug 2010 in (a) and 0000 UTC 1 Aug–1800 UTC 15 Aug 2010 in (b) and are spatially averaged over the entire domain, but not including points within 4 grid points of domain boundary. See text for further details.

At 200 hPa, both the forecast tendencies and the prior difference terms are negative, large in magnitude, and much larger in magnitude than the analysis increment (Fig. 10a). The relatively small analysis increment indicates that the ensemble envelope has not diverged from observations as much as the forecasts in E2. The large magnitude of the prior difference term, and the fact that this term is much larger than the magnitude of the analysis increment, indicate that forecasts in E2 are not representative forecasts of the ensemble itself. That is, forecasts initialized from the ensemble mean exhibit drastically different behavior than any individual ensemble forecast member, and likely do not fall within the entire ensemble envelope at 200 hPa (Fig. 11a). This could indicate that the analysis is fitting to the bias derived from ACARS observations, and thus the analysis has a large bias of opposite

sign relative to radiosondes. Model forecasts systematically move away from ACARS and toward radiosondes (and perhaps the truth), and in this regard, the forecasts initialized from the ensemble mean may be less biased than any of those from the ensemble members. At 500 hPa, the magnitude of the prior difference term is again much larger than the magnitude of the analysis increment, indicating that forecasts in E2 may not lie within the ensemble envelope (Fig. 11b). The positive (negative) analysis increment at 200 (500) hPa implies that ensemble θ drifts to values that are colder (warmer) than observations over time. However, since tendencies are negative at both 200 and 500 hPa, the forecasts in E2 must drift toward observations over time at 500 hPa, but away from observations at 200 hPa. In other words, forecast biases in E2 decrease in time at 500 hPa while they increase in the control. At 850 hPa, the magnitudes of the analysis increment and the tendencies are more comparable, while the prior difference term is smaller in magnitude. This implies that forecasts in E2 are similar to the control at 850 hPa, and remain within the ensemble envelope (Fig. 11c).

Note that E2 is a transpose-AMIP-type experiment similar to E1, since forecasts using the same model as in the control are initialized from different ICs. However, the short-term tendencies in E2 do not bear much resemblance to those in either the control or E1. At all levels below 175 hPa, the mean forecast tendencies are negative and are much lower in E2 than in the control (cf. Figs. 8 and 12a). Indeed, at the levels where E2 forecasts drift closer to observations, the largest reductions are in convective heating (Fig. 12b). Most of the reductions at 500 hPa occur near the intertropical convergence zone (ITCZ), and the reduction pattern indicates that it is primarily localized regions of convection that perhaps either do not occur, or are largely reduced in the forecasts initialized from the ensemble mean (Fig. 13). In conclusion, the MITA increment method clearly is not a useful diagnostic of model error when initializing forecasts from the ensemble mean.

d. Experiment E3: Changing the model

We now ask whether the apparent bias deriving from the Kain–Fritsch convective scheme can easily be alleviated by using a different convective parameterization. Experiment E3 shows that “switching” to a different cumulus parameterization scheme reduces the positive tendencies, but leads to an overall greater model imbalance since the ICs were created using the Kain–Fritsch parameterization (Figs. 7b and 8). The convective heating tendencies in E3 do not exhibit the same oscillatory pattern as those from the control case (not shown), though the sudden decreases in Kain–Fritsch each hour do not appear to affect the

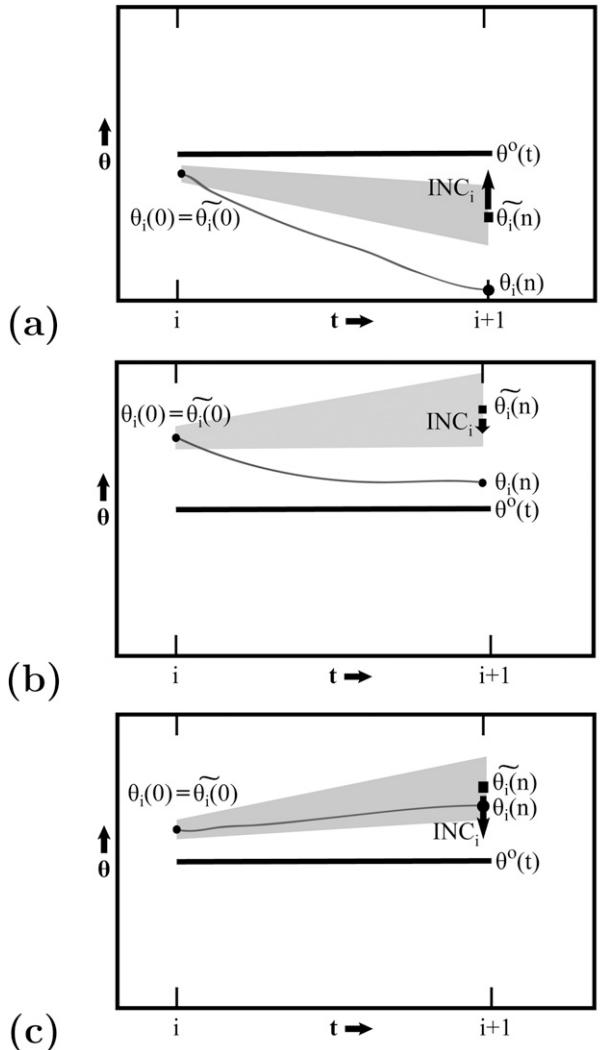


FIG. 11. Cartoon of the potential temperature forecast evolution at (a) 200, (b) 500, and (c) 850 hPa. The data assimilation cycle is indicated by index i , while time is given by t , ranging from time steps 0 to n . Observations are indicated by the thick, black line labeled $\theta^o(t)$. The gray line denotes a deterministic forecast evolution in potential temperature, where the initial state is equal to the ensemble mean values. The gray shading denotes an ensemble forecast envelope. The deterministic potential temperature at the analysis time and time step n are given by $\theta_i(0)$ and $\theta_i(n)$, respectively. The ensemble mean potential temperature at the analysis time and time step n are given by $\tilde{\theta}_i(0)$ and $\tilde{\theta}_i(n)$, respectively. The potential temperature ensemble mean analysis increment is indicated by INC_i .

overall net bias or its diagnosis, so long as a sufficient number of time steps are used when computing model bias with the MITA increment method. Torn and Davis (2012) showed that when cycling the AHW–DART DA system using the Tiedtke cumulus parameterization, forecast bias is indeed reduced by $\sim 1 \text{ K day}^{-1}$ at 700 hPa, with a net zero bias when averaged over the

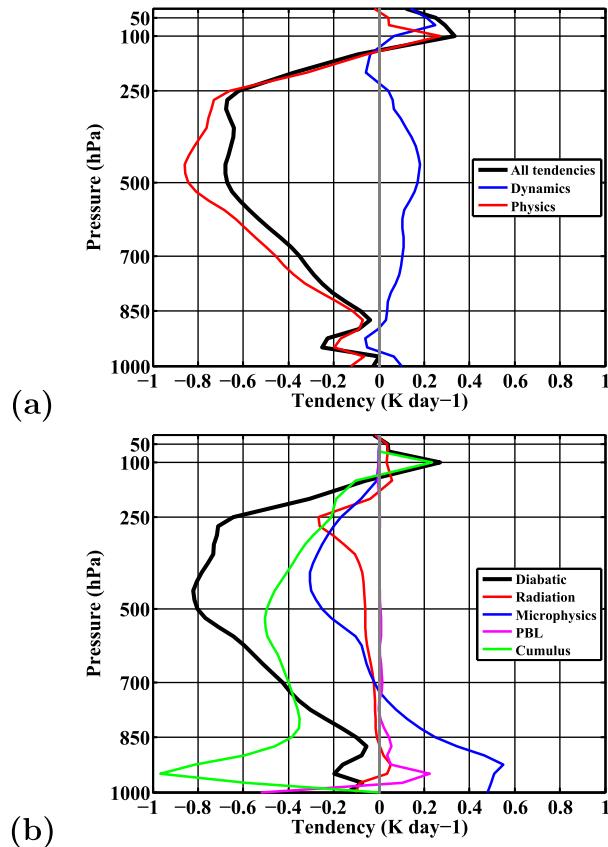


FIG. 12. Comparison of vertical profile differences of (a) net dynamics (black line), large time-step dynamics (black dashed line), and small time-step dynamics (gray dashed line) tendencies and (b) net diabatic (black line), radiation (thick black dashed line), microphysics (black “O”), planetary boundary layer (black “x”), and cumulus (thin black) between the forecasts initialized from the ensemble mean and the forecasts initialized from ensemble member 31 (ensemble mean – member 31). Tendencies are averaged over the 60–90-min period of each simulation over the period 1–15 Aug 2010 and over the entire domain, but not including points within 4 grid points of the domain boundary.

vertical column. Thus, even a slight modification to the model with the change of convective parameterization renders a slow adjustment from initial conditions to the model’s own attractor, such that short-term forecast tendencies do not represent the true model bias.

4. Summary and conclusions

The mean initial tendency analysis (MITA) increment method was applied here for the first time in a limited-area model (LAM) to diagnose systematic forecast biases in the Advanced Hurricane WRF Data Assimilation (DA) Research Testbed (AHW–DART) forecasting

system during the 2010 Atlantic hurricane season. The AHW–DART system performed cycling DA every 6 h using a 96-member DART implementation of the ensemble adjustment Kalman filter (EAKF) technique from 28 July to 7 November 2010. Forecast tendencies of an individual ensemble member were analyzed here during a subset of the cycling period from 0000 UTC 1 August to 1800 UTC 15 August 2010. To isolate model error from the model’s adjustment from initial conditions, we sought an analysis of the time steps after which the net tendencies converged to a nearly constant value and that was represented by the mean analysis increment (e.g., Judd et al. 2008; Klocke and Rodwell 2014).

Results showed that there were systematic model biases in potential temperature, including a warm bias near the surface and a smaller but vertically elongated warm bias in the middle troposphere. The warm bias near the surface derived from the PBL parameterization, and was due to erroneously fixed sea surface temperatures in AHW–DART. This led to strong upward sensible heat fluxes and, hence, warmer low-level temperatures than observed. A cold bias at 200 hPa is likely related to a documented instrument bias in ACARS.

The middle tropospheric warm bias was found to derive from the convective heating parameterization. Evaluation of this bias revealed several interesting aspects of the model’s spinup from initial conditions. First, there was an initial fast component of spinup consisting of ~2 model time steps related to a time lag for the convective heating parameterization to initiate, similar to the first time-step tendencies described by Rodwell and Palmer (2007), where it was noted that the first time step differed due to structural differences when beginning forecasts from the analysis. Second, changing the convective heating parameterization from the Kain–Fritsch to Tiedtke scheme reduced the magnitude of convective heating; however, the model failed to reach an equilibrium tendency within the DA cycling period since the Tiedtke scheme was not used in the forecast model used in creating the initial conditions. This confirmed that short-term forecasts could only successfully isolate model and initial condition error when using a native analysis. Third, short-term forecast tendencies in experiments using GFS and ensemble mean ICs both failed to reach equilibrium states within the DA cycling period, and the GFS IC experiment did not yield nearly constant forecast tendencies until around forecast day 3. Interestingly, the pattern of the time-averaged vertical profiles of the GFS IC experiment were similar to the control experiment using native ICs by forecast hour 5, yielding the possibility that the tendencies could still be beneficial in qualitatively diagnosing model bias. Since these results

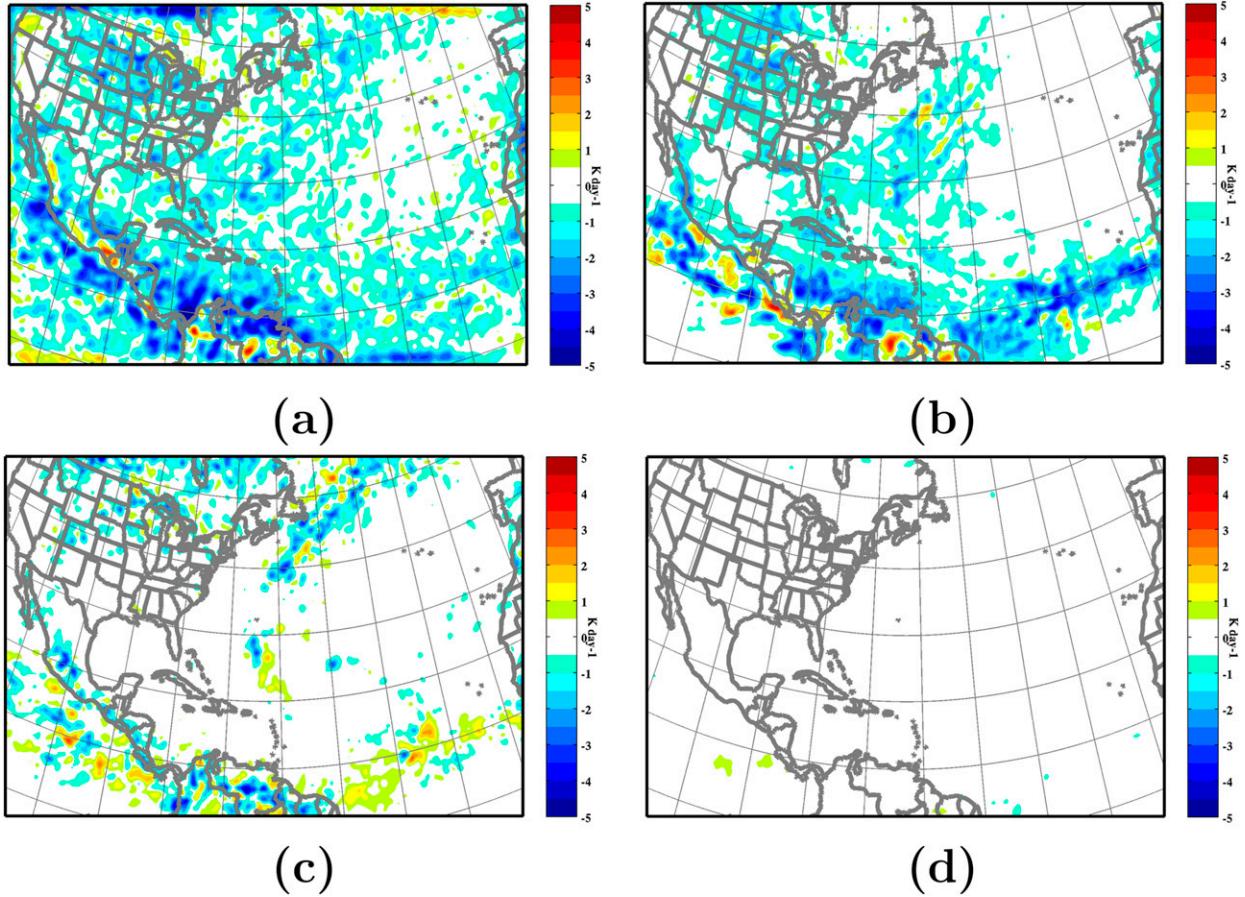


FIG. 13. AHW–DART differences at 500 hPa between the 60–90-min tendencies from forecasts initialized from the ensemble mean and the forecasts initialized from ensemble member 31 (ensemble mean – member 31) with the (a) net, (b) cumulus, (c) microphysical, and (d) radiative tendencies.

are valid with one modeling configuration, more investigation is needed to make a stronger generalization. In general, due to the relatively short forecast lead times between DA cycles in typical mesoscale LAM configurations, it may only be possible to diagnose the systematic model bias using short-term forecast tendencies when initializing forecasts from a native analysis.

In any case, it is helpful if the net forecast tendencies converge to a nearly constant value when the forecast tendencies represent the mean analysis increment. Such convergence does not necessarily occur in the first few time steps, even when using a native analysis, and there are time steps throughout the model integration where tendencies may not represent the net model bias due to convective adjustments in the convective parameterization scheme. Forecast tendencies in preliminary work with the WRF Model in a similar configuration over a different region where $\Delta t = 144$ s do not equilibrate until around 25 time steps. Since this is the first time applying the MITA increment method in a LAM with relatively

small time steps, we cannot make a general conclusion regarding the number of time steps, and future work would be helpful to determine this aspect.

The implementation of the MITA increment method here was used as a postprocessing tool requiring large files to accommodate tendencies at every time step. Analyzing tendencies from shorter time intervals can be successful as long as the forecast model has appropriately spun up to represent the mean analysis increment. An alternative computational approach is to retain a running sum of the tendencies within the modeling integration framework; the authors plan to implement this approach in future WRF releases.

The warm middle tropospheric bias has previously been connected to tropical cyclone track errors in the AHW system during the 2010 North Atlantic hurricane season by Torn and Davis (2012) and Galarneau and Davis (2013). In particular, Torn and Davis (2012) showed that a temperature bias of 1 K centered at 700 hPa over much of the western Atlantic Ocean could

be attributed to a lack of shallow convection from the Kain–Fritsch scheme. This heating bias led to an easterly wind bias and westerly TC forecast track bias. In contrast to deep convection, shallow convection produces cooling (heating) above (below) the top of the PBL (Han and Pan 2011). Thus, if convection is “triggered,” but shallow convection is underrepresented, a warm bias will be evident above the PBL top where there would otherwise be cooling in a shallow convection regime. Our tests showed that while switching forecasts from using the Kain–Fritsch to Tiedtke convective heating parameterization reduced the magnitude of the convective heating tendencies, systematic improvement in forecast bias could not be isolated from the differences in initial conditions. Given that the use of the MITA increment method would require removing these differences in initial conditions by cycling DA with the Tiedtke parameterization, the results of Torn and Davis (2012) support these results since they did not similarly find a warm temperature bias in the middle troposphere when they cycled the AHW–DART DA system using the Tiedtke parameterization.

Given the complexities of parameterizing physical processes in numerical models, it is greatly beneficial to employ a mesoscale forecasting system for which forecasts are initialized from its own DA system. This provides a method to properly diagnose the source of model error, and substantially alleviates initial condition spinup. Although this method cannot automatically determine causation, it is a necessary step that objectively narrows the source of error so that “educated” hypothesis tests can be subsequently devised. An ideal application with this method is in data-sparse regions where uncertainty in model parameterizations is relatively high and where flow-dependent background error information in the EAKF could be advantageous (e.g., Whitaker et al. 2009). One such region is over the Southern Ocean, where there are relatively few in situ data near an active storm track.

Acknowledgments. We would like to acknowledge high-performance computing support provided by NCAR’s Computational and Information Systems Laboratory sponsored by the National Science Foundation. We also thank Jimy Dudhia (NCAR), Jack Kain (NSSL), William Skamarock (NCAR), Ryan Torn (SUNY-Albany), and Wei Wang (NCAR) for their helpful conversations regarding this work. Satellite atmospheric motion vector data were obtained from Chris Velden and Dave Stettner (CIMSS/SSEC). ACARS data used in this study were made available to the Earth System Research Laboratory/Global Systems Division by the following commercial airlines: American, Delta,

Federal Express, Northwest, United, and United Parcel Service.

REFERENCES

- Anderson, J., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano, 2009: The Data Assimilation Research Testbed: A community data assimilation facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, doi:[10.1175/2009BAMS2618.1](https://doi.org/10.1175/2009BAMS2618.1).
- Anthes, R. A., and Coauthors, 2008: The COSMIC/FORMOSAT-3 mission: Early results. *Bull. Amer. Meteor. Soc.*, **89**, 313–333, doi:[10.1175/BAMS-89-3-313](https://doi.org/10.1175/BAMS-89-3-313).
- Ballish, B. A., and V. K. Kumar, 2008: Systematic differences in aircraft and radiosonde temperatures. *Bull. Amer. Meteor. Soc.*, **89**, 1689–1707, doi:[10.1175/2008BAMS2332.1](https://doi.org/10.1175/2008BAMS2332.1).
- Barker, D. M., W. Huang, R.-R. Guo, A. J. Bourgeois, and Q. N. Xiao, 2004: A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.*, **132**, 897–914, doi:[10.1175/1520-0493\(2004\)132<897:ATVDAS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<897:ATVDAS>2.0.CO;2).
- Bretherton, C. S., E. Klinker, A. K. Betts, and J. A. Coakley Jr., 1995: Comparison of ceilometer, satellite, and synoptic measurements of boundary-layer cloudiness and the ECMWF diagnostic cloud parameterization scheme during ASTEX. *J. Atmos. Sci.*, **52**, 2736–2751, doi:[10.1175/1520-0469\(1995\)052<2736:COCSAS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<2736:COCSAS>2.0.CO;2).
- Bullock, O. R. J., K. Alapaty, J. A. Herwehe, and J. S. Kain, 2015: A dynamically computed convective time scale for the Kain–Fritsch convective parameterization scheme. *Mon. Wea. Rev.*, **143**, 2105–2120, doi:[10.1175/MWR-D-14-00251.1](https://doi.org/10.1175/MWR-D-14-00251.1).
- Cavallo, S. M., J. Dudhia, and C. Snyder, 2011: A multilayer upper-boundary condition for longwave radiative flux to correct temperature biases in a mesoscale model. *Mon. Wea. Rev.*, **139**, 1952–1959, doi:[10.1175/2010MWR3513.1](https://doi.org/10.1175/2010MWR3513.1).
- , R. T. Torn, C. Snyder, C. Davis, W. Wang, and J. Done, 2013: Evaluation of the Advanced Hurricane WRF data assimilation system for the 2009 Atlantic hurricane season. *Mon. Wea. Rev.*, **141**, 523–541, doi:[10.1175/MWR-D-12-00139.1](https://doi.org/10.1175/MWR-D-12-00139.1).
- Chou, M.-D. C., and M. J. Suarez, 1994: An efficient thermal infrared radiation parameterization for use in general circulation models. NASA Tech. Memo. 104606, Vol. 3, NASA, 85 pp.
- Davis, C. A., and Coauthors, 2008: Prediction of landfalling hurricanes with the Advanced Hurricane WRF Model. *Mon. Wea. Rev.*, **136**, 1990–2005, doi:[10.1175/2007MWR2085.1](https://doi.org/10.1175/2007MWR2085.1).
- , W. Wang, J. Dudhia, and R. Torn, 2010: Does increased horizontal resolution improve hurricane wind forecasts? *Wea. Forecasting*, **25**, 1826–1841, doi:[10.1175/2010WAF2222423.1](https://doi.org/10.1175/2010WAF2222423.1).
- Dudhia, J., 1996: A multi-layer soil temperature model for MM5. Preprints, Sixth PSU/NCAR Mesoscale Model Users’ Workshop, Boulder, CO, NCAR, 49–50. [Available online at http://www2.mmm.ucar.edu/wrf/users/phys_refs/LAND_SURFACE/5_layer_thermal.pdf.]
- Dyer, A. J., and B. B. Hicks, 1970: Flux-gradient relationships in the constant flux layer. *Quart. J. Roy. Meteor. Soc.*, **96**, 715–721, doi:[10.1002/qj.49709641012](https://doi.org/10.1002/qj.49709641012).
- Galarneau, T. J., and C. A. Davis, 2013: Diagnosing forecast errors in tropical cyclone motion. *Mon. Wea. Rev.*, **141**, 405–430, doi:[10.1175/MWR-D-12-00071.1](https://doi.org/10.1175/MWR-D-12-00071.1).
- Han, J., and H. L. Pan, 2011: Revision of convection and vertical diffusion schemes in the NCEP global forecast system. *Wea. Forecasting*, **26**, 520–533, doi:[10.1175/WAF-D-10-05038.1](https://doi.org/10.1175/WAF-D-10-05038.1).

- Hannay, C., D. Williamson, J. J. Hack, J. T. Kiehl, J. G. Olson, S. A. Klein, C. S. Bretherton, and M. Köhler, 2009: Evaluation of forecasted southeast Pacific stratocumulus in the NCAR, GFDL, and ECMWF models. *J. Climate*, **22**, 2871–2889, doi:[10.1175/2008JCLI2479.1](https://doi.org/10.1175/2008JCLI2479.1).
- Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.*, **124**, 2322–2329, doi:[10.1175/1520-0493\(1996\)124<2322:NBLVDI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<2322:NBLVDI>2.0.CO;2).
- , J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.*, **132**, 103–120, doi:[10.1175/1520-0493\(2004\)132<0103:ARATIM>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2).
- Judd, K., C. A. Reynolds, T. A. Rosmond, and L. A. Smith, 2008: The geometry of model error. *J. Atmos. Sci.*, **65**, 1749–1772, doi:[10.1175/2007JAS2327.1](https://doi.org/10.1175/2007JAS2327.1).
- Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining detraining plume model and its application in convective parameterization. *J. Atmos. Sci.*, **47**, 2784–2802, doi:[10.1175/1520-0469\(1990\)047<2784:AODEPM>2.0.CO;2](https://doi.org/10.1175/1520-0469(1990)047<2784:AODEPM>2.0.CO;2).
- , and —, 1993: Convective parameterization for mesoscale models: The Kain–Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models*, Meteor. Monogr., No. 46, Amer. Meteor. Soc., 165–170.
- Kay, J. E., K. Raeder, A. Gettelman, and J. Anderson, 2011: The boundary layer response to recent Arctic sea ice loss and implications for high-latitude climate feedbacks. *J. Climate*, **24**, 428–447, doi:[10.1175/2010JCLI3651.1](https://doi.org/10.1175/2010JCLI3651.1).
- Klinker, E., and P. D. Sardeshmukh, 1992: The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.*, **49**, 608–627, doi:[10.1175/1520-0469\(1992\)049<0608:TDOMDI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1992)049<0608:TDOMDI>2.0.CO;2).
- Klocke, D., and M. J. Rodwell, 2014: A comparison of two numerical weather prediction methods for diagnosing fast-physics errors in climate models. *Quart. J. Roy. Meteor. Soc.*, **140**, 517–524, doi:[10.1002/qj.2172](https://doi.org/10.1002/qj.2172).
- Martin, G. M., S. F. Milton, C. A. Senior, M. E. Brooks, and S. Ineson, 2010: Analysis and reduction of systematic errors through a seamless approach to modeling weather and climate. *J. Climate*, **23**, 5933–5957, doi:[10.1175/2010JCLI3541.1](https://doi.org/10.1175/2010JCLI3541.1).
- Medeiros, B., D. L. Williamson, C. Hannay, and J. G. Olson, 2012: Southeast Pacific stratocumulus in the Community Atmosphere Model. *J. Climate*, **25**, 6175–6192, doi:[10.1175/JCLI-D-11-00503.1](https://doi.org/10.1175/JCLI-D-11-00503.1).
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated- k model for the long-wave. *J. Geophys. Res.*, **102**, 16 663–16 682, doi:[10.1029/97JD00237](https://doi.org/10.1029/97JD00237).
- Paulson, C. A., 1970: The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer. *J. Appl. Meteor.*, **9**, 857–861, doi:[10.1175/1520-0450\(1970\)009<0857:TMROWS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2).
- Phillips, T. J., and Coauthors, 2004: Evaluating parameterizations in general circulation models. *Bull. Amer. Meteor. Soc.*, **85**, 1903–1915, doi:[10.1175/BAMS-85-12-1903](https://doi.org/10.1175/BAMS-85-12-1903).
- Rodwell, M. J., and T. N. Palmer, 2007: Using numerical weather prediction to assess climate models. *Quart. J. Roy. Meteor. Soc.*, **133**, 129–146, doi:[10.1002/qj.23](https://doi.org/10.1002/qj.23).
- , and T. Jung, 2008: Understanding the local and global impacts of model physics changes: An aerosol example. *Quart. J. Roy. Meteor. Soc.*, **134**, 1479–1497, doi:[10.1002/qj.298](https://doi.org/10.1002/qj.298).
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., doi:[10.5065/D68S4MVH](https://doi.org/10.5065/D68S4MVH). [Available online at http://www.mmm.ucar.edu/wrf/users/docs/arw_v3_bw.pdf.]
- Torn, R. D., 2010: Performance of a mesoscale ensemble Kalman filter (EnKF) during the NOAA high-resolution hurricane test. *Mon. Wea. Rev.*, **138**, 4375–4392, doi:[10.1175/2010MWR3361.1](https://doi.org/10.1175/2010MWR3361.1).
- , and C. A. Davis, 2012: The influence of shallow convection on tropical cyclone track forecasts. *Mon. Wea. Rev.*, **140**, 2188–2197, doi:[10.1175/MWR-D-11-00246.1](https://doi.org/10.1175/MWR-D-11-00246.1).
- , G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.*, **134**, 2490–2502, doi:[10.1175/MWR3187.1](https://doi.org/10.1175/MWR3187.1).
- Velden, C. S., and Coauthors, 2005: Recent innovations in deriving tropospheric winds from meteorological satellites. *Bull. Amer. Meteor. Soc.*, **86**, 205–223, doi:[10.1175/BAMS-86-2-205](https://doi.org/10.1175/BAMS-86-2-205).
- Webb, E. K., 1970: Profile relationships: The log-linear range, and extension to strong stability. *Quart. J. Roy. Meteor. Soc.*, **96**, 67–90, doi:[10.1002/qj.49709640708](https://doi.org/10.1002/qj.49709640708).
- Whitaker, J. S., G. P. Compo, and J. N. Thépaut, 2009: A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observations. *Mon. Wea. Rev.*, **137**, 1991–1999, doi:[10.1175/2008MWR2781.1](https://doi.org/10.1175/2008MWR2781.1).
- Wicker, L. J., and W. C. Skamarock, 2002: Time-splitting methods for elastic models using forward time schemes. *Mon. Wea. Rev.*, **130**, 2088–2097, doi:[10.1175/1520-0493\(2002\)130<2088:TSMFEM>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2088:TSMFEM>2.0.CO;2).
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Academic Press, 676 pp.
- Williams, K. D., and M. Brooks, 2008: Initial tendencies of cloud regimes in the Met Office unified model. *J. Climate*, **21**, 833–840, doi:[10.1175/2007JCLI1900.1](https://doi.org/10.1175/2007JCLI1900.1).
- , and Coauthors, 2013: The transpose-AMIP II experiment and its application to the understanding of Southern Ocean cloud biases in climate. *J. Climate*, **26**, 3258–3274, doi:[10.1175/JCLI-D-12-00429.1](https://doi.org/10.1175/JCLI-D-12-00429.1).
- Williamson, D. L., and J. G. Olson, 2007: A comparison of forecast errors in CAM2 and CAM3 at the ARM Southern Great Plains site. *J. Climate*, **20**, 4572–4585, doi:[10.1175/JCLI4267.1](https://doi.org/10.1175/JCLI4267.1).
- , and Coauthors, 2005: Moisture and temperature balances at the Atmospheric Radiation Measurement Southern Great Plains site in forecasts with the Community Atmosphere Model (CAM2). *J. Geophys. Res.*, **110**, D15S16, doi:[10.1029/2004JD005109](https://doi.org/10.1029/2004JD005109).
- Xie, S., H.-Y. Ma, J. S. Boyle, S. A. Klein, and Y. Zhang, 2012: On the correspondence between short-and long-time-scale systematic errors in CAM4/CAM5 for the year of tropical convection. *J. Climate*, **25**, 7937–7955, doi:[10.1175/JCLI-D-12-00134.1](https://doi.org/10.1175/JCLI-D-12-00134.1).
- Zhang, C., Y. Wang, and K. Hamilton, 2011: Improved representation of boundary layer clouds over the southeast Pacific in ARW-WRF using a modified Tiedtke cumulus parameterization scheme. *Mon. Wea. Rev.*, **139**, 3489–3513, doi:[10.1175/MWR-D-10-05091.1](https://doi.org/10.1175/MWR-D-10-05091.1).