SPECIAL
Community Climate System Model CCSM4
COLLECTION

DART/CAM: An Ensemble Data Assimilation System for CESM Atmospheric Models

KEVIN RAEDER, JEFFREY L. ANDERSON, NANCY COLLINS, AND TIMOTHY J. HOAR

IMAGe, CISL, National Center for Atmospheric Research, Boulder, Colorado*

JENNIFER E. KAY AND PETER H. LAURITZEN

CGD, NESL, National Center for Atmospheric Research, Boulder, Colorado*

ROBERT PINCUS

University of Colorado, and NOAA/Earth System Research Laboratory, Boulder, Colorado

(Manuscript received 14 July 2011, in final form 2 April 2012)

ABSTRACT

The Community Atmosphere Model (CAM) has been interfaced to the Data Assimilation Research Testbed (DART), a community facility for ensemble data assimilation. This provides a large set of data assimilation tools for climate model research and development. Aspects of the interface to the Community Earth System Model (CESM) software are discussed and a variety of applications are illustrated, ranging from model development to the production of long series of analyses. CAM output is compared directly to real observations from platforms ranging from radiosondes to global positioning system satellites. Such comparisons use the temporally and spatially heterogeneous analysis error estimates available from the ensemble to provide very specific forecast quality evaluations. The ability to start forecasts from analyses, which were generated by CAM on its native grid and have no foreign model bias, contributed to the detection of a code error involving Arctic sea ice and cloud cover. The potential of parameter estimation is discussed. A CAM ensemble reanalysis has been generated for more than 15 yr. Atmospheric forcings from the reanalysis were required as input to generate an ocean ensemble reanalysis that provided initial conditions for decadal prediction experiments. The software enables rapid experimentation with differing sets of observations and state variables, and the comparison of different models against identical real observations, as illustrated by a comparison of forecasts initialized by interpolated ECMWF analyses and by DART/CAM analyses.

1. Introduction

Data assimilation (DA) has long been recognized as an indispensable tool in numerical weather forecasting for generating realistic initial and boundary conditions, for melding diverse observations into gridded analyses that have been used for model forecast verification (Lynch 2006) and for added quality control of observational systems. Until recently, its usefulness for climate model development has not been compelling enough to warrant the effort of implementing the best available DA algorithms. That effort has been greatly reduced by the advent of ensemble DA, so that climate model development and research can now benefit greatly and directly from the variety of tools available from DA. Several generations of the Community Atmosphere Model [CAM, the atmospheric component of the Community Earth System Model (CESM)] can now be used with ensemble DA using the Data Assimilation Research Testbed (DART).

The DART algorithm and software are described briefly here (section 2), but the focus of this paper is a description of the DART/CAM interface (section 3) and the application to CAM of the tools available in DART (sections 4 and 5). A more detailed description

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

Corresponding author address: Kevin Raeder, IMAGe, CISL, NCAR, 1850 Table Mesa Dr., Boulder, CO 80305.
E-mail: raeder@ucar.edu

DOI: 10.1175/JCLI-D-11-00395.1

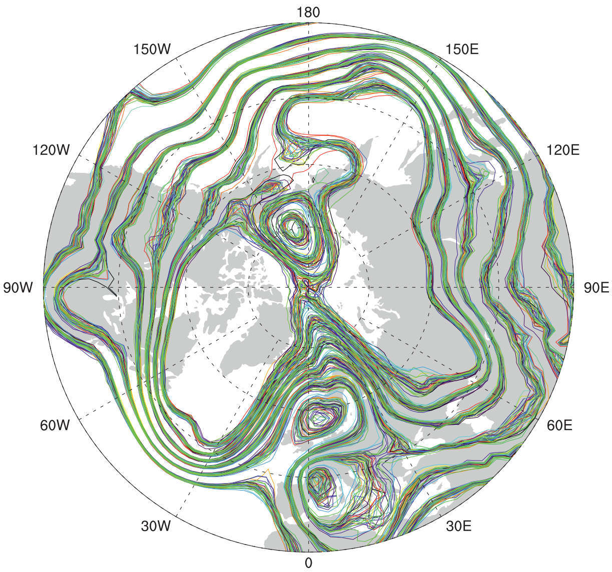

FIG. 1. Geopotential heights (500 hPa) for half (40) of the ensemble members typically used in DART/CAM assimilations for 1200 UTC 17 Feb 2003. Areas of high model/forecast uncertainty are apparent where the contours are not tightly clustered. This happens in both data-rich and data-sparse areas.

of DA and DART is available in the references below, and in an overview paper (Anderson et al. 2009).

2. The Data Assimilation Research Test bed

a. Overview of ensemble data assimilation

Data assimilation is the combination of information from a model forecast and observations of a physical system to produce an improved model estimate called an analysis. The analysis can be used as initial conditions for subsequent model forecasts. Ensemble DA uses a set (ensemble) of model forecasts to compute the sample covariance between model state variables and forecast estimates of observations. These covariances are then used in Bayes' theorem to compute an ensemble of model state analyses (Anderson 2003). Each analysis can be regarded as an approximately equally likely draw from an analysis probability distribution. An illustration of such an ensemble is presented in Fig. 1. The ensemble mean analysis is often used as a best estimate of the model state. Ensemble Kalman filter algorithms for geophysics were described in Evensen (1994) and Burgers et al. (1998).

Most operational centers for numerical weather prediction use variational methods of DA. However, the Meteorological Service of Canada has been running an ensemble DA system since 2005 (Houtekamer et al. 2009). Recent results have suggested that ensemble systems are now competitive with variational systems for geophysical DA (Buehner et al. 2010; Miyoshi et al. 2010). In addition to the analysis, ensemble assimilation systems provide an estimate of the analysis error distribution.

b. DART

DART has been developed since 2002 at the National Center for Atmospheric Research (NCAR) as a community facility that makes high-quality, theoretically based data assimilation available to the geoscience research community while requiring little expertise in DA on the part of users (www.image.ucar.edu/DAReS/DART).

The modularity of DART makes the interface to new models and observations straightforward and clean. As an example, CAM, version 3.5 (CAM3.5) + Chem was interfaced to DART in less than a month by a scientist unfamiliar with both the model and DART (section 3a). CAM3.5 refers to a version developed before the Community Climate System Model, version 4 (CCSM4) release (Gent et al. 2011) to focus on the implementation of the finite-volume core developed by Lin (2004). Hereafter ''-FV'' will refer to CAM executables using the finite-volume core. In addition, ''CAM4'' and ''CAM5'' will refer to the FV version, unless noted otherwise. ''Chem'' refers to a simple chemistry model interfaced to CAM to enable forecasts of some chemical species (Arellano et al. 2007). As a result of this straightforward interface, DA with the most current versions of CAM and other full-complexity models has been possible for several years.

The basic ensemble assimilation algorithm is straightforward to implement for low-order models that assimilate simple, synthetic observations but can nonetheless yield useful results for theoretical studies. For full-complexity models, it is necessary to use additional algorithms to maximize assimilation performance while minimizing the computational expense. DART provides a flexible suite of these algorithms, some of which are briefly described below. In the past such algorithms have required tuning of assimilation parameters, but our experience has shown that the DART versions are either self-tuning or the assimilation is sufficiently insensitive to the parameter, so that almost no tuning is required from users, especially for the CAM lineage of models.

1) ENSEMBLE INFLATION

Assimilation of each observation causes the spread of the ensemble to decrease, as more information is added to the state of knowledge of the system. Any error sources that are not included in the observation error or model spread cause the spread to decrease more than is justified. This can be corrected for the large atmospheric models used in DART by inflating the spread periodically.

DART can use user-provided inflation fields or employ a hierarchical Baysian filter (HBF) to use the observations to determine appropriate inflation values as a function of time, space, and model variable (Anderson 2009). Each observation can increase the inflation via the hierarchical Baysian filter, but during periods of fewer observations the HBF cannot respond to the relative lack of observations. To prevent inflation from remaining too large when observation counts decrease, the inflation is damped to some percentage of its previous value during each assimilation. So, the adaptive inflation algorithm has two basic tuning parameters: one that controls how rapidly inflation values adapt to inconsistencies between model forecasts and observations and one that controls how rapidly inflation is damped. Tuning experiments in CAM and several other global and regional atmospheric models revealed that good performance was obtained in all cases with a standard pair of values. This experience suggests that it is not necessary to explore the inflation parameters when implementing DART with a new version of CAM. An example of such an inflation field can be found in Anderson et al. (2009, their Fig. 13).

2) Localization

The probability distribution of a variable is represented in ensemble DA by an ensemble of draws from that distribution. Large models, such as CAM, motivate the use of ensembles that are too small to accurately represent the Gaussian distribution assumed by the DA algorithm. This results in nonnegligible correlations between widely separated variables, which are believed, a priori, to be uncorrelated (Anderson 2007). Localization is a mechanism for telling the assimilation algorithm that it should ignore such spurious correlations. DART can employ a user-specified localization distance [the half-width of a Gaspari–Cohn distribution (Gaspari and Cohn 1999)] or modify this value based on the local density of observations (Torn 2010). Tuning exercises for the localization have been performed in several CAM versions and in other global atmospheric models, and the user-specified values of 0.2 radians in the horizontal and 1000 hPa in the vertical have resulted in good performance in all cases. This suggests that tuning of the horizontal localization is not necessary when implementing DART with new versions of CAM. However, tuning of the vertical localization may be required to use new types of remote sensing observations with CAM.

3) Sampling error correction

The leading source of error in many ensemble DA applications is sampling error due to the use of small ensembles. DART provides an algorithm that can reduce sampling error by applying a correction factor for state variable increments that is a function of the ensemble size and the sample correlation of a state variable with an observation (Anderson 2012). This correction improves the performance of relatively small ensembles, for example, 20 members, but it has negligible effect on the performance of larger ensembles, for example, 80 members.

4) DART/CAM leverages code resources

DART has a large user community of modelers and observations experts. New interfaces and features contributed from this community are added to the DART software repository for general use. These contributions and the modularity of DART enable two kinds of comparison studies, with one study consisting of using CAM to assimilate differing sets of observations.

As an example, DART/CAM was one of the first global assimilation systems to assimilate observations from the Constellation Observing System for Meteorology Ionosphere and Climate (COSMIC) constellation of global positioning system (GPS) radio occultation instruments. Figure 2 shows the fit to observations for three experiments, with one having assimilated the observations used in the (National Centers for Environmental Prediction) NCEP–NCAR reanalyses. Those observations include temperature and winds from radiosondes, aircraft, Aircraft Communications Addressing and Reporting System (ACARS), and satellite drift winds. The others assimilated those plus COSMIC GPS observations (Anthes et al. 2008). Note that Fig. 2 shows the fit to observations *at the locations of the radiosonde observations,* but that the GPS observations at *other* locations are able to improve that fit. The fit of 6-h forecasts to observations is shown because the forecasts have less potential for being ''overfit'' to the observations, which would give an inflated and erroneous estimate of the agreement between the observations and the model. Another example is the novel hyperspectral radiance retrieval (Liu and Li 2010), which is being studied in NCAR's Weather Research and Forecasting Model (WRF) but can immediately be used in CAM assimilations.

Another comparison consists of using the identical assimilation algorithm and observation set but assimilating into CAM and another model. For instance, comparisons to assimilations in the Geophysical Fluid Dynamics Laboratory (GFDL)'s Atmospheric Model, version 2 (AM2) have provided insight into CAM biases (Pincus et al. 2011).

3. DART/CAM

a. CAM versions supported

The DART/CAM interface was originally written for CAM2.0.1, which used the Eulerian core (Kiehl et al. 1996)

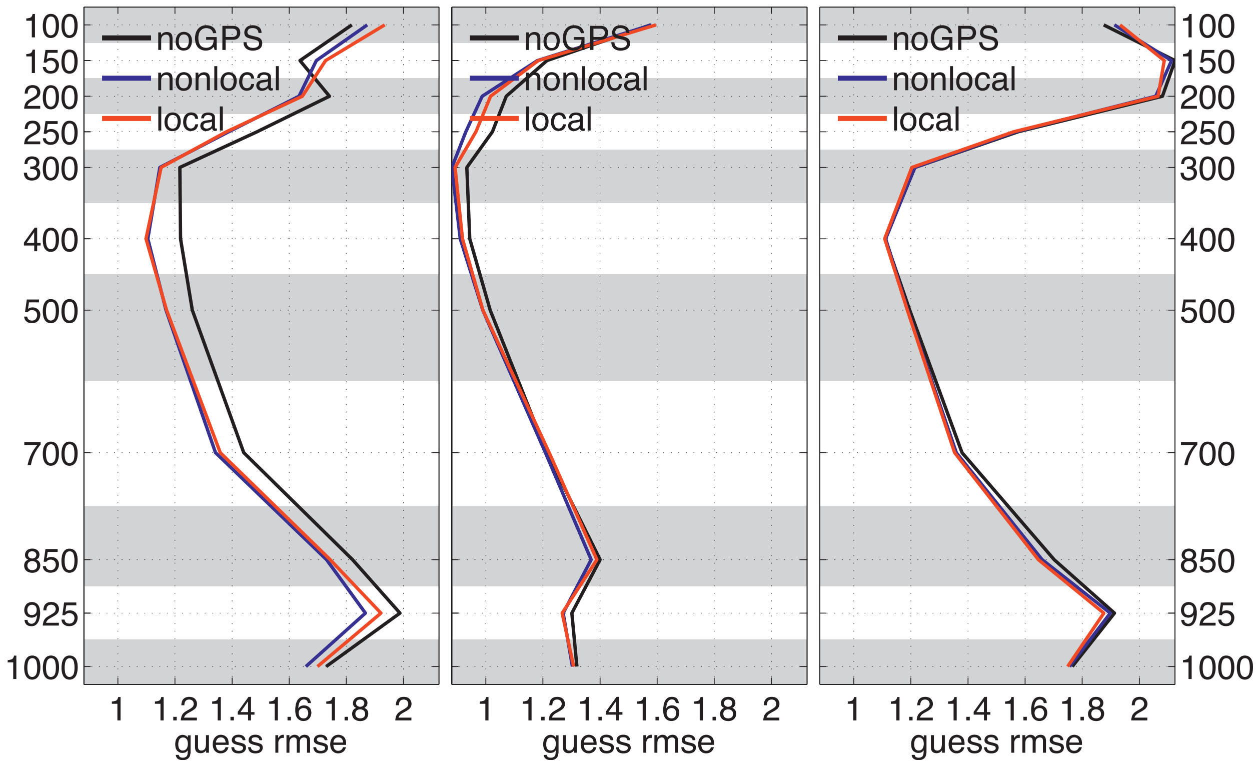

FIG. 2. RMS error of the 6-h forecasts (''priors'' or ''guess''), relative to the radiosonde T observations, from DART/CAM assimilation of December 2006 at the locations of the observations in three regions: (left) south of 20°S, (middle) 20°S–20°N, and (right) north of 20°N. Black curve represents assimilation of only NCEP–NCAR reanalysis observations, while the red and blue curves are for NCEP–NCAR plus COSMIC GPS radio occultation observations by two different forward operators.

converted to the Arakawa A grid for output. Arellano et al. (2007) used that as the basis for an interface to CAM3.5-FV on the Arakawa D grid to enable chemical data assimilation. The FV core promised enhanced transport characteristics for improved chemistry. The interface has since been used extensively with CAM3.1, CAM3.5, and several development versions leading up to CAM4 (Gent et al. 2011) and CAM5 (Neale et al. 2010). To date, resolutions used include T42 and T85 for the Eulerian core [''T'' means triangular truncation (Kiehl et al. 1996)], and 4° × 5° (for testing), 1.9° × 2.5°, and 0.9° × 1.25° for the FV core. The interface has been extensively tested with the Eulerian (Žagar et al. 2010) and FV cores (Arellano et al. 2007; Pincus et al. 2011; Lauritzen et al. 2012).

b. Adaptation to CAM features

1) SUPPORTS ARBITRARY NUMBER OF ADDITIONAL TRACERS

The definition of the ''state'' of the CAM atmosphere is somewhat flexible; besides the usual dynamical variables and temperature, it can include an arbitrary number of tracers, such as moisture and chemical species. DART/CAM accommodates this flexibility by allowing specification of the model state without recompiling DART. An example of how this could be used is that assimilations could be run with a full set of chemical species supporting a complex chemical model, then with a reduced set of species supporting a simpler chemical model.

While the interface has been coded to be as insensitive as possible to CAM's dynamical core, resolution, and model state variables, some aspects of current CAMs have required special treatment.

2) MODEL TOP DAMPED REGION

CAM is strongly damped near the model top, which challenges the assimilation algorithm. The variables there are not free to adjust to the observations, and the ensemble spread can become too small for the assimilation algorithm to work. Two mechanisms exist to sidestep this problem. The primary tool allows the user to specify a height, above which the influence of observations on state variables will be reduced as a function of the distance above the specified height. This height is usually set to 150 hPa, and the influence of observations falls to

0 by about 54 hPa. In the 26-level resolution (CAM4), these correspond to levels 11 and 6, as counted from the top. The secondary mechanism enables a user to specify a height, above which no observations will be used. This is usually quite flexible, but it must be set to 100 hPa or less for GPS radio occultation observations. These two mechanisms interact with the localization to determine how much influence an observation will have on a given state variable.

3) Bounded model variables

Another challenge to the assimilation algorithm comes from the bounded nature of some model variables, for example, tracers, which must be nonnegative, or cloud fraction, which must remain between 0 and 1. The spread inflation algorithm can push the values of these variables outside of their ranges for some ensemble members. In many cases the damage from this is minimal because CAM will reset them to the proper range when it starts, but if such variables are the primary interest of the assimilation (Pincus et al. 2011), DART includes a rank histogram assimilation algorithm, which can accommodate such variables (Anderson 2010).

4) The poles

The current CAM grids (Arakawa A for the Eulerian core, Arakawa D for the FV core) have no grid points poleward of the first and last latitudes, but there can be observations there. Rather than perform a relatively expensive interpolation involving all the most poleward grid points, the user can specify a maximum latitude for observations to be assimilated.

5) The surface

The topography of CAM only approximates the real topography, near which some observations are taken. For observations that are below the lowest model level, no interpolation from the model grid to the observation location is possible. Currently, rather than take the risk of extrapolating the CAM model state to the observation location, the assimilation algorithm ignores those observations.

c. *Continuing development*

The future of geophysical computing appears to be on massively parallel hardware that is leading to the adoption of dynamical core(s) on horizontally nonrectangular grids, which offer better scaling on thousands of processors. Examples include the spectral element core on the cubed sphere (Taylor et al. 2008; Evans et al. 2012) and other, even less structured grids. This is not a fundamental problem for DART/CAM because DART and CAM are separate programs that communicate with each other via a conceptually simple interface, and DART uses the parallel assimilation algorithm (Anderson and Collins 2007) to facilitate excellent scaling up to thousands of processors for CAMs with high resolution or a large number of tracers.

In the opposite direction from increasingly complex grids, the DART/CAM interface is compatible with the single-column CAM (S. Park and J. Truesdale 2011, unpublished manuscript), but no additional development or testing has been done to investigate the ramifications for DA of having a single column in CAM. This mode has been used in WRF (Hacker and Rostkier-Edelstein 2007).

The new ''multi-instance'' capability in CESM enables it to advance an ensemble of states of a model component for the same forecast period. DART/CAM adapts this feature to data assimilation by making CESM stop and call DART to assimilate observations whenever they are available. This will facilitate coupled assimilations using the atmosphere, ocean, and land as active components. Each CESM component will potentially be influenced directly by all available observations, even observations of a different part of the earth system.

4. Using DART/CAM for climate model evaluation and development

a. *Forecast verification, analysis error, and observation space diagnostics*

Climate model fidelity is typically measured against gridded analyses that are a merging of observations with model forecasts using data assimilation. In most cases, no estimate of the errors associated with the analyses is provided, so it is impossible to objectively judge whether a model result is consistent with the analyses. An ensemble analysis like those produced by DART can provide an estimate of analysis and forecast uncertainty.

Climate model performance can also be evaluated in ''observation space'' by using the gridded model variables to compute estimates of available observations. Comparing these estimates to actual observations is referred to as evaluating performance in observation space. For observations such as a radiosonde temperature, this only requires spatial interpolation from the model grid, while observations such as a COSMIC radio occultation may require much more complicated functions of the model variables.

Evaluating the significance of differences between actual observations and model estimates requires error estimates for both the observations and the model. The former come from instrument designers, while the ensemble spread available from a DART assimilation provides the latter. Spaghetti plots like Fig. 1 are one way to

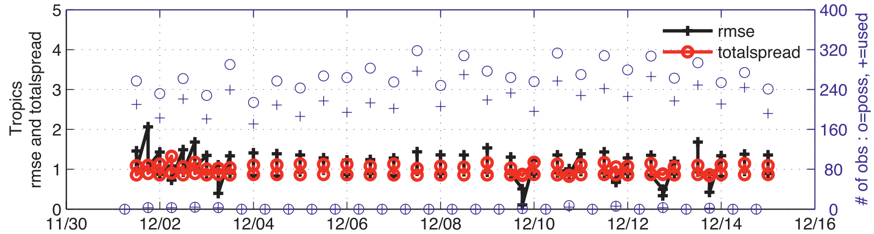

FIG. 3. RMS error of the ensemble of estimates of a temperature observation is calculated for each observation in a time slot in the region spanning 20°S–20°N and 800–887 hPa. These are averaged to yield the value represented by a black plus sign at that time. There are two black pluses at each time: one each for the prior (6-h forecast) and posterior (analysis) estimates. Total spread (section 4a) is shown by the red circles. (left axis) RMSE and total spread (K), and (right axis) blue symbols show the number of observations available (°) and used (+) at each time. Time series span early December 2010 and come from the 2° CAM assimilation described in section 5a.

visualize the spatial heterogeneity of the ensemble. The spread also evolves with time as a function of the atmospheric state and the available observations.

Given estimates of uncertainty for both the ensemble model estimates of an observation and the observation itself, the significance of differences between them can be estimated. The ''total spread,'' the square root of the sum of the ensemble variance and the error variance associated with the observation, is the expected value of the difference between the model ensemble mean and the observed value.

DART has tools to use the extensive information available in the model ensemble to evaluate model performance in observation space. Examples can be seen in Anderson et al. (2009, their Figs. 3 and 12). Figure 3 here provides information about a CAM assimilation in the tropics. This assimilation is stable, as shown by the limited variability of the RMS error as a function of time. The DART observation quality control algorithm discards observations that are unexpectedly different from the model ensemble estimate. Figure 3 shows that the number of rejected observations is not large, generally less than 10%. This fraction does not vary significantly in time, providing further evidence that the assimilation is working appropriately. In a carefully tuned ensemble assimilation system using a model and observations with small systematic errors (bias), the ensemble total spread should provide a good estimate of the RMS error between the ensemble mean and the observations. Figure 3 shows generally good agreement for this CAM assimilation. Instances of large differences between the RMS error and the total spread are confined to 0600 and 1800 UTC. At these times, the number of radiosondes available in the tropics is very small, so that sampling error can lead to larger differences. This temporal variability highlights the difficulty of assigning a single ''quality'' score to a small set of model analyses or forecasts. Similar plots for other regions, times, and observation types show an even richer variety of behaviors. Similar plots of model time-mean error (bias) relative to the observations are particularly useful for identifying model (or observational) error.

In contexts where comparison to analyses is still preferred, the frequent analyses produced by DART/CAM enables evaluation of CAM forecast error without the confounding error of a foreign model being used in the analyses (see also section 5a; Kay et al. 2011, hereafter Kay2011).

b. *CAPT*

For much of the last decade, the Cloud-Associated Parameterization Testbed (CAPT) project (Phillips et al. 2004; www-pcmdi.llnl.gov/projects/capt) has used short-term forecasts to gain insights into parameterization errors in climate models (Klein et al. 2006; Xie et al. 2008). To date CAPT has initialized these forecasts from analyses provided by operational weather centers for reasons that are partially pragmatic (ensemble DA methods were just emerging when the project started, while variational methods require adjoint and tangent–linear models not normally available for climate models) and partially reflect the belief that DA was best left to experts. Nonexpert users of DA are unlikely to exploit all the observation types used in operational analyses, particularly not the satellite radiance observations that must be carefully processed to remove biases. Using fewer observations in conjunction with an assimilating model whose short-term biases have not been explored (as is typical for climate models) might lead to substantially less realistic initial conditions.

But there are several reasons why identifying parameterization deficiencies can be enhanced by using the same model in the forecast and data assimilation processes. The first reason is that the numerical weather

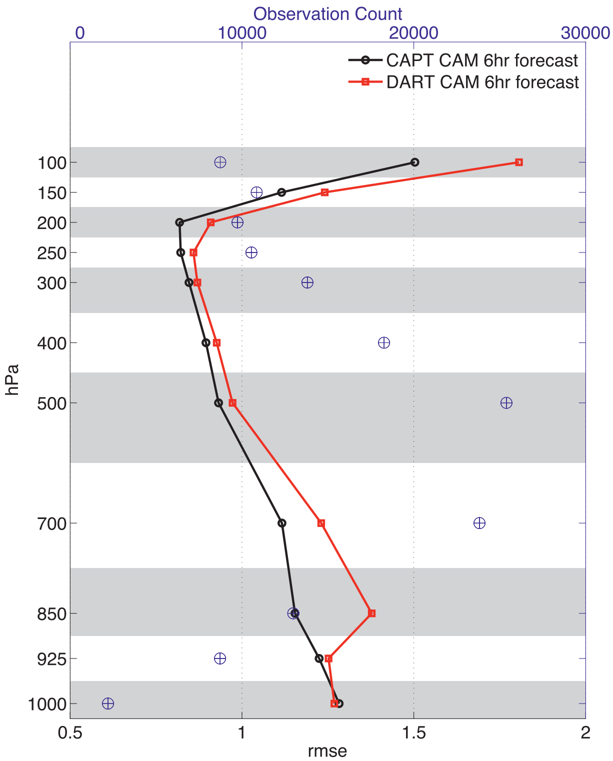

FIG. 4. RMS error in CAM4 6-h forecasts of tropical temperature evaluated against radiosonde observations. Black line shows results from CAM4 running at (nominal) 1° resolution from analyses provided by the European Centre for Medium-Range Weather Forecasts; red line shows results from a lengthy reanalysis from DART using the same forecast model at 2° resolution. Blue crossed circles show the number of observations used in the comparison (i.e., the number remaining after quality control), as numbered on the top axis.

prediction model that generates the initial conditions for the climate model forecast is certain to have its own set of systematic errors. These errors will result in errors in the climate model forecast that are related to the external model. It is extremely difficult to isolate the errors due to the climate model in this situation. The second reason is that even if the analyzing model is unbiased, interpolation can lead to initial conditions that are not in large-scale balance, so that very short-term (1 day or less) forecasts are dominated by spinup not associated with forecast model error. Finally, using foreign initial conditions in the atmosphere usually requires special efforts (normally ‘‘nudged’’ runs) to initialize the land model to be consistent with the state of the atmosphere.

For these reasons the CAPT project has begun using DART to provide ensemble DA and ensemble forecasts. Analyses produced using CAPT and DART will be useful in diagnosing model error only to the extent that the initial conditions are themselves accurate. Figure 4 shows 6-h forecast errors in tropical temperature for two sets of forecasts with CAM4-FV: one with the model running at nominal 1° resolution from analyses provided by the European Centre for Medium-Range Weather Forecasts (interpolated from the original T799 resolution; roughly ¼° resolution) and another from an ensemble reanalysis (described in section 5a) using DART and a 2° version of the model. The evaluation is performed in observation space (see section 4a) against radiosonde temperature observations that fell within three standard deviations of the DART/CAM ensemble mean. Errors in forecasts from the DART ensemble reanalysis are larger than from the ECWMF conditions (and in extratropical regions, the differences are larger still) due to some combination of reduced model resolution and smaller numbers of observations used in the analyses. This comparison illustrates the ability of DART to readily provide a direct comparison of different models (or versions) against the same set of observations. We plan to increase the DART/CAM resolution to 1° and extend this comparison to cover one or more years. Such changes in the model(s) used in DART are simple to implement, and facilitate model development.

c. *Evaluating boundary layer response to sea ice loss*

Kay2011 used short-term CAM4 forecasts starting from DART/CAM analyses to evaluate the boundary layer response in CAM to observed sea ice loss during the 2007 melt year, the most extreme melt year on record (Stroeve et al. 2008). The Kay2011 work was motivated in part by the contrast between observations, which revealed little cloud response to midsummer sea ice loss (Kay and Gettelman 2009), and the response in CAM4, which showed large low cloud increases over newly open water. Indeed, *reduced* clouds associated with large-scale circulation are a factor that enhanced sea ice loss during the 2007 melt season (Kay et al. 2008). The Kay2011 study found that although the September CAM4 boundary layer response to sea ice loss was qualitatively consistent with observations, the July CAM4 forecasts developed unrealistic cloud increases over regions that became ice free. This was traced to a parameterization error: the failure to explicitly require an unstable boundary layer for the formation of low clouds. Kay2011 implemented a physically motivated improvement to the CAM4 low cloud parameterization and were able to verify that it improved the cloud response to sea ice loss (Fig. 5). This discovery and solution was greatly facilitated by the ability to start CAM from DART/CAM analyses.

More recently, work has been undertaken to examine the boundary layer response to sea ice loss in winter. Open polar water in winter provides a source of both heat, which destabilizes the boundary layer, and moisture,

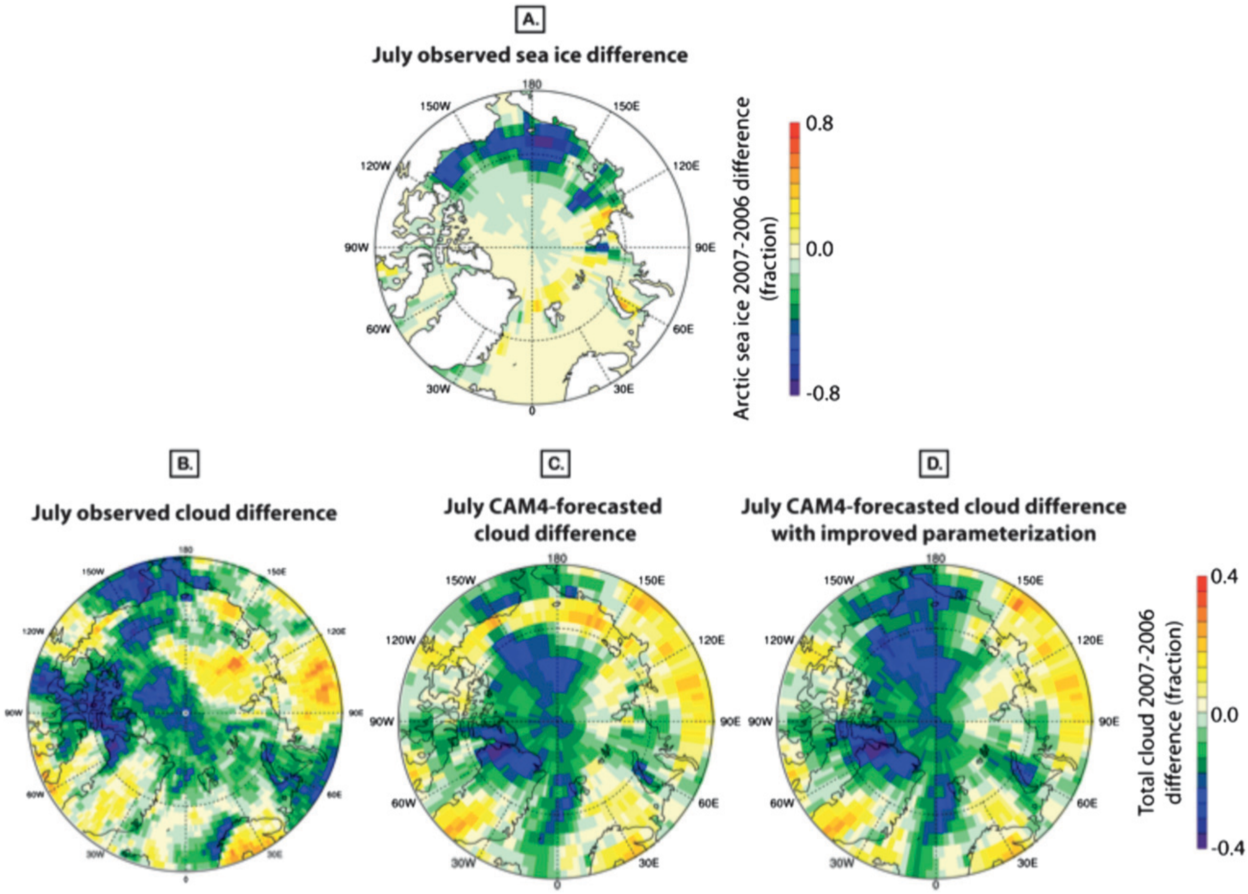

FIG. 5. July 2007 minus July 2006 sea ice and cloud patterns: (a) observed sea ice difference, (b) observed cloud response to sea ice loss, (c) CAM4-forecasted cloud difference with no modifications showing the unrealistic and ubiquitous cloud increases in regions that had less sea ice in 2007 than in 2006, (d) CAM4-forecasted cloud difference with improved cloud parameterization, a much closer match to the observations in regions of sea ice loss. Adapted from Figs. 2 and 8 of Kay2011.

the combination of which should lead to a warmer boundary layer and increased cloudiness. The influence of variations in sea ice extent and atmospheric circulation on such winter temperature and cloud anomalies can be evaluated with short-term DART forecasts. For example, Fig. 6 shows the influence of December 2007 sea ice anomalies on the atmosphere by comparing pairs of 24-h forecasts, which start from the same analyses but use either climatological or December 2007 ocean boundary conditions, as in Kay2011. Figures 6d–f show that over regions of sea ice loss, the temperature and cloud amounts increase up to 10 K and 25% respectively, consistent with the expected boundary layer response to winter sea ice loss. Figure 6c shows that the sea level pressure (SLP) changes caused by the ice anomaly are <1 hPa everywhere except over the regions of largest ice anomaly, and even there they are <2 hPa. These SLP anomalies appear to be thermodynamically driven and a local response to the ice anomaly. This is consistent with previous studies showing that a baroclinic atmospheric circulation response to sea ice loss takes days to appear (Deser et al. 2007).

d. Noise detection with DART

Climate model dynamics are usually evaluated in terms of statistical measures, such as mean sea level pressure and zonal wind speed averaged in time over several years and in space along latitudes/longitudes. While such diagnostics are indeed relevant for assessing the fidelity of climate models, this averaging may make it harder to detect potential numerical noise problems (unless, of course, the noise is excessive or stationary in space/time).

DART was used as an efficient tool to detect noise problems in CAM; two examples are discussed below. Assimilation runs over only a few days using DART can reveal noise. A key point is that the assimilation constrains the model to the same synoptic situations when the model is modified in an attempt to correct problems. Both noise problems below were visible at the same date and time for repeated tests of model versions. Both noise

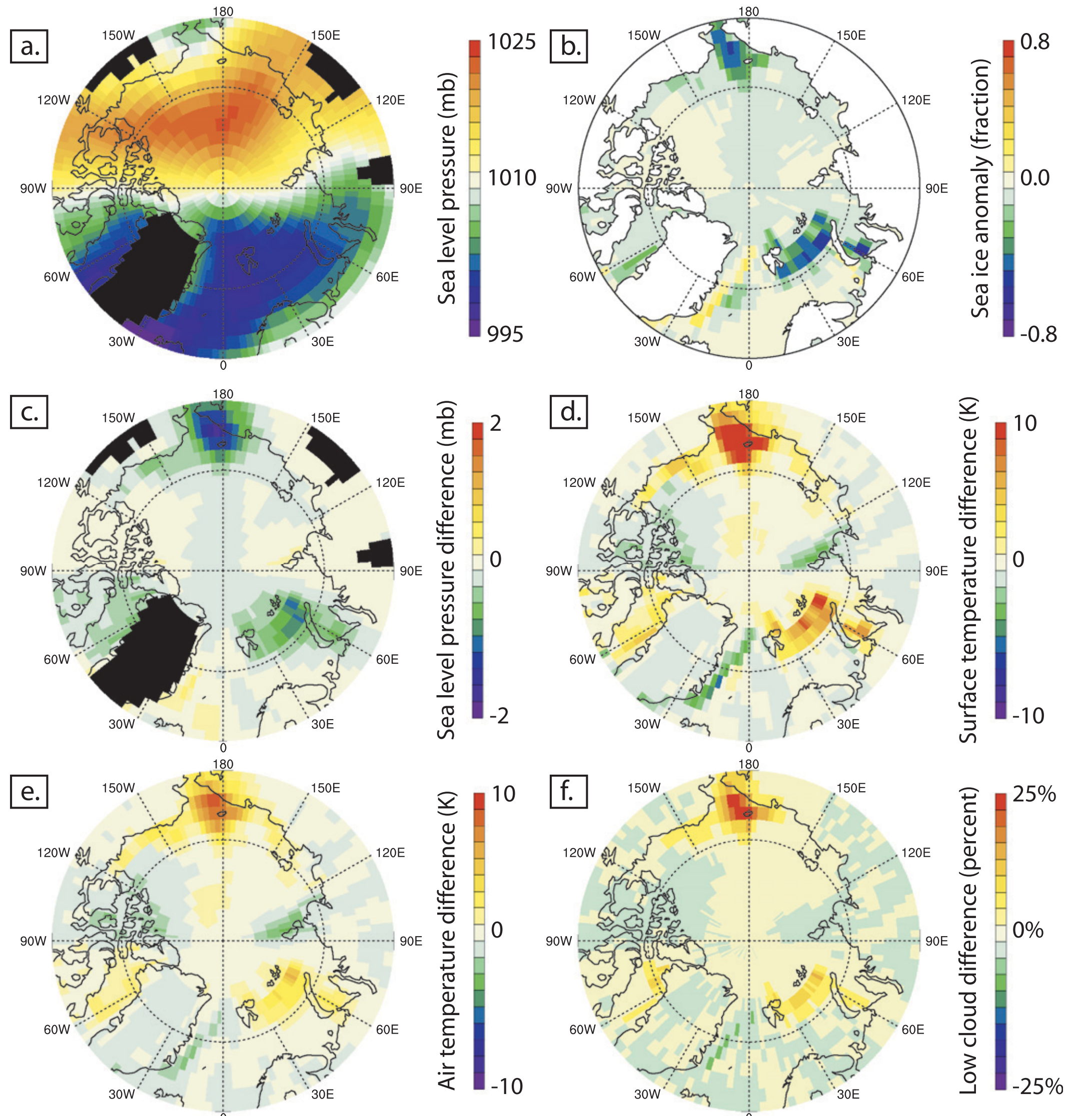

FIG. 6. Influence of the December 2007 sea ice anomaly (actual month − climatology) on the Arctic atmosphere in 24-h CAM forecasts started from DART/CAM analyses: (a) December 2007 SLP, (b) December 2007 sea ice anomaly, and the resulting anomalies of (c) SLP (mb), (d) surface temperature (K), (e) air temperature in the lowest model layer (K), and (f) low cloud (percent).

issues can also be seen in long climate runs but occur only sporadically in time and space. Because of chaotic error growth in free climate runs, the time and place where noise may appear is not reproducible between different model versions that contain the same error. Instead, careful examination of many days' worth of model output is required to determine if model noise has been reduced.

1) Noise associated with "polar filter" transition

The CAM-FV model uses explicit time-stepping based on a flux-form finite-volume scheme [Lin 2004; for a review see Machenhauer et al. (2009)]. Because of the convergence of the meridians near the poles of the regular latitude–longitude grid, it is necessary to apply filtering to stabilize the fast-moving gravity waves to avoid the need for using excessively short time steps. In the midlatitudes CAM3 uses a one-dimensional digital filter along latitudes (approximately between 36°N/S and 69°N/S for the 1.9° × 2.5° horizontal resolution version of CAM-FV) and a fast Fourier transform (FFT) filter in the polar regions poleward of 69°N/S. This 1D filtering along latitudes is referred to as the polar filter, although filtering is also applied away from the polar regions in the midlatitudes. A simple linear analysis of the damping properties of the two filtering methods will show that not all wavelengths are damped equally in the transition region, where the digital filtering is replaced with FFT filtering. Experiments with DART have shown that this noncontinuous transition may lead to grid-scale noise (see Fig. 8a in Anderson et al. 2009). Therefore, in CAM4 (and later versions of CAM) the digital filter was replaced with the FFT filter so that FFT filtering would be applied from approximately 36°N/S all the way to the poles. The noise detected in DART/CAM3 associated with the digital-to-FFT transition zone is no longer present in DART/CAM4 and later versions (see Fig. 8b in Anderson et al. 2009).

2) NOISE ALIGNED WITH LATITUDES/LONGITUDES

Noise was also detected outside of the filtering transition zone in CAM3-FV (described above) using the DART/CAM setup, and later also identified in ''free running'' CAM (no data assimilation) as well as in idealized steady-state experiments (see Fig. 5 in Lauritzen et al. 2010). The noise is particularly visible in the meridional wind field in the upper part of the model vertical domain (approximately 200 hPa and above) and is aligned with grid lines (latitudes/longitudes). Examples are given in Figs. 4a and 6a in Lauritzen et al. (2012) for free-running CAM and DART/CAM, respectively. Closely related to the grid-scale noise is the divergence damping used in CAM-FV. In the original version of CAM-FV, second-order divergence damping is used. A logical first attempt at reducing noise levels would be to increase the second-order divergence damping strength; it results in a slight reduction in the amplitude of the spurious grid-scale waves aligned with grid lines (Fig. 4b in Lauritzen et al. 2012) at the expense of increased damping of divergence associated with longer, well-resolved wavelengths (Fig. 3 in Lauritzen et al. 2012). A more scale-selective damping operator was therefore implemented, that is, a fourth-order divergence damping scheme. This higher-order damping operator provides, in general, a more scale-selective damping of divergent modes near the grid scale, although its implementation on the regular latitude–longitude grid may provide less damping in the polar regions (Whitehead et al. 2011). The fourth-order divergence damping more effectively reduced grid-scale noise in the divergence field while not damping the divergence of ''well resolved'' scales (Figs. 4c and Fig. 3 in Lauritzen et al. 2012) both in free-running CAM and DART/CAM. For more details on the stability properties of the divergence damping as implemented on the regular latitude–longitude grid, see Whitehead et al. (2011) and, more generally, for the damping/dispersion properties of the CAM-FV scheme, see Lauritzen (2007) and Skamarock (2008).

e. Parameter estimation

DART/CAM can, in principle, be used to estimate the best values of model parameters, which is an active area of research with important, outstanding questions about the proper application to geophysical models and parameterizations (Anderson 2001; Evensen 2009; DelSole and Yang 2010; Aksoy et al. 2006). The most straightforward technique is the ''augmentation'' method, in which the model parameter of interest is incorporated into the state vector (Jazwinski 1970). Then the assimilation process adjusts the distribution of parameter values to make it more consistent with the observations. This requires modification of the model code to convert the parameter into a variable that can be exported to, and imported from, DART. An early, unpublished, study indicated that application of this technique to estimate the value of parameters in the CAM gravity wave drag scheme does result in analyses with reduced RMS error in observation space. However, the resulting parameters were inconsistent with implicit assumptions made by the developers of the parameterization. For example, the parameter representing the efficiency of momentum transfer into the atmosphere took on negative values in some locations, which contradicted the developer's intention that it represent a quantity that ranges from 0 to 1. This may represent a flaw in the parameterization scheme, or it may merely illustrate the tendency of assimilation schemes to use any available free parameter to improve the fit of the model state to the observations.

5. Applications of DART

a. Ensemble atmospheric reanalysis

DART/CAM has been used to generate an 80-member ensemble reanalysis for the years 2000–10 plus every fifth year from 1974 through 1994 using CAM4 FV with $1.9° \times 2.5°$ resolution. The observations assimilated include all those used in the NCEP–NCAR reanalysis (temperature and wind components from radiosondes, aircraft, and satellite drift winds) plus radio occultation observations from the COSMIC satellites (Anthes et al. 2008) starting in late 2006. There are typically 10^6 observations per day for the 2000–10 period, ramping down to only one-tenth of that for the earliest year, 1974. Analysis of the ensemble mean reveals that the fit to observations for the 2000–10 period is at least as good as that obtained in the NCEP–NCAR reanalysis. A variety of products from this reanalysis was archived and is publicly available.

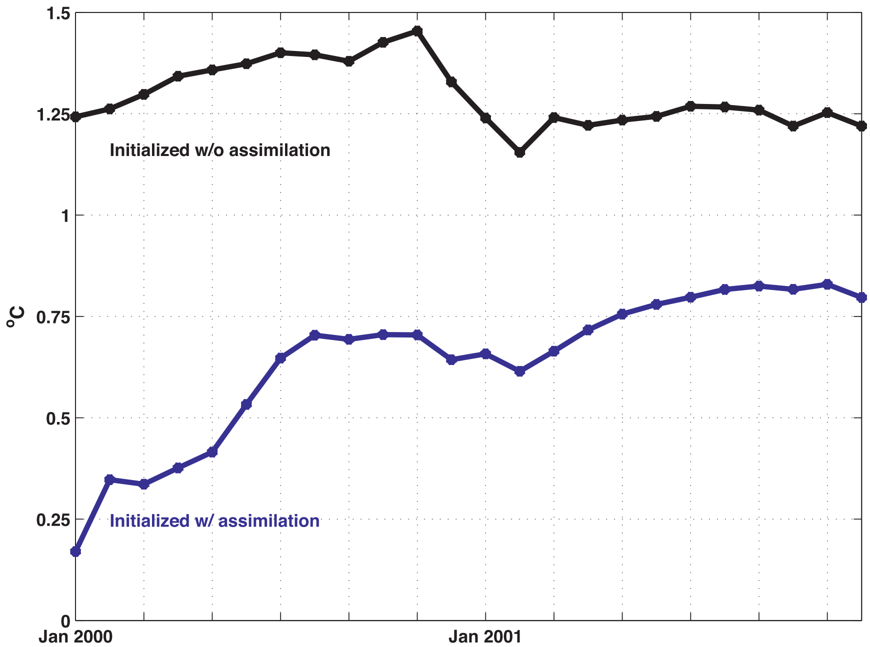

FIG. 7. RMS error of the temperature in the top 250 m of the ocean, from POP monthly-mean forecasts started from hindcast initial conditions (black), and DART/POP analysis (blue), relative to DART/POP analyses.

Ensemble forcing for other CESM climate system component models are available from the CAM reanalyses. These ensemble forcings are essential to produce high-quality ensemble assimilations for the ocean or land surface. The next subsection describes an ocean ensemble reanalysis using DART and version 2 of the Parallel Ocean Program (POP2; Danabasoglu et al. (2012)) with forcing from the CAM reanalysis.

b. Coupled ocean analysis

The next Intergovernmental Panel on Climate Change (IPCC) report (Taylor et al. 2012) will include a section about decadal lead-time predictions of the climate system. It is well known that most of the skill in atmospheric predictions is exhausted in forecasts for longer than a few weeks' lead. For longer lead-time forecasts, the details of the initial conditions of the atmosphere are no longer important. Instead, initial conditions for more slowly varying components of the climate system—in particular, the ocean—become the source of most forecast skill. An ocean ensemble data assimilation using the POP2 ocean component of CESM is one method that has been used to produce initial conditions for decadal predictions.

Maintaining sufficient variability is one of the key problems in ensemble data assimilation; this leads to the need for inflation as noted in section 2a. If all ensemble members of a DART assimilation with POP2 are forced by a single estimate of atmospheric forcing, the resulting analysis spread is found to be far too small. Forcing each POP2 member with a different forcing estimate from the ensemble CAM reanalysis leads to much increased spread and a significantly improved POP2 analysis.

Figure 7 compares the results of one example of a coupled CESM forecast using DART/POP2 initial conditions with a forecast produced from a more traditional method. The black curve represents the previous state-of-the-art product: a forecast initialized from "hindcast" initial conditions, which are generated by forcing the combined CCSM4 ocean (POP2) and sea ice models at the surface with atmospheric data from version 2 of the Common Ocean Reference Experiment (CORE2) dataset (Yeager et al. 2012; Griffies et al. 2009) for a spinup period consisting of four repetitions of the 60-yr span from 1948 to 2007. The long time span is intended to allow the surface forcing to be communicated to the subsurface ocean. The blue curve represents the forecast started from DART/POP2 analyses. These curves of RMSE are calculated relative to the DART/POP2 analyses because there are no independent ocean analyses below the surface that incorporate subsurface observations. While it is standard practice at operational numerical weather prediction centers to compare forecasts against analyses generated using the same forecast model, care is still needed in interpreting these results.

The DART/POP2 forecast RMSE is notably smaller than the hindcast forecast for at least 2 yr. As would be expected, the free DART/POP2 forecasts drift toward the climatological POP model bias with time. There is also evidence that in later years of the 1998–2008 DART/POP2 reanalysis, the ensemble analyses drift away from the observations and toward the model bias. This is partly due to the absence of ensemble spread inflation in the DART/POP2 assimilations, so that the model becomes overconfident and rejects an increasing number of observations. Further analysis and verification of decadal-scale, coupled, hindcast forecasts is available in Yeager et al. (2012), where techniques such as ''bias correction'' are applied to improve decadal forecasts.

c. *Evaluating the utility of novel observations across models*

The DART/CAM system has also proved useful in addressing questions about the utility of unconventional observations. An ensemble DA system is a natural venue to explore this question because it samples the covariance among variables directly from the ensemble. This makes it easy, in principle, to exploit new types of observations, though the observations can only reduce analysis errors if they act across scales resolved by the forecast model.

Observations of clouds, for example, are notoriously hard to exploit in data assimilation systems (Errico et al. 2007) for several reasons. First, available observations (e.g., radar reflectivity, optical thickness) are related indirectly and often nonlinearly to the model state (normally condensate mixing ratios or similar). Second, neither the observed quantities nor the model state are likely to follow Gaussian distributions, as DA has assumed. Indeed, many moisture-related quantities are bounded (with minimum zero) and depend strongly and nonlinearly on temperature. The rank histogram algorithm [section 3b(3)] can help with this issue. Finally, the evolution of cloud properties is largely controlled by parameterized processes with local control, and it has been unclear how much benefit such observations might be expected to provide to analyses and short-term forecasts.

Pincus et al. (2011) focused on the last of these issues with ''perfect model'' experiments and identity observations. Perfect-model experiments (fully supported within DART) draw observations from a single free-running model instance, thus removing systematic model bias, while using identity observations (i.e., observing precisely those variables used to represent clouds in the forecast model) eliminates observational uncertainties. DART was coupled identically to both CAM3.5 and the GFDL climate model AM2 (Anderson et al. 2004), two models with fairly different cloud schemes. Short-term forecasts of all quantities were mildly improved, assimilating perfect observations of clouds but the benefit was more widespread and persistent in CAM than in AM2, in which the cloud parameterization contains an extra degree of freedom (a prognostic equation for cloud fraction) that produces a climatology of clouds more consistent with observations.

6. Summary

CAM, and CESM in general, can now benefit directly from the many tools available in the Data Assimilation Research Testbed. These include direct comparison of model output to real observations from a wide variety of platforms, detection of short-term model biases and some code errors, the ability to start forecasts from analyses that are compatible with CAM and have no foreign model bias, the use of temporally and spatially inhomogeneous analysis error estimates when comparing model output to such analyses, ensemble-based sensitivity analysis of model variables, rapid experimentation with a variety of observations and state variables, straight-forward interchange of model versions, and others not considered in this paper. Such tools complement the traditional evaluation of CAM performance using time and spatial averages of model output. The interfacing of CAM with DART also extends data assimilation beyond immediate numerical weather prediction needs into areas of interest to climate studies, including chemical DA and long-term coupled model forecasting. It appears likely that DA has the ability to improve the speed and quality of climate model development, and should become an integral part of that process.

Acknowledgments. Computing resources were provided by the Climate Simulation Laboratory at NCAR's Computational and Information Systems Laboratory. This research was enabled by CISL compute and storage resources. Bluefire, a 4064-processor IBM Power6 resource with a peak of 77 teraFLOPS, provided almost a million computing hours, and CISL's 12-PB HPSS archive provided over 50 TB of storage. The CESM project is supported by the National Science Foundation and the Office of Science (BER) of the U.S. Department of Energy. Kevin Raeder was partly supported by NASA Grants NNX08AI23G and NNX09AJ05G. We thank Mariana Vertenstein and the Software Engineering Working Group at NCAR for the helpful discussions and multi-instance development work. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Aksoy, A., F. Zhang, and J. Nielsen-Gammon, 2006: Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Mon. Wea. Rev.,* **134,** 2951–2970.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.,* **129,** 2884–2903.

——, 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.,* **131,** 634–642.

——, 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D,* **230,** 99–111.

——, 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus,* **61,** 72–83.

——, 2010: A non-Gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.,* **138,** 4186–4198.

——, 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.,* **140,** 2359–2371.

——, and N. Collins, 2007: Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.,* **24,** 1452–1473.

——, and Coauthors, 2004: The new GFDL global atmosphere and land model AM2–LM2: Evaluation with prescribed SST simulations. *J. Climate,* **17,** 4641–4673.

——, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.,* **90,** 1283–1296.

Anthes, R. A., and Coauthors, 2008: The COSMIC/FORMOSAT-3 mission: Early results. *Bull. Amer. Meteor. Soc.,* **89,** 313–333.

Arellano, A. F., and Coauthors, 2007: Evaluating model performance of an ensemble-based chemical data assimilation system during INTEX-B field mission. *Atmos. Chem. Phys.,* **7,** 5695–5710.

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Mon. Wea. Rev.,* **138,** 1902–1921.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.,* **126,** 1719–1724.

Danabasoglu, G., S. C. Bates, B. P. Briegleb, S. R. Jayne, M. Jochum, W. G. Large, S. Peacock, and S. G. Yeager, 2012: The CCSM4 ocean component. *J. Climate,* **25,** 1361–1389.

DelSole, T., and X. Yang, 2010: State and parameter estimation in stochastic dynamical models. *Physica D,* **239,** 1781–1788.

Deser, C., R. A. Tomas, and S. Peng, 2007: The transient atmospheric circulation response to North Atlantic SST and sea ice anomalies. *J. Climate,* **20,** 4751–4767.

Errico, R. M., P. Bauer, and J.-F. Mahfouf, 2007: Issues regarding the assimilation of cloud and precipitation data. *J. Atmos. Sci.,* **64,** 3785–3798.

Evans, K. J., P. H. Lauritzen, S. Mishra, R. B. Neale, M. A. Taylor, and J. J. Tribbia, 2012: AMIP simulation with the CAM4 spectral element dynamical Core. *J. Climate,* in press.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

——, 2009: The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Syst.,* **29,** 83–104, doi:10.1109/MCS.2009.932223.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.,* **125,** 723–757.

Gent, P. R., and Coauthors, 2011: The Community Climate System Model version 4. *J. Climate,* **24,** 4973–4991.

Griffies, S. M., and Coauthors, 2009: Coordinated Ocean-Ice Reference Experiments (CORES). *Ocean Modell.,* **26,** 1–46, doi:10.1016/j.ocemod.2008.08.007.

Hacker, J. P., and D. Rostkier-Edelstein, 2007: PBL state estimation with surface observations, a column model, and an ensemble filter. *Mon. Wea. Rev.,* **135,** 2958–2972.

Houtekamer, P. L., H. L. Mitchell, and X. Deng, 2009: Model error representation in an operational ensemble Kalman filter. *Mon. Wea. Rev.,* **137,** 2126–2143.

Jazwinski, A., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.

Kay, J. E., and A. Gettelman, 2009: Cloud influence on and response to seasonal Arctic sea ice loss. *J. Geophys. Res.,* **114,** D18204, doi:10.1029/2009JD011773.

——, T. L'Ecuyer, A. Gettelman, G. Stephens, and C. O'Dell, 2008: The contribution of cloud and radiation anomalies to the 2007 Arctic sea ice extent minimum. *Geophys. Res. Lett.,* **35,** L08503, doi:10.1029/2008GL033451.

——, 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.

Kiehl, J. T., J. Hack, G. B. Bonan, B. A. Boville, B. P. Briegleb, D. L. Williamson, and P. J. Rasch, 1996: Dynamics. Description of the NCAR Community Climate Model (CCM3), NCAR Tech. Note TN-420+STR, 29–33.

Klein, S. A., X. Jiang, J. Boyle, S. Malyshev, and S. Xie, 2006: Diagnosis of the summertime warm and dry bias over the U.S. southern Great Plains in the GFDL climate model using a weather forecasting approach. *Geophys. Res. Lett.,* **33,** L18805, doi:10.1029/2006GL027567.

Lauritzen, P. H., 2007: A stability analysis of finite-volume advection schemes permitting long time steps. *Mon. Wea. Rev.,* **135,** 2658–2673.

——, C. Jablonowski, M. Taylor, and R. D. Nair, 2010: Rotated versions of the Jablonowski steady-state and baroclinic wave test cases: A dynamical core intercomparison. *J. Adv. Model. Earth Syst.,* **2,** 15, doi:10.3894/JAMES.2010.2.15.

——, A. Mirin, J. Truesdale, K. Raeder, J. Anderson, J. Bacmeister, and R. B. Neale, 2012: Implementation of new diffusion/filtering operators in the CAM-FV dynamical core. *Int. J. High Perform. Comput. Appl.,* **26,** 63–73, doi:10.1177/1094342011410088.

Lin, S.-J., 2004: A ''vertically Lagrangian'' finite-volume dynamical core for global models. *Mon. Wea. Rev.,* **132,** 2293–2307.

Liu, H., and J. Li, 2010: An improvement in forecasting rapid intensification of Typhoon Sinlaku (2008) using clear-sky full spatial resolution advanced IR soundings. *J. Appl. Meteor. Climatol.,* **49,** 821–827.

Lynch, P., 2006: *The Emergence of Numerical Weather Prediction: Richardson's Dream*. Cambridge University Press, 279 pp.

Machenhauer, B., E. Kaas, and P. H. Lauritzen, 2009: Finite-volume methods in meteorology. *Computational Methods for the Atmosphere and the Oceans,* P. Ciarlet, R. Temam, and J. Tribbia, Eds., Handbook of Numerical Analysis, Vol. 14, North-Holland, 3–120.

Miyoshi, T., Y. Sato, and T. Kadowaki, 2010: Ensemble Kalman filter and 4D-Var intercomparison with the Japanese operational global analysis and prediction system. *Mon. Wea. Rev.,* **138,** 2846–2866.

Neale, R. B., and Coauthors, 2010: Description of the NCAR Community Atmosphere Model (CAM 5.0). NCAR/TN-486+STR, 268 pp.

Phillips, T. J., and Coauthors, 2004: Evaluating parameterizations in general circulation models: Climate simulation meets weather prediction. *Bull. Amer. Meteor. Soc.,* **85,** 1903–1915.

Pincus, R., R. J. P. Hofmann, J. L. Anderson, K. Raeder, N. Collins, and J. S. Whitaker, 2011: Can fully accounting for clouds in data assimilation improve short-term forecasts? *Mon. Wea. Rev.,* **139,** 946–957.

Skamarock, W. C., 2008: A linear analysis of the NCAR CCSM finite-volume dynamical core. *Mon. Wea. Rev.,* **136,** 2112–2119.

Stroeve, J., M. Serezze, S. Drobot, S. Gearheard, M. Holland, J. Maslanik, W. Meier, and T. Scambos, 2008: Arctic sea ice extent plummets in 2007. *Eos, Trans. Amer. Geophys. Union,* **89,** 13, doi:10.1029/2008EO020001.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2009: A summary of the CMIP5 experimental design. World Climate Research Programme Tech Rep., 33 pp. [Available online at http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf.]

——, ——, and ——, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.,* **93,** 485–498.

Taylor, M. A., J. Edwards, and A. St. Cyr, 2008: Petascale atmospheric models for the Community Climate System Model: New developments and evaluation of scalable dynamical cores. *J. Phys.,* **125,** 012023, doi:10.1088/1742-6596/125/1/012023.

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.

Whitehead, J., C. Jablonowski, R. B. Rood, and P. H. Lauritzen, 2011: A stability analysis of divergence damping on a latitude–longitude grid. *Mon. Wea. Rev.,* **139,** 2976–2993.

Xie, S., J. Boyle, S. A. Klein, X. Liu, and S. Ghan, 2008: Simulations of Arctic mixed-phase clouds in forecasts with CAM3 and AM2 for M-PACE. *J. Geophys. Res.,* **113,** D04211, doi:10.1029/2007JD009225.

Yeager, S., A. Karspeck, G. Danabasoglu, J. Tribbia, and H. Teng, 2012: A decadal prediction case study: Late twentieth-century North Atlantic ocean heat content. *J. Climate,* **25,** 5173–5189.

Žagar, N., J. Tribbia, J. Anderson, K. Raeder, and D. Kleist, 2010: Diagnosis of systematic analysis increments by using normal modes. *Quart. J. Roy. Meteor. Soc.,* **136,** 61–76.