Using Analysis Corrections to Address Model Error in Atmospheric Forecasts

WILLIAM CRAWFORD

Naval Research Laboratory, Monterey, California

SERGEY FROLOV

Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, NOAA/ESRL/PSD, Boulder, Colorado

JUSTIN MCLAY, CAROLYN A. REYNOLDS, NEIL BARTON, AND BENJAMIN RUSTON

Naval Research Laboratory, Monterey, California

CRAIG H. BISHOP

School of Earth Sciences, and Centre of Excellence for Climate Extremes, University of Melbourne, Melbourne, Victoria, Australia

ABSTRACT

This paper illustrates that analysis corrections, when applied as a model tendency term, can be used to improve nonlinear model forecasts and are consistent with the hypothesis that they represent an additive 6-h accumulation of model error. Comparison of mean analysis corrections with observational estimates of bias further illustrates the fidelity with which mean analysis corrections capture the model bias. While most previous implementations have explored the use of analysis corrections to correct forecast biases in short-range forecasts, this is the first implementation of the correction method using both a seasonal mean and random analysis correction for medium-range forecasts (out to 10 days). Overall, the analysis correction–based perturbations are able to improve forecast skill in ensemble and deterministic systems, especially in the first 5 days of the forecast where bias and RMSE in both lower-tropospheric temperature and 500 hPa geopotential height are significantly improved across all experiments. However, while the method does provide some significant improvement to forecast skill, some degradation in bias can occur at later lead times when the average bias at analysis time trends toward zero over the length of the forecast, leading to an overcorrection by the analysis correction–based additive inflation (ACAI) method. Additionally, it is shown that both the mean and random component of the ACAI perturbations play a role in adjusting the model bias, and that the two components can have a complicated and sometimes nonlinear interaction.

1. Introduction

Recent work by Bowler et al. (2017) proposed a novel way to account for model error in an ensemble forecast. In their work, which we will refer to as analysis correction–based additive inflation (ACAI), an estimate of model error is derived from an archive of analysis corrections and added as a tendency term to the model equations every six hours of the model integration. The ACAI method assumes that analysis corrections represent a reasonable approximation to the short-term

model error and on average that such corrections represent an approximation to the model bias. The origins of the ACAI method can be traced back to Saha (1992) where average differences between forecast and analysis states are used to mitigate the development of model biases in deterministic forecasts. In contrast to the application of a systematic error term to correct model biases, Batté and Déqué (2016) illustrate the utility of using additive perturbations derived from randomly sampled model error corrections to address bias in seasonal forecasts of a global coupled model. Piccolo et al. (2019) illustrate the utility of using random analysis increments to account for model uncertainty in ensemble

Corresponding author: william.crawford@nrlmry.navy.mil

forecasts and present a comparison with other stochastic methods (i.e., SKEB and SPT). Their results indicate that the analysis increment method outperforms other methods in terms of ensemble spread and reliability. Both in deterministic and ensemble forecasts, the ACAI method demonstrates added skill and reduced forecast bias in short-term model forecasts. The method introduced by Bowler et al. (2017) combines these prior applications to simultaneously address model bias and model uncertainty.

In this paper, we further examine the applicability of the ACAI method to error reduction in short-term forecasts, but also, for the first time, up to 10-day model forecasts. We investigate the performance of ACAI in three operationally relevant settings: 1) An ensemble prediction system (EPS) based on the ensemble transform (ET) initial conditions (Bishop and Toth 1999; McLay et al. 2010), 2) an EPS system based on an ensemble of data assimilations (EDA) generated by perturbing the assimilated observations [similar to Houtekamer et al. (1996) and Kucukkaraca and Fisher (2006)], and 3) a deterministic forecast system. All systems use the Navy Global Environmental Model (NAVGEM) atmospheric model.

We investigate several properties of the ACAI method. We first evaluate the ability of analysis corrections to approximate short-term model error computed from observations (i.e., radiosondes; which are considered to be minimally biased) followed by a structural description of the mean analysis corrections. We evaluate the impacts of the ACAI perturbations on short and long-term error in three forecast systems listed above, as well as, the individual impacts of the mean and random components of the ACAI method on bias reduction in the ET-based EPS. We conclude by discussing deficiencies of the ACAI method and suggest ways to improve the ACAI method and the analysis system that generates the archive of the perturbations.

2. Methods

a. Numerical model

NAVGEM (Hogan et al. 2014) is used in all of the presented experiments at the T359 resolution (Gaussian grid of 1080×540 grid points, ~ 37 km equatorial resolution) with 60 hybrid pressure levels in the vertical (model top at 0.04 hPa). It is a primitive equation, spectral atmospheric model with a semi-Lagrangian/semi-implicit dynamical core, a comprehensive set of physical parameterizations and an accompanying hybrid 4D-Var data assimilation system (NAVDAS-AR; Kuhl et al. 2013; Rosmond and Xu 2006; Xu et al. 2005). NAVGEM is the

global atmospheric NWP model used operationally by the Fleet Numerical Meteorological and Oceanography Center (FNMOC) for both ensemble and deterministic forecasting.

b. Ensemble transform method

The first ensemble forecast system is based on initial perturbations generated using the ensemble transform (ET) method. Several experiments were conducted using the ET method described by McLay et al. (2008). The ET generates ensemble perturbations through a linear combination of short-term (6-h) forecasts under the constraint of an estimated analysis error variance produced by the NAVDAS-AR data assimilation system. McLay et al. (2010) describe a local formulation of the ET that enables a better fit of the analysis perturbations to the analysis error constraints compared to the global formulation in McLay et al. (2008). The local formulation performs the linear transformations using weighting matrices that are a function of latitude band, and this formulation is employed in the presented experiments. In each of the short 6-h ET-based ensemble forecasts, the model is run 4 times daily at 0000, 0600, 1200, and 1800 UTC. For each initialization, a 20-member ensemble is integrated to the 6-h lead time to produce the next set of ensemble initial conditions via the ET.

c. Ensemble of data assimilations (EDA)

An additional ensemble forecast system is also tested using an EDA-based system. The EDA presented here is initialized using the method of perturbed observations (Houtekamer et al. 1996; Kucukkaraca and Fisher 2006) where perturbations are randomly sampled from the observation error probability density function and added to each variable. The analysis equation for each ensemble member i then becomes

$$x_i^a = x_i^f + \mathbf{K}_i[(y^o + \xi_i) - H(x_i^f)], \quad (1)$$

where x_i^a and x_i^f are the analysis and forecasts states, respectively; $\mathbf{K}_i$ is the Kalman gain matrix; and H is the observation operator; $\xi_i \sim N(0, \mathbf{R})$ are random perturbations to the observations y^o, drawn from the normal distribution with zero mean and variance equal to the observation error variance, $\mathbf{R}$. The ensemble is then comprised of N_e independent data assimilation runs that differ at initial time according to the perturbed observations assimilated by each member. Each ensemble member is run with the 6-h cycling NAVDAS-AR hybrid 4D-Var data assimilation system. Note that the ensemble covariances used in the hybrid 4D-Var come from an ET ensemble (Kuhl et al. 2013) and *not* from the

ensemble of data assimilations. Before testing the EDA perturbations in the Hybrid DA, one would ideally like to add a model error component to the EDA perturbations that ensures their covariances are a good approximation to the forecast error covariance. The ACAI model error study described here is viewed as a stepping-stone toward that goal.

d. ACAI perturbations

We produce the ACAI perturbations using a 1-year archive of analysis corrections from 2015. The archive is generated from a series of deterministic NAVGEM atmospheric reforecasts at the T359 resolution with 60 vertical levels. The reforecasts are initialized from analysis states produced by the NAVDAS-AR system. Because the perturbations will be added to the right hand side of the forecast equations (i.e., as tendencies), the archive is produced from differences between analysis and prior states at each model grid point as

$$\delta x^a = x^a(t) - M[x^a(t - \Delta t^a)], \quad (2)$$

where $x^a(t)$ is an analysis state valid at time t and $M[x^a(t - \Delta t^a)]$ is a 6-h forecast state produced by the nonlinear forecast model M using an analysis state valid at $t - \Delta t^a$, where Δt^a is the length of the assimilation window. This will produce analysis corrections at the full model resolution as opposed to using the true analysis corrections from the DA system, which are computed at a reduced T119 resolution (~110 km equatorial resolution). The reforecasts are initialized every 0000, 0600, 1200, and 1800 UTC producing analysis corrections 4 times per day.

The ACAI perturbations used during the ensemble forecasts are computed in a similar fashion to those used by Bowler et al. (2017) in that they are a combination of a mean analysis correction and a randomly sampled correction drawn from the archive. Perturbations to surface pressure, temperature, humidity, zonal wind speed, and meridional wind speed at each model grid point are computed as

$$\delta x_i = \frac{1}{N_s} \sum_{k=1}^{N_s} \delta x_k^a + \alpha \left(\delta x_{r_i}^a - \frac{1}{N_e} \sum_{j=1}^{N_e} \delta x_{r_j}^a \right), \quad (3)$$

where the first term on the right-hand side represents a 3-month, seasonal average analysis correction centered on the month of the forecast and is aimed at correcting model bias; N_s is the total number of corrections in that time period where a correction is computed every 6 h. The second term is intended to be an added source of stochastic perturbation during the forecast. For each ensemble member i, a random sample is drawn from the same 3-month time period used to compute the seasonal

average and is represented in (3) by $\delta x_{r_i}^a$. N_e is the number of ensemble forecast members, and the mean of the N_e random samples is subtracted from the random sample for each individual ensemble member ($\delta x_{r_i}^a$) before adding to the seasonal average component. Following the work of Bowler et al. (2017), a scaling term, α, is included in order to control the impact of the random perturbations on the forecast and is set to 0.5 in the presented experiments. The random portion of the perturbations is intended to increase the spread of the ensemble forecasts and is applied only in the ensemble forecast experiments. The perturbations computed using (3) are divided by the total number of time steps per 6-h period of the forecast (N_t) and added as tendencies to the model solution at each time step of the integration as

$$\frac{dx_i}{dt} = f(x_i) + \frac{\delta x_i}{N_t}, \quad (4)$$

where $f(x_i)$ is the tendency term of the prognostic equation. A 6-min numerical time step is used in all of the presented experiments giving a total of 60 time steps per 6-h period. A new set of perturbations is computed for each ensemble member for each 6-h portion of the forecast. Figure 1 gives an example of what an ACAI-based tendency term might look like in zonal wind, temperature, and surface pressure for one ensemble member at a random point on the globe over the length of a single 10-day forecast. The total ACAI forcing (mean plus randomly sampled correction) is indicated by the red line and is shown to change every 6 h of the forecast. The mean component of the forcing is shown in black and remains constant over the 10-day forecast. While the second term in (3) is defined by random samples from the archive of analysis corrections, as noted in Piccolo et al. (2019) and indicated in Fig. 1, the perturbations are constant and perfectly autocorrelated (i.e., no decorrelation in time) over a 6-h period. The black line in Fig. 1 would remain the same between all ensemble members; however, the red line would differ between members according to the random sample.

In the deterministic system, only the seasonal average analysis correction [i.e., first term on the rhs of (3)] is used to correct model biases during the forecast. It is a seasonal (3 month) mean analysis correction and is applied as a tendency at each model time step as in (4). In this setting, the random component of the ACAI perturbations is set to zero, and therefore the tendency added will be the same at each time step throughout the entire forecast.

e. List of experiments

We designed our experiments based on three types of forecast systems that are commonly used in operational
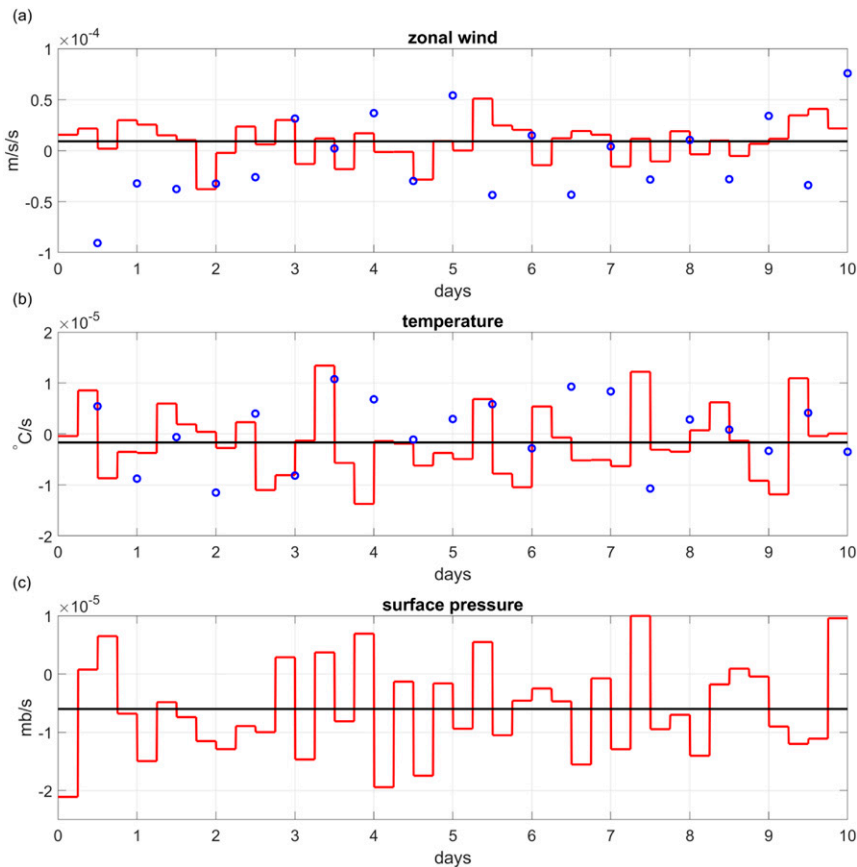
FIG. 1. Example time series of the ACAI tendency perturbations to (a) zonal wind and (b) temperature in the lower troposphere (~950 hPa) and (c) surface pressure at a single point on the globe for one ensemble member over the length of a 10-day forecast. The black line indicates the mean correction, and the red line represents the mean plus a random sample. For reference, (a) and (b) also show the total model tendency in the same location as the perturbation (blue circles).

centers, including the U.S. Navy's operational forecast center [the Fleet Numerical Meteorological and Oceanography Center (FNMOC)]. The prime objective of these experiments is not to compare the skill of the differing forecast systems but to examine the effect of ACAI on each of them individually. In the first system, the ET method is used to create the initial ensemble perturbations. This method tightly controls the average variance of the initial perturbations. In the second system, the EDA method is used to create ensembles of initial conditions. This method uses perturbed observations and the initial spread of the ensemble depends on a multitude of factors including the forecast error covariance model of the data assimilation scheme and the values of the assumed observation error variances. The third system is a standard deterministic forecasting system. Details of each experiment are summarized in the Table 1 with the further rationale described below.

ET-based EPS experiments with and without the ACAI perturbations were conducted for two 1-month periods; a boreal winter (15 December 2016–14 January 2017) and summer (1–31 July 2016) period. At each 0000 and 1200 UTC for each day, a 20-member ensemble is initialized and integrated out to 10 days. All ET-based experiments use stochastic kinetic energy backscatter (SKEB) as an additional source of stochastic error representation. The ET-based ensembles are centered on an external analysis from the NAVDAS-AR hybrid 4D-Var data assimilation system and are not part of a cycling data assimilation system as in the EDA experiments.

In the EDA-based EPS experiments, the same 1-month winter period is used as in the ET-based ensemble. The summer period was not repeated in these experiments due to the similarity between the summertime and wintertime results in the ET-based EPS experiments. Both EDA experiments were run using 5 ensemble members, with 10 day forecasts issued at 0000 and 1200 UTC. While the EDA-based EPS is not currently used operationally, the

TABLE 1. List of experiments. ET/EDA experiments: Winter (15 Dec 2016–14 Jan 2017); Summer (1–31 Jul 2016). DET experiments: Winter (15 Dec 2016–31 Mar 2017). N/A: not applicable.

Name	Season	Ensemble initialization	Model error	ACAI
ETctrl	Winter/summer	Ensemble transform	SKEB	N/A
ETacai	Winter/summer	Ensemble transform	SKEB and ACAI	$\alpha = 0.5$
EDActrl	Winter	Perturbation to obs	None	N/A
EDAacai	Winter	Perturbation to obs	ACAI	$\alpha = 0.5$
ETacai0	Summer	Ensemble transform	SKEB and ACAI	$\alpha = 0$; bias only
ETacaiR	Summer	Ensemble transform	SKEB and ACAI	$\alpha = 0.5$; no bias
DETctrl	Winter	N/A	None	N/A
DETacai	Winter	N/A	ACAI	$\alpha = 0$; bias only

Navy is developing a global coupled ensemble forecasting system that uses an EDA configuration, motivating an investigation on the impacts of the ACAI method on an EDA type system. It is anticipated that the global coupled ensemble system will likely be run with fewer members than the operational ET-based system, motivating the EDA-based system to be tested with a reduced number of members.

To investigate the contribution of the mean and random components of the ACAI perturbations on the forecast, we rerun the summer ET-based experiments with either the random (Table 1: ETacai0) or the bias (Table 1: ETacaiR) component set to zero.

The initial conditions for the deterministic runs are provided by a 6-hourly cycling hybrid data assimilation system, where 5-day forecasts are initialized every 0000 and 1200 UTC. The deterministic experiments span the wintertime period from 15 December 2016–31 March 2017. Note that since our interest is in comparing the three systems performances with and without ACAI, our use of differing time periods and forecast integration time for each experiment does not affect this comparison.

In all ACAI-based experiments, the perturbations are added throughout the entire length of all short and extended-range forecasts. In the cycling data assimilation systems (i.e., EDA and deterministic), because the perturbations are included starting at analysis time, the perturbations influence the background of the subsequent cycle potentially influencing the skill metrics at initial time.

3. Results

a. Comparison of analysis corrections with an observational estimate of bias

As described above, a key motivation for ACAI is the *assumption* that seasonally averaged analysis corrections are an approximation to the model bias. To examine the validity of this assumption more closely, we compare an estimate of the bias based on radiosonde observations with the mean analysis corrections. Figure 2
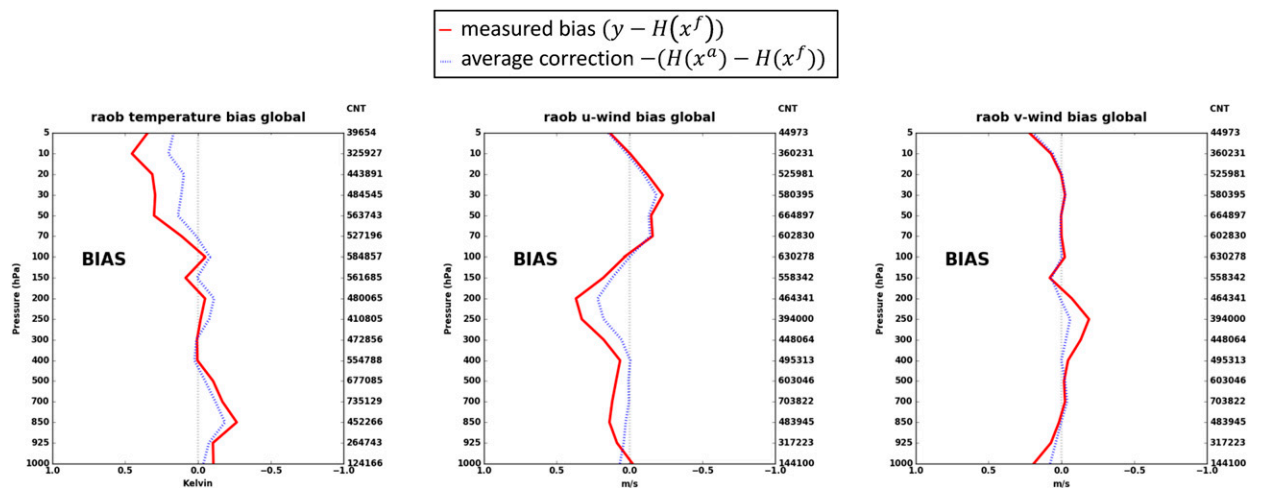

FIG. 2. Global mean profiles of background departures from radiosondes (red) and negative of analysis corrections at observation locations $[H(x^a) - H(x^f)]$ (blue dashed) averaged over 15 Dec 2016–31 Mar 2017.

shows globally averaged profiles of 6-h forecast departures from radiosonde observations over the period 15 December 2016–31 March 2017. The radiosonde profiles are assumed unbiased, and therefore, the departures are assumed to present a reliable estimate of the bias in the background. Figure 2 also shows globally average profiles of analysis corrections at the same locations as the radiosonde profiles (which are primarily over the continental regions) over the same time period. The average analysis corrections represent a negative of the bias, and therefore the sign has been reversed in Fig. 2. The two lines track well and show that the analysis corrections generally capture the sign of the bias. However, the average analysis corrections consistently underestimate the magnitude of the bias. Dee and Da Silva (1998) discuss this point from a theoretical standpoint and show that when forecast errors and observations errors are about equal, the analysis corrections underestimate the bias by 1/2. It can, therefore, be expected that while the average analysis corrections will reduce the forecast bias, it will not remove it entirely. It should be noted that while the mean analysis corrections match well with the bias captured from radiosondes, this does not necessarily mean the analysis corrections are representative of the bias throughout the entire forecast. To illustrate, Fig. 3 shows the bias in 500 hPa geopotential height as a function of forecast lead time in the northern extratropics, southern extratropics, and tropics (regions defined in appendix). The bias in northern extratropical geopotential height is fairly stable with a positive bias throughout the entire forecast. The southern extratropical bias, on the other hand, maintains a fairly consistent positive bias through time, but decreases in amplitude and even switches sign slightly by the end of the forecast. Bias in the tropics decreases rapidly through the first 5 days of the forecast, switches sign and is of equal negative amplitude by the end of the forecast. Figure 3 illustrates that while the bias can be fairly stable in one region, bias in another region may evolve in a very different way. This complex evolution of bias can be seen in other variables as well, and illustrates that a short-term estimate of bias may not represent the true model bias at longer lead times.

b. Structure of the average corrections

Before we examine the impact of ACAI on the model forecasts, we find it instructive to examine the statistics of the analysis corrections. Figure 4 shows an average analysis correction to surface pressure over the winter (DJF) and summer (JJA) time period of 2015. Average analysis corrections are computed as the seasonal (3-month) mean of the corrections computed in Eq. (2). In Fig. 4, red means that the 6-h forecast of surface pressure is too low
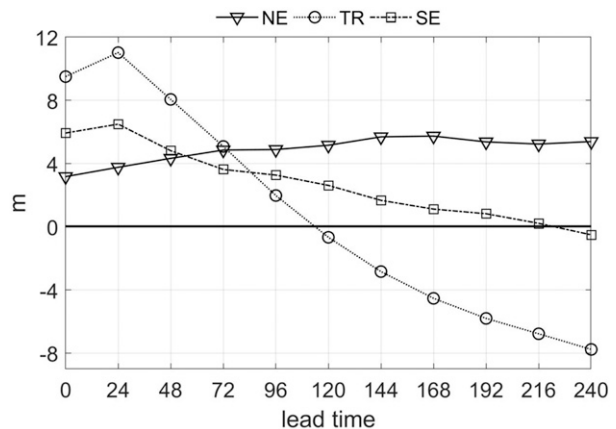

FIG. 3. Ensemble mean bias of 500 hPa geopotential height as a function of forecast lead time (hours) in the northern extratropics (NE; triangles), tropics (TR; circles), and southern extratropics (SE; squares) in the ET-based control experiment (ETctrl).

whereas blue means that it is too high. The mean correction shows that, on average, the surface pressure of the 6-h NAVGEM forecast is too low in the tropics and too high in the extratropics. Also shown in Fig. 4 are contours of the seasonal mean surface pressure, which indicate a meridional pressure gradient from the tropics to the midlatitudes. The average analysis correction will then act to decrease the relatively high pressure in the midlatitudes and increase the low pressure in the tropics, potentially modifying the strength of the meridional pressure gradient and the associated pressure driven flow (i.e., easterly trade winds). There is also some evidence of a land–sea gradient in the correction to pressure, particularly in the tropical regions, which is consistent with Bhargava et al. (2018) where they examine analysis corrections derived from GFS. However, in their case, the land–sea gradient in pressure appears to be the dominant structure of the average pressure corrections. Figure 4 illustrates little seasonality to the general structure of the mean pressure corrections, which is also consistent with the findings of Bhargava et al. (2018). However, the largest negative increments to surface pressure occur in the winter hemisphere with the Northern (Southern) Hemisphere having largest corrections in DJF (JJA).

As in the case of the mean surface pressure corrections shown in Fig. 4, zonally averaged profiles of temperature, specific humidity, zonal velocity and meridional velocity (Fig. 5) exhibit a general consistency in the structure of the increments between seasons, particularly in the case of specific humidity. The zonally averaged temperature (T) correction indicates that the 6-h forecast is generally too warm in the lower troposphere with some slight variation above ~600 hPa. A slight cold bias is also
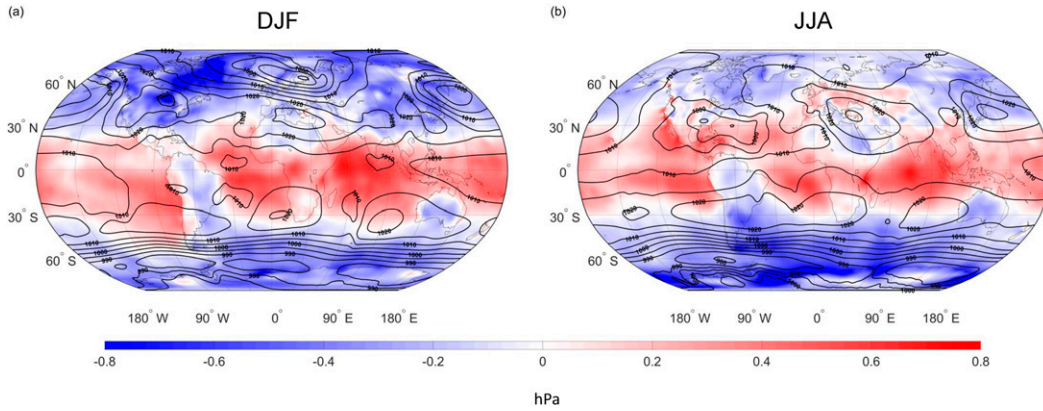
FIG. 4. Mean (a) winter (DJF 2015) and (b) summer (JJA 2015) analysis corrections to surface pressure (color). Seasonal averages of analysis surface pressure (contours).

evident in the upper troposphere ($\sim$200 hPa) of the summer hemisphere. Specific humidity (Q) contains a dry bias in the tropical regions below 600 hPa with the largest magnitude adjustments focused along the equator. Figures 5c,g and 5d,h show the mean analysis correction for the zonal (U) and meridional (V) component of the wind velocity with contours of the average total analysis velocity overlaid. The mean flow contours in Figs. 5c and 5g illustrate the easterly component of the surface level trade winds with negative (dashed line) contours between 30°S and 30°N. In Figs. 5c and 5g, red (blue) colors indicate a correction that increases the westerly (easterly) component of the wind. Since the mean correction is increasing the westerly component in a region where the mean is easterly, the mean correction is actually decreasing the strength of the easterly trade winds. As in the case of the mean increment to surface pressure shown in Fig. 4, the largest magnitude correction to the surface level winds occurs in the winter hemisphere. In Figs. 5d and 5h, red (blue) colors indicate

corrections that increase the northward (southward) component of the meridional wind. The low-level mean meridional wind contours on Figs. 5d and 5h are consistent with Hadley's simple model of low-level tropical convergence. Specifically, it indicates a low-level flow toward the equator between 30°S and 30°N and below $\sim$600 hPa, and the upper-level flow away from the equator at 200 hPa at these same latitudes. While the mean correction to the meridional winds exhibits a complex structure, many of the largest amplitude adjustments occur in the vicinity of the upper and lower branches of this Hadley-like circulation, particularly in the Northern Hemisphere. Comparison of the mean analysis and prior (or background) state (not shown) indicates the large amplitude positive and negative adjustment to the upper-level divergent flow centered at 15°N and 200 hPa is primarily due to a downward shift in altitude of the northward flow. Similarly, the adjustment (centered at 15°N and 700 hPa) to the equatorward flow is primarily caused by an upward and northward shift in
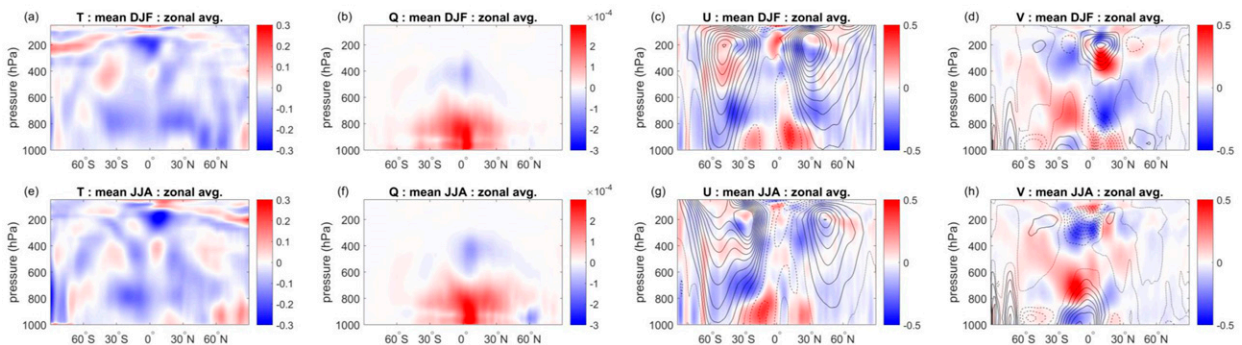

FIG. 5. Zonal averages of mean (a)–(d) winter (DJF 2015) and (e)–(h) summer (JJA 2015) analysis corrections to (a),(e) temperature (°C), (b),(f) humidity (kg kg^{-1}), (c),(g) zonal wind speed (m s^{-1}), and (d),(h) meridional wind speed (m s^{-1}). Contours on (c), (d), (g), and (h) represent mean analysis velocity. Positive (negative) direction indicated by solid (dashed) lines. Contour interval is 3 (0.5) m s^{-1} in U (V).
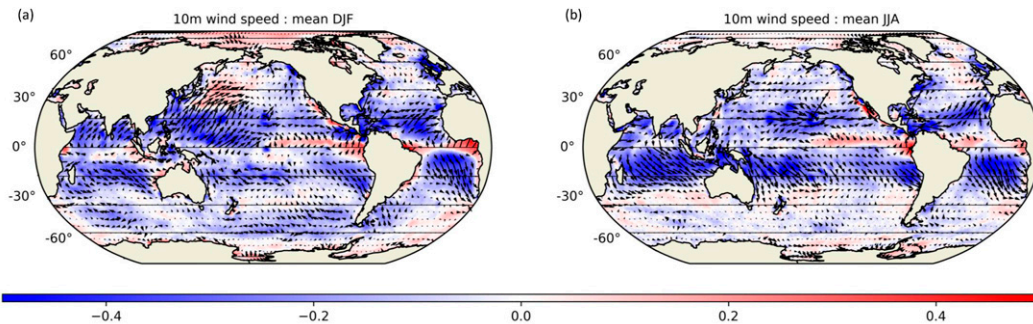
FIG. 6. Mean (a) winter (DJF 2015) and (b) summer (JJA 2015) analysis corrections to 10 m wind speed. Direction (vectors) and magnitude/sign (color, m s^{-1}).

the equatorward flow. Figures 5d and 5h show that the Hadley-like circulation strengths in the winter hemisphere, as does the magnitude of the corrections to the circulation. Simpson et al. (2018) found similar structures in the boreal winter average of analysis corrections to the wind in the ERA-Interim reanalysis (Dee et al. 2011). However, while Simpson et al. (2018) shows a coherent increase in the convergent portion of the Hadley cell by the mean analysis correction to V (surface–600 hPa), the mean NAVGEM correction indicates a sign reversal below ~900 hPa with the mean near-surface correction acting to reduce convergence (Figs. 5d,h).

Figure 6 shows maps of seasonally averaged 10 m wind speed corrections (magnitude and direction). The direction of the mean correction is in opposition to the direction of the tropical trade winds, also evident as a negative correction to the wind speed magnitude. There is also some evidence for average increases to the wind speed, particularly along the equator in the eastern Pacific and Atlantic Ocean and in the northeastern North Pacific. The magnitude of the corrections the winds speed is again largest in the winter hemisphere.

c. Impact of ACAI on short-term forecast scores

Here we compare the performance of forecasts using the ACAI perturbations to control experiments without the perturbations. All metrics, regions and variables are defined in the appendix. We find that ACAI has an overall positive impact on short-term (0–5 days) forecast scores in all experiments: ETacai, EDAacai, and DETacai. All three systems see reduction in both bias (Fig. 7, top) and the RMSE (Fig. 7, bottom). Changes in bias are computed as

$$100 \times \frac{|\text{experimental bias}| - |\text{control bias}|}{|\text{control bias}|},$$

where | | represents the absolute magnitude of the bias. Differences in RMSE between the experiments are also

computed as percent changes. The maximum change denoted in bias and RMSE in Fig. 7 are 100% and 10%, respectively; however, the actual change in RMSE may be larger than 10% and the degradation in bias may be larger than 100%. In Fig. 7, a 100% improvement in bias scorecard indicates the average bias has been forced to zero in the ACAI-based forecast. The largest positive impact is on 500 hPa geopotential height (Z500) scores and lower-tropospheric temperatures (T850). The global 5-day average reduction in Z500 bias in the three systems (ETacai, EDAacai, and DETacai) is 37%, 35%, and 23%, respectively. For T850, the reduction for the 5-day average bias is 50%, 49%, and 27%. The average global improvement to RMSE in Z500 (T850) is 6%, 7%, and 3% (3%, 6%, and 4%). Improvements to Z500 and T850 are consistent across all experiments and regions, except in the deterministic system where tropical biases are only marginally improved. Tropical surface conditions are also improved across all experiments with a reduction of bias in 10 m wind speeds (W10m: 17%, 24%, and 31%) and 2 m air temperature (T2m: 9%, 29%, and 10%). The RMSE of W10m is reduced by ~3% in each experiment and T2m is reduce by ~2% in the ETacai and DETacai experiments, while the EDAacai experiment saw a much large reduction (~7%). Note, the perturbations used in the ETacai and EDAacai experiments contain both a mean and random component, whereas the DETacai experiment perturbations only contain the mean correction which is likely responsible for the generally larger impact of the ACAI perturbations in the ensemble systems.

Some variables did show significant degradation in scores in the first 5 days, particularly in the case of the EDA-based system. Figure 7e shows that northern extratropical 2 m air temperature RMSE scores are degraded by ~6% and Fig. 7b indicates a substantial increase (~100%) occurred in southern extratropical 10 m wind speed bias. However, the bias in southern extratropical 10 m wind speed in the EDActrl experiment is only
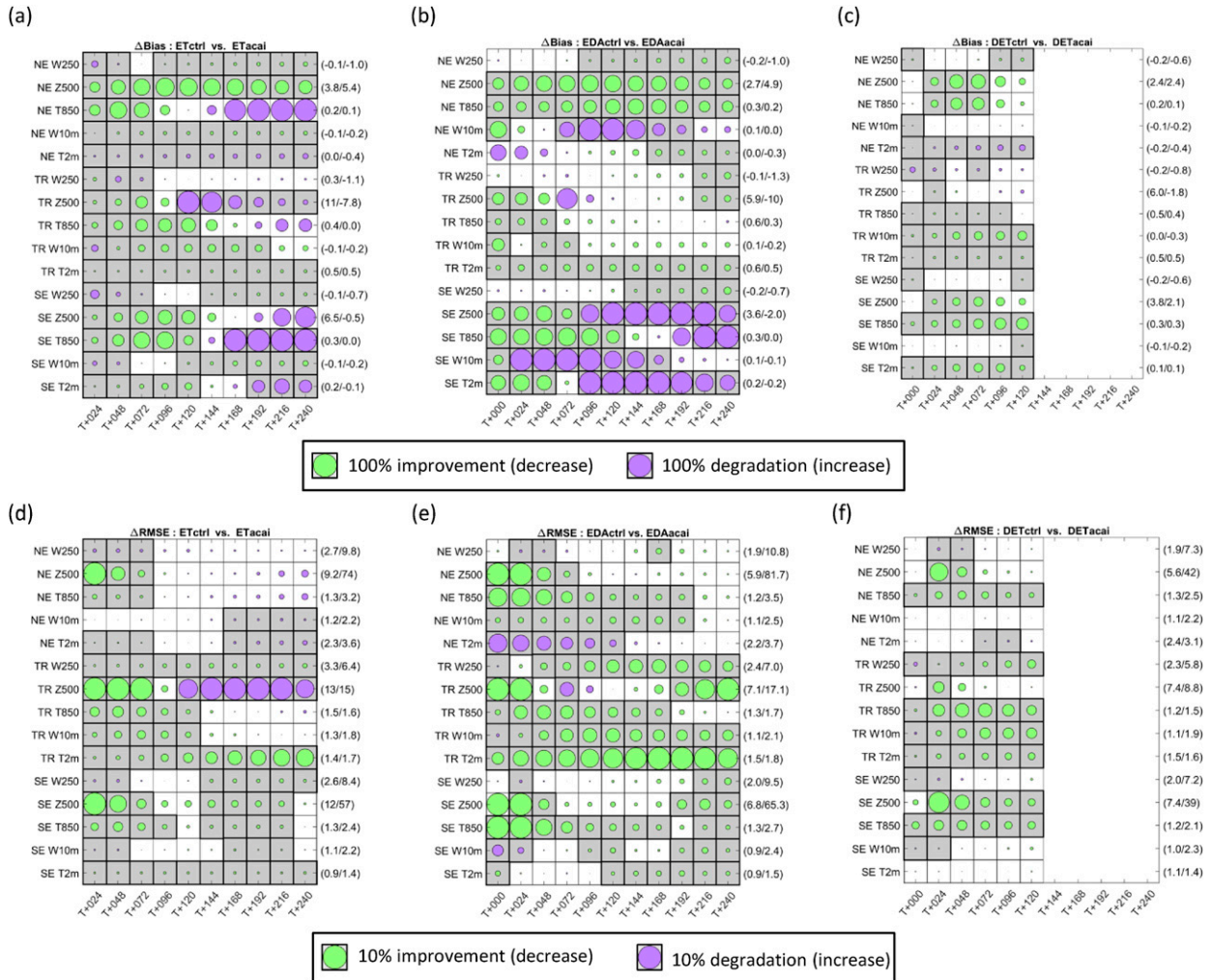
FIG. 7. Scorecards showing percent change in (top) bias and (bottom) RMSE for (left) ETacai, (middle) EDAacai, and (right) DETacai for winter 2017 period compared to control experiments. Green (purple) circles represent improvement (degradation). Gray shading and bold outline represents statistical significance at the 95% level. Circle size maximum for bias (RMSE) is 100% (10%). Scores for individual variables (see the appendix) shown on the vertical axis with forecast lead time on the horizontal axis. For context, the true values of each metric from the control experiment at initial and final lead time are shown at the outside right of each line of the scorecard. All scores computed against ECMWF analyses.

~0.02 m s^{-1} on average over the first 5 days, therefore this only represents a modest increase in wind speed bias. Overall, degradations in ETacai are not as great as for EDAacai while DETacai is almost free of degradations. The relative impact of each component of the ACAI perturbations (mean and random) on bias correction in the ensemble systems is explored further in section 3e.

d. Impact of ACAI on extended-range forecast scores

We find that in the extended-range forecasts (days 5–10), the ACAI perturbations continue to improve RMSE scores for most variables/experiments. All EDAacai RMSE scores are improved by an average of 3% across all variables/regions with the largest impact

occurring in the tropics (Fig. 7e). In the ETacai experiment, RMSE scores continue to be improved for most SE and TR variables (except for TR Z500), but are degraded for most of the NE variables (though several of these are shown to fall below the 95% significance threshold).

We also find that several of the bias scores that are positive in the short range, are degraded in the extended range. For example, in the ETacai experiment, T850 biases in all regions switch from improved to degraded after day 5. A similar switch from improved to degraded biases is also present in TR and SE Z500. While many of the large amplitude degradations in bias fall below the significance threshold, a near equivalent number do not,
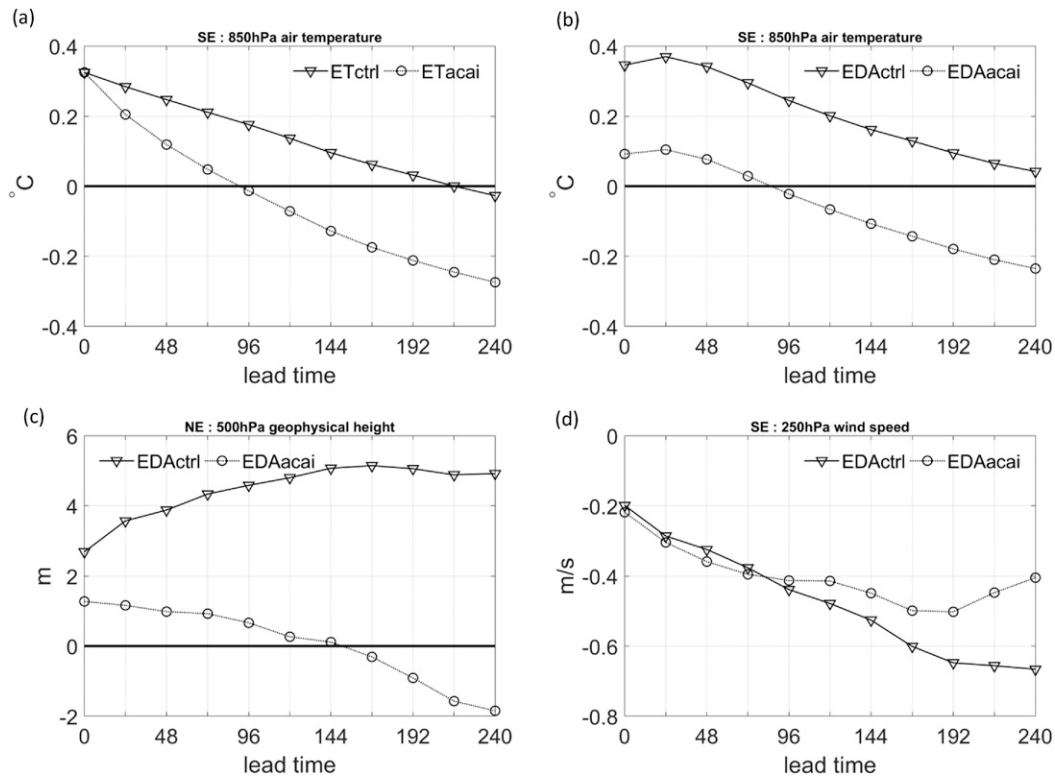
FIG. 8. Ensemble mean bias as a function of forecast lead time (hours) for the control (triangle) and ACAI-based (circle) experiments in the ET-based [(a) southern extratropical 850 hPa air temperature] and EDA-based [(b) southern extratropical 850 hPa air temperature, (c) northern extratropical 500 hPa geopotential height, and (d) southern extratropical 250 hPa wind speed] systems.

warranting further exploration. Figure 8a shows lower-tropospheric (850 hPa) bias in the southern extratropics for the ETctrl and ETacai experiment. The figure illustrates that the ACAI perturbations are improving the temperature bias as intended in the first half of the forecast; however, the bias is being pushed through zero increasing the magnitude of the bias at later lead times. We find that this mechanism is responsible for most of the degradations in bias at later lead times of the ET-based experiments.

With the exception of NE 10 m wind speed, NE and TR biases in the EDAacai experiment continue to be improved relative to the control in days 5–10. However, there is a significant degradation in bias for the SE region. The cross-over from improved to degraded biases is similar to that experienced in ETacai; however, the problem experienced in the ETacai experiment illustrated in Fig. 8a is exacerbated by the fact that the initial time biases are improved in the EDAacai experiment (Fig. 8b). The EDA experiments are based on cycling of individual members, and since the ACAI perturbations are added starting at analysis time, the perturbations have the potential to affect the background of the subsequent cycle and reduce the bias at initial time. In contrast, ET

initial biases are fixed because the ET ensemble is centered on an external analysis that did not use ACAI. The reduction in the area-averaged bias at initial time places the EDAacai ensemble closer to zero, and therefore the trend toward negative bias results in a degradation at later lead times.

It should be noted that while Fig. 7b indicates improvement in the bias metric for SE T850, comparison with Fig. 8b indicates that ACAI is doing very little to slow down the cold-bias tendency in the model. In which case, the positive impacts to bias shown in the scorecard are due almost entirely to the impact ACAI has at initial time and less to do with the reduction of bias throughout the forecast. This behavior appears most pronounced in the southern extratropical temperature and geopotential height fields. On the other hand, ACAI appears quite capable of altering bias trends in other region/variable combinations as shown by the changes to bias throughout the forecast in NE Z500 and SE W250 in Figs. 8c and 8d, respectively.

e. Impact of ACAI on the ensemble spread

We find that ACAI improves the ensemble variance over ensemble mean squared error (MSE) ratio (σ^2/MSE)

of the EDAacai relative to the control for all combinations of variable/region/lead time. Increases in spread (Fig. 9b) are as large as 30%–50% for many northern extratropical and tropical variables out to ~6 days and this increase in the spread of the EDAacai experiment contributes to an improved spread–skill ratio of the EDAacai (Fig. 9d). Note, changes in the MSE can also influence the spread–skill relationship and should be considered in conjunction with ensemble spread when looking at changes in variance ratio (VARR) shown in Figs. 9c and 9d. While ACAI does contribute significantly to the spread of the ensemble, as described in Bowler et al. (2017), ACAI alone is insufficient to achieve skillful ensemble spread statistics for the EDAacai experiment and has therefore been combined with "relaxation-to-prior" methods to further improve the spread–skill relationship. Initial testing using the method of relaxation to prior perturbations (RTPP; Zhang et al. 2004; Whitaker and Hamill 2012) in our EDA-based EPS results in further improvement in spread–skill of the ensemble (not shown). The combination of RTPP and ACAI is particularly useful in generating additional spread in the EDA system since relaxing toward the forecast (or prior) perturbations will allow some of the variance introduced by ACAI in the first 6 h of the forecast to affect the subsequent cycle.

ACAI has a considerably smaller impact on the spread of the ET ensemble (Fig. 9a; Note, the maximum representable value of change in spread for Fig. 9a is 5%, compared to 50% in Fig. 9b.) In fact, while the impacts are quite small (-1% on average), the spread in many variables is actually reduced by ACAI. We attribute the lack of additional growth in spread of the ETacai ensemble to the amplitude and dynamical conditioning of the ET initial conditions. Not only are the perturbations added by the ET much larger than those added by ACAI, the ET is specifically formulated to isolate the fastest growing modes of variability whereas ACAI is not. For these reasons, the perturbations added by ACAI will likely be overwhelmed by the perturbations added by the ET. Slightly less clear is the *reduction* of spread in the ETacai experiments; however, a likely reason for this is the consistency guaranteed by the ET between the ensemble perturbation variance and analysis error variance across the entire state vector. Localized modification of the forecast variance by ACAI can lead to an adjustment of the ET perturbations between experiments, potentially in a manner that reduces the ensemble variance.

f. Impact of individual ACAI components on bias

As described in section 2a, the additive perturbations defined by (3) are comprised of two terms; a seasonal average analysis correction and a random component derived from a random sample from the analysis correction archive. To investigate the relative contributions of each term on the correction of bias, additional tests were run using only the seasonal average or the randomly sampled correction. These experiments were conducted for the summer period only using the ET system and are listed in Table 1 as ETacai0 (bias only; $\alpha = 0$) and ETacaiR (random only). We find that the relative impact of each component is variable dependent. Figure 10 illustrates the reduction of bias at day 10 in each experiment relative to the control (ETctrl) and shows that the random component of the ACAI perturbations is more effective at reducing bias in 850 hPa (~50%) and 2 m air temperature (~28%), while the mean component is more effective in 250 mb (~55%) and 10 m wind speeds (~50%). It should be noted that, as indicated by Fig. 1, the random component of the ACAI perturbations are often much larger than the mean component. This is illustrated by the fact that, in Fig. 1, the distance between the horizontal black line and zero is often much smaller than the distance between the black and varying red line. Following the work presented in Berner et al. (2017), strong additive noise can be expected to adjust the mean state of a nonlinear system, and therefore, given the relative amplitude of the random portion of the ACAI perturbations, one would expect some effect on the mean state.

It is less obvious that the random component of the ACAI perturbations would affect the bias in the same way as the mean component of the ACAI perturbations. As demonstrated in Fig. 8a, when the baseline system contains a bias at initial time and tends toward zero over the length of the forecast, the ACAI perturbations often end up overcorrecting the bias resulting in an increase in the magnitude of the bias at later lead times. This concept is further illustrated in Fig. 11a, which shows NE Z500 bias in the four summertime ET experiments (ETctrl, ETacai, ETacai0, and ETacaiR). The baseline ET experiment (ETctrl) and the full implementation of ACAI (ETacai) are shown by the triangles and circles, respectively. With the addition of the ACAI perturbations containing both the mean and random component, the bias is driven through zero resulting in a substantially larger bias at the 240-h lead time. Also shown in Fig. 11a are the Z500 biases in the experiments using only the bias (squares) and random (diamonds) components of the ACAI perturbations. While both the mean and random component of the ACAI perturbations are contributing to the effect of running through zero in the ETacai experiment (circles), it is clear that the random perturbations are having a larger effect than the mean component. Thus, while the random portion of
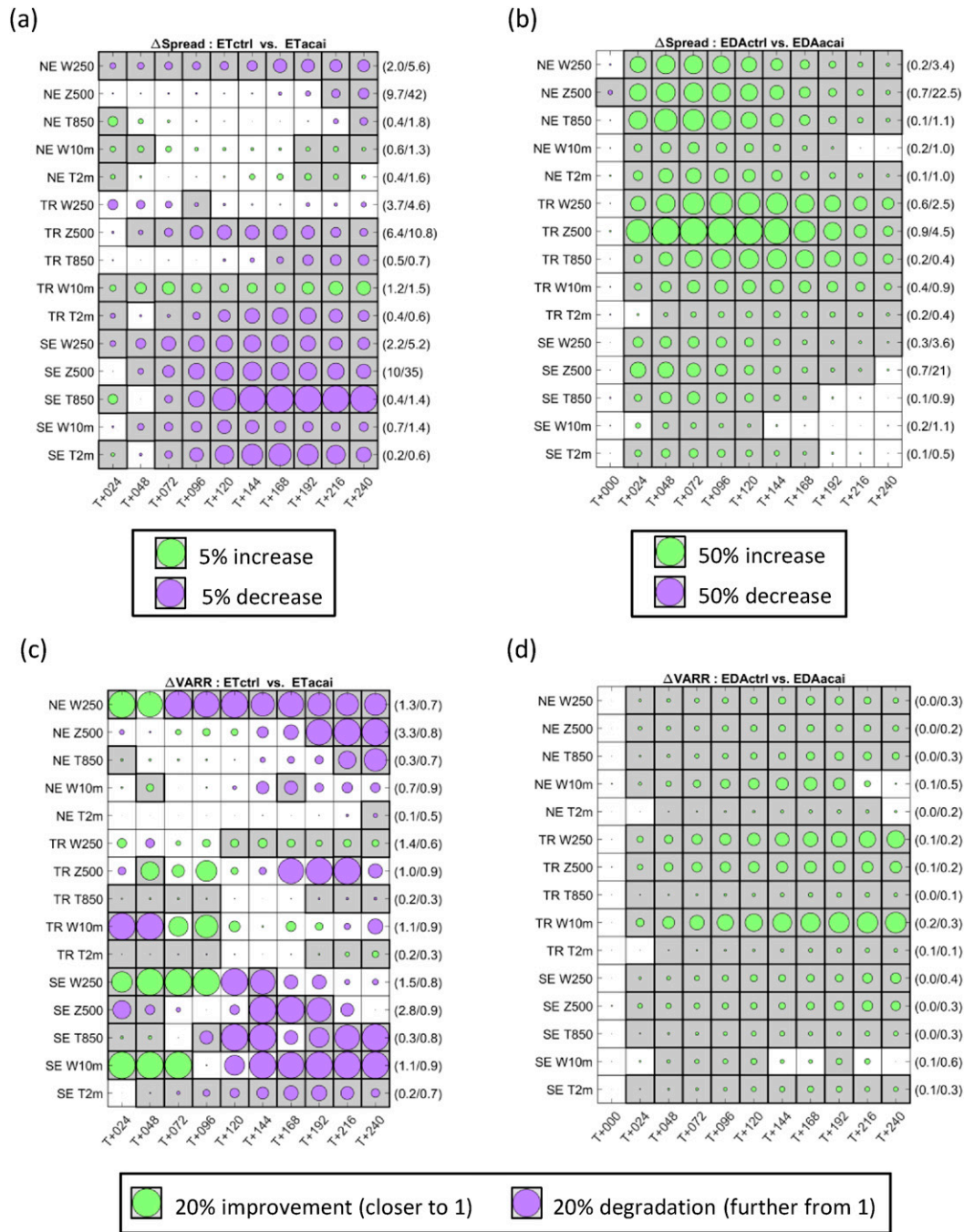
FIG. 9. Scorecards showing percent change in (top) ensemble spread and (bottom) variance ratio (σ^2/MSE) for (left) ETacai and (right) EDAacai for winter 2017 period compared to control experiments. Green (purple) circles represent improvement (degradation). Gray shading and bold outline represents statistical significance at the 95% level. Circle size maximum for ETacai (EDAacai) spread is 5% (50%). Circle size maximum for VARR is (20%). Scores for individual variables (see appendix) shown on the vertical axis with forecast lead time on the horizontal axis. For context, the true values of each metric from the control experiment at initial and final lead time are shown at the outside right of each line of the scorecard.
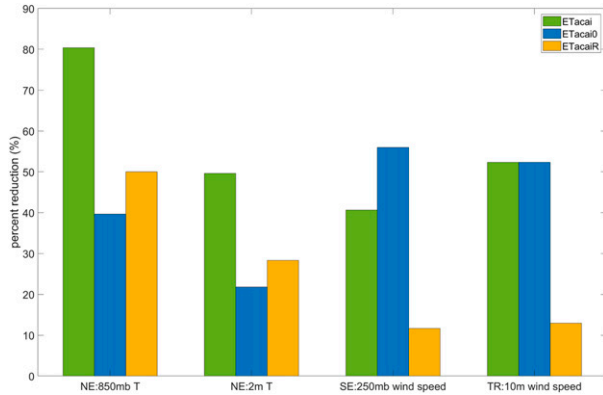
FIG. 10. Percent reduction in day-10 bias in ETacai (green), ETacai0 (bias only; blue), and ETacaiR (random only; orange) experiments. Reduction computed as change relative to ECMWF analyses from ETctrl to ACAI-based experiments.

the perturbations has the potential to have a larger positive impact on the bias (Fig. 10), they also have the potential to contribute to an overcorrection of the bias (Fig. 11a).

In the case of Z500 bias shown in Fig. 11a, the mean and random component of the ACAI perturbations are correcting the bias in the same direction; however, the relative impacts on tropical T850 bias shown in Fig. 11b illustrates that this is not always the case. Here the mean and random components are affecting the bias in opposing directions with the mean component (squares) acting to increase the bias relative to the control (triangles). Interestingly, the combination of the two components (circles) results in the smallest T850 bias overall. Figure 11 illustrates the complicated (and sometimes nonlinear) interaction between the mean and random components of the ACAI perturbations on correcting the bias and suggests more effort needs to be place on understand this interplay in order to arrive at the best possible implementation.

g. Impact of ACAI on mean correction in the deterministic system

Figures 7c and 7f indicate a consistent reduction in bias and RMSE by the bias only component of the ACAI perturbations applied in the deterministic system. In the extratropics, the reduction in bias is most pronounced in Z500 and T850. Figure 12 gives an illustration of the reduction of Z500 bias by showing the actual bias in the DETctrl and DETacai experiments. The largest reduction in bias occurs in the northern and southern extratropics and can be on the order of 10 m at days 3 and 5. Figure 7c indicates that in the tropical region, the largest reduction in bias is for 10 m wind speed at all forecast lead times. To illustrate the localized
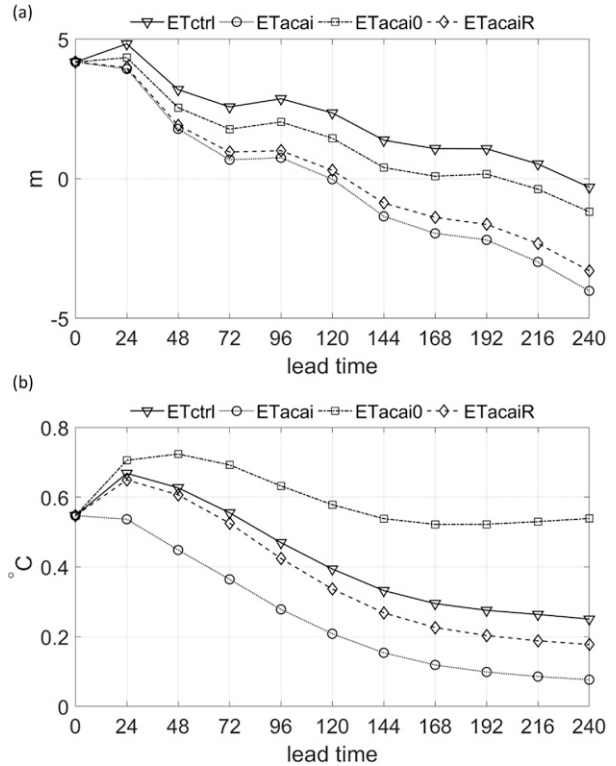

FIG. 11. Ensemble-mean bias as a function of forecast lead time (hours) in the (a) northern extratropical 500 hPa geopotential height and (b) tropical 850 hPa temperature in the ETctrl (triangle), ETacai (circle), ETacai0 (bias only; square), and ETacaiR (random only; diamond) EPS experiments.

effects of ACAI on 10 m wind speed bias, Fig. 13 shows the difference in the absolute magnitude of 10 m wind speed bias in the DETctrl and DETacai experiments, and shows that the reduction is mostly centered in the tropical regions and can be reduced up to $\sim 1.25\,\mathrm{m\,s^{-1}}$.

Last, it is expected that if the application of the average analysis correction during the forecast is indeed reducing bias over the first 6-h period of the integration, then the resulting prior for the next cycle should be closer to the observations compared to the control case. We would then expect the data assimilation system to produce smaller corrections to the prior on average. Figure 14 shows globally averaged profiles of the corrections made to temperature, specific humidity, zonal winds (U), and meridional winds (V) over the entire extent of the experimental period, and indicates an across the board reduction in the magnitude of the corrections made by the DETctrl case versus the DETacai case. The magnitude of the corrections is reduced by $\sim 19\%$ on average with a maximum reduction of $\sim 30\%$–35%.

As noted in section 2e, the time period of the deterministic runs are extended relative to the ensemble runs; however, the results remain relatively unchanged if
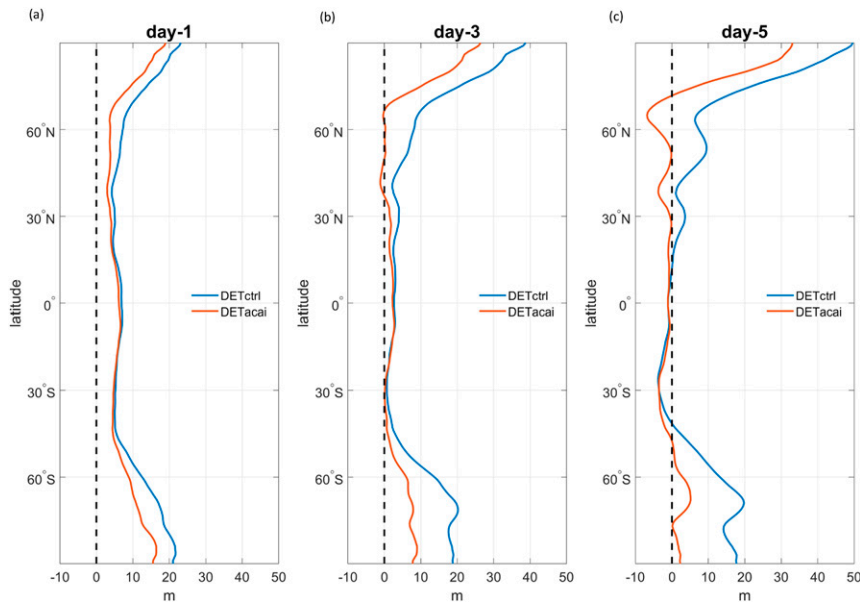
FIG. 12. Zonal averages of 500 hPa geopotential height bias in meters at days 1, 3, and 5 from
deterministic DETctrl (blue) and DETacai (red) experiments.

analysis of the deterministic results are performed over the same time period as the ensemble runs but are less statistically robust.

4. Summary and conclusions

Much of the presented analysis is predicated on the idea that mean analysis corrections are a valid approximation of model bias. Our comparison of estimated model biases from analysis corrections with measured bias computed using radiosondes suggests that average analysis corrections correctly capture the sign of the bias for 6-h forecasts, but that they are generally an underestimate of the true bias. This is in agreement with the theoretical result first shown by Dee and Da Silva (1998) and suggests that the relationship between model bias and mean analysis

corrections is fairly insensitive to model and data assimilation scheme. It should be noted that while the average analysis corrections match well with the observed bias, the similarity is representative of short-term (6 h) model bias and may not match the model biases at extended lead times. This point is demonstrated in Fig. 3, which shows that model bias can change in amplitude or even sign as a function of forecast lead time. In which case, the bias correction component of the ACAI perturbations may not be valid for some combinations of region, variable and forecast lead time.

Despite the apparent underestimate of the systematic error by average analysis corrections, we find that ACAI improved short-term (0–5 days) bias scores in all three experiments. In the ensemble-based experiments, we
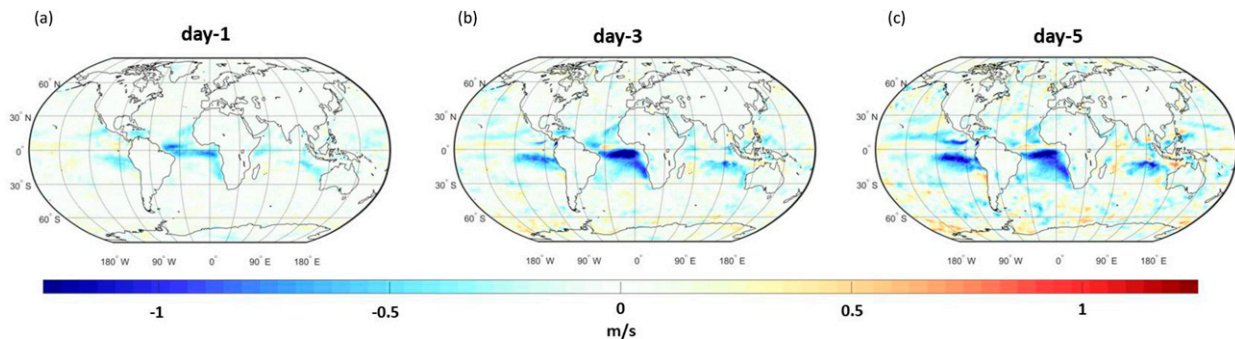

FIG. 13. Difference in magnitude of 10 m wind speed bias in deterministic forecasts at day 1, 3, and 5 |DETacaibias| − |DETctrlbias|. Blue (red) colors indicate an improvement (degradation) to the bias.
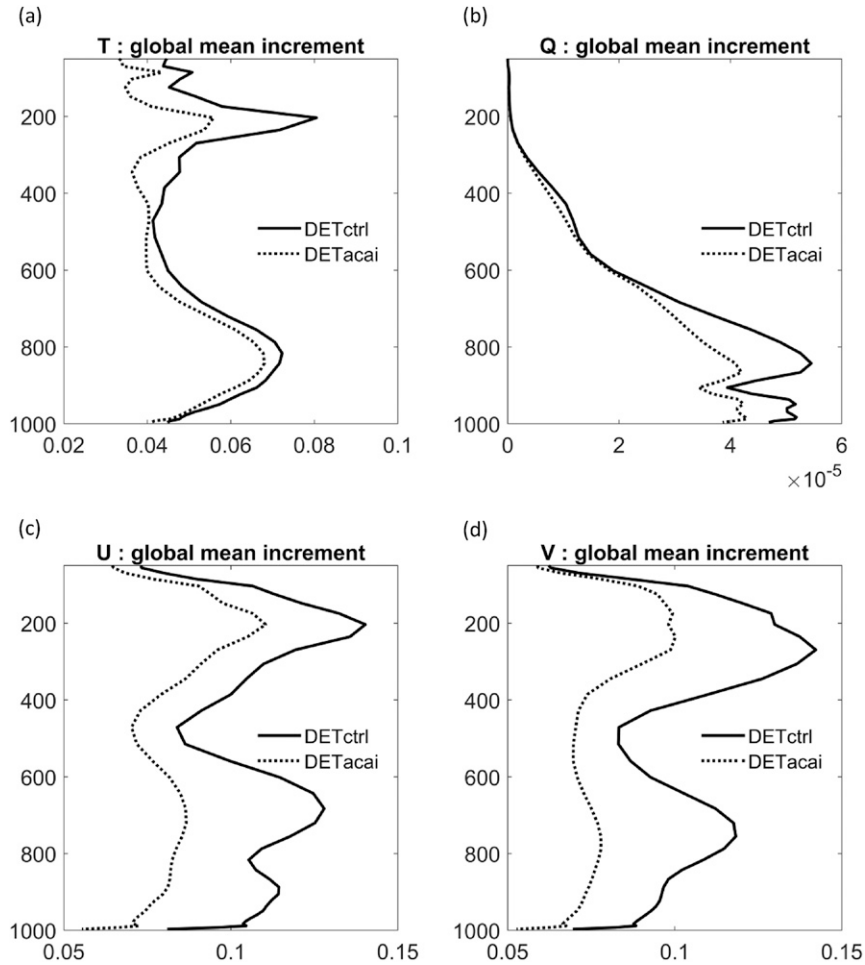
FIG. 14. Global mean profile of absolute magnitude of temperature (T), humidity (Q), zonal velocity (U), and meridional velocity (V) analysis corrections. DETctrl (solid); DETacai (dotted).

find that rectification of the mean error by the random portion of the ACAI perturbations is as important as the mean correction for temperature and geopotential scores, but is not as important for wind scores. The random portion of the ACAI perturbations is shown to be as large as or larger than the mean component, helping to explain the impact of the random portion on the mean state. The addition of ACAI also leads to a reduction of the RMSE over the entire length of the ensemble (10 days) and deterministic (5 days) forecasts for most variables and regions. The ACAI perturbations had relatively little impact on the ensemble spread in the ET system; however, the result may be somewhat anticipated given that the ET-based EPS initializes from dynamically conditioned initial conditions that have an amplitude and growth characteristics that can easily overwhelm any additional growth added by the ACAI perturbations. On the other hand, the ACAI perturbations

are shown to have a significant positive impact on the spread–skill of the EDA system, which has no prior method to account for model uncertainty, as well as, being run with unconditioned ensemble members, further explaining the disparate impacts of ACAI on the ET and EDA systems.

We also document shortcomings of ACAI in its present form. Specifically, we find that ACAI sometimes overcorrected bias after ~5 days of the forecast, leading to the degradation of some of the NWP scores at 10 days. It is shown that both the mean and random portion of the ACAI perturbations contribute to this effect, with the random portion likely playing a larger role. Some portion of the bias degradation is likely also due to a mismatch in the sign of the bias at early versus later lead times causing the ACAI perturbations to increase the magnitude of the bias at the later lead times. In order for the ACAI method to be implemented at extended time

scales, this issue would most certainly need to be re-solved and suggests that it might be appropriate to use a forecast lead time dependent version of the ACAI method. Furthermore, we illustrate a complicated and sometimes nonlinear interaction of the mean and random components of the ACAI perturbations in cor-recting the forecast bias, lending weight to a need for further experimentation to understand and control the interplay in a way to maximize the impact.

To help synthesize the positive and negative impacts of ACAI on the various systems and metrics analyzed, below is a list of the more salient findings.

- Except for the southern extratropics in the EDAacai experiment, day 0–5 bias and RMSE scores are gener-ally improved by ACAI in the ensemble experiments.
- Degradations of bias in temperature and geopotential height at later lead times of the ETacai experiment are primarily driven by an acceleration of the model bias tendency (cf. Fig. 8a).
- Degradations of bias in southern extratropical tem-perature and geopotential height at later lead times of the EDAacai experiment are primarily driven by a reduction of initial time biases and an inability of ACAI to rectify the model bias tendency (cf. Fig. 8b).
- Day 5–10 RMSE scores continue to be improved for many region/variable combinations in the ensemble systems, especially in the EDA experiments.
- In the ET-based system, the ACAI perturbations provide no additional growth of ensemble spread and can even act to reduce it. This is likely due to the ET perturbations overwhelming the growth added by ACAI and feedback on the ET perturbations between cycles.
- The ACAI perturbations have a large positive impact on spread–skill relationship in the EDA-based system.
- In the deterministic system, bias and RMSE scores are improved by ACAI for nearly all region/variable combinations.

It should be noted that while methods such as ACAI present a potential means to minimize systematic model errors, these methods do not render model de-velopment unnecessary. Rather, examination of the true model biases without the use of ACAI or other online bias correction methods must be conducted periodically (especially as upgrades are made to the model) to help identify and possibly eliminate sources of model error.

As mentioned in section 2e, the Navy is currently developing a fully coupled ensemble forecasting system, and this development helped motivate the need to ex-plore the use of ACAI in an EDA system. While the implementation here only explores the correction of

biases in the atmosphere, when implemented in a cou-pled system, ACAI has the potential to arrest the de-velopment of the feedback loops between biases in the ocean and atmosphere. This is the subject of ongoing research at the Naval Research Laboratory and will be presented in future publications.

APPENDIX

Scorecard Metrics

Scorecard metrics are computed for three regions: northern extratropics (NE; 20°–90°N), tropics (TR; 20°N–20°S), and southern extratropics (SE; 20°–90°S). For each region, we focus on five variables at varying atmospheric geopotential heights. These include: 2-m air temperature (T2m), 10-m wind speed (V10m), 850 hPa air tempera-ture (T850), 500 hPa atmospheric heights (Z500), and wind speed at 250 hPa (V250).

a. Bias and root-mean-squared error (RMSE)

1-degree gridded fields are output from the NAVGEM control and experimental model runs, and bias and RMSE are computed at each 24-h lead time against an-alyses from the European Centre for Medium-Range Weather Forecasts (ECMWF) provided by the TIGGE archive. Over a particular region, the bias and RMSE are computed as

$$\text{Bias} = \frac{1}{N_p} \sum_{p=1}^{N_p} \mathcal{O}_p (\overline{x}_p - y_p),$$

$$\text{RMSE} = \sqrt{\frac{1}{N_p} \sum_{p=1}^{N_p} \mathcal{O}_p (\overline{x}_p - y_p)^2},$$

where N_p is the number of points in the region; $\overline{x}_p$ and y_p are the ensemble mean and value of the ECMWF analysis at point p, respectively; and $\mathcal{O}_p$ is a weighting applied according to the cosine of the latitude at point p. Bias and RMSE at each forecast lead time are averaged across all available forecasts.

b. Ensemble spread and variance ratio (VARR)

Over a particular region, the spread in the ensemble is computed as

$$\sigma_e = \frac{1}{N_p} \sum_{p=1}^{N_p} \sqrt{\frac{1}{N-1} \sum_{m=1}^{N} \emptyset_p (x_p^m - \overline{x}_p)^2},$$

where N is the number of ensemble members; N_p is the number of points in the region; x_p^m and $\overline{x}_p$ are the value of ensemble member m and the ensemble mean at point p, respectively; and $\emptyset_p$ is a weighting applied according to the cosine of the latitude at point p. The variance ratio (VARR) is then the ratio of the squared ensemble spread (σ_e^2) to the bias corrected mean squared error. Spread and VARR at each forecast lead time are averaged across all available forecasts.

REFERENCES

Batté, L., and M. Déqué, 2016: Randomly correcting model errors in the ARPEGE-Climate v6.1 component of CNRM-CM: Applications for seasonal forecasts. *Geosci. Model Dev.*, **9**, 2055–2076, https://doi.org/10.5194/gmd-9-2055-2016.

Berner, J., and Coauthors, 2017: Stochastic parameterization: Toward a new view of weather and climate models. *Bull. Amer. Meteor. Soc.*, **98**, 565–588, https://doi.org/10.1175/BAMS-D-15-00268.1.

Bhargava, K., E. Kalnay, J. A. Carton, and F. Yang, 2018: Estimation of systematic errors in the GFS using analysis increments. *J. Geophys. Res. Atmos.*, **123**, 1626–1637, https://doi.org/10.1002/2017jd027423.

Bishop, C. H., and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, **56**, 1748–1765, https://doi.org/10.1175/1520-0469(1999)056<1748:ETAAO>2.0.CO;2.

Bowler, N. E., and Coauthors, 2017: Inflation and localization tests in the development of an ensemble of 4D-ensemble variational assimilations. *Quart. J. Roy. Meteor. Soc.*, **143**, 1280–1302, https://doi.org/10.1002/qj.3004.

Dee, D. P., and A. M. Da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.*, **124**, 269–295, https://doi.org/10.1002/qj.49712454512.

——, and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Hogan, T., and Coauthors, 2014: The Navy global environmental model. *Oceanography*, **27**, 116–125, https://doi.org/10.5670/oceanog.2014.73.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242, https://doi.org/10.1175/1520-0493(1996)124<1225:ASSATE>2.0.CO;2.

Kucukkaraca, E., and M. Fisher, 2006: Use of analysis ensembles in estimating flow-dependent background error variances. ECMWF Tech. Memo. 492, ECMWF, 18 pp., https://doi.org/10.21957/36n2z0p1p.

Kuhl, D. D., T. E. Rosmond, C. H. Bishop, J. McLay, and N. L. Baker, 2013: Comparison of hybrid ensemble/4DVar and 4DVar within the NAVDAS-AR data assimilation framework. *Mon. Wea. Rev.*, **141**, 2740–2758, https://doi.org/10.1175/MWR-D-12-00182.1.

McLay, J. G., C. H. Bishop, and C. A. Reynolds, 2008: Evaluation of the ensemble transform analysis perturbation scheme at NRL. *Mon. Wea. Rev.*, **136**, 1093–1108, https://doi.org/10.1175/2007MWR2010.1.

——, ——, and ——, 2010: A local formulation of the ensemble transform (ET) analysis perturbation scheme. *Wea. Forecasting*, **25**, 985–993, https://doi.org/10.1175/2010WAF2222359.1.

Piccolo, C., M. J. P. Cullen, W. J. Tennant, and A. T. Semple, 2019: Comparison of different representations of model error in ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, **145**, 15–27, https://doi.org/10.1002/qj.3348.

Rosmond, T., and L. Xu, 2006: Development of NAVDAS-AR: Non-linear formulation and outer loop tests. *Tellus*, **58A**, 45–58, https://doi.org/10.1111/j.1600-0870.2006.00148.x.

Saha, S., 1992: Response of the NMC MRF model to systematic-error correction within integration. *Mon. Wea. Rev.*, **120**, 345–360, https://doi.org/10.1175/1520-0493(1992)120<0345:ROTNMM>2.0.CO;2.

Simpson, I. R., J. T. Bacmeister, I. Sandu, and M. J. Rodwell, 2018: Why do modeled and observed surface wind stress climatologies differ in the trade wind regions? *J. Climate*, **31**, 491–513, https://doi.org/10.1175/JCLI-D-17-0255.1.

Whitaker, J. S., and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, https://doi.org/10.1175/MWR-D-11-00276.1.

Xu, L., T. Rosmond, and R. Daley, 2005: Development of NAVDAS-AR: Formulation and initial tests of the linear problem. *Tellus*, **57A**, 546–559, https://doi.org/10.1111/j.1600-0870.2005.00123.x.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.