An OSSE-Based Evaluation of Hybrid Variational–Ensemble Data Assimilation for the NCEP GFS. Part I: System Description and 3D-Hybrid Results

DARYL T. KLEIST
Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland

KAYO IDE
Department of Atmospheric and Oceanic Science, and Earth System Science Interdisciplinary Center, and Institute for Physical Science and Technology, and Center for Scientific Computation and Mathematical Modeling, University of Maryland, College Park, College Park, Maryland

(Manuscript received 5 November 2013, in final form 12 October 2014)

ABSTRACT

An observing system simulation experiment (OSSE) has been carried out to evaluate the impact of a hybrid ensemble-variational data assimilation algorithm for use with the National Centers for Environmental Prediction (NCEP) global data assimilation system. An OSSE provides a controlled framework for evaluating analysis and forecast errors since a truth is known. In this case, the nature run was generated and provided by the European Centre for Medium-Range Weather Forecasts as part of the international Joint OSSE project. The assimilation and forecast impact studies are carried out using a model that is different than the nature run model, thereby accounting for model error and avoiding issues with the so-called identical-twin experiments.

It is found that the quality of analysis is improved substantially when going from three-dimensional variational data assimilation (3DVar) to a hybrid 3D ensemble–variational (EnVar)-based algorithm. This is especially true in terms of the analysis error reduction for wind and moisture, most notably in the tropics. Forecast impact experiments show that the hybrid-initialized forecasts improve upon the 3DVar-based forecasts for most metrics, lead times, variables, and levels. An additional experiment that utilizes 3DEnVar (100% ensemble) demonstrates that the use of a 25% static error covariance contribution does not alter the quality of hybrid analysis when utilizing the tangent-linear normal mode constraint on the total hybrid increment.

1. Introduction

To combine the advantages of ensemble and variational methods while at the same time attempting to minimize the effects of their weaknesses, hybrid assimilation methods have been proposed and developed by supplementing the ensemble with a static background error covariance (Hamill and Snyder 2000; Lorenc 2003; Buehner 2005). Typically, these hybrid methods utilize the variational framework for the purposes of calculating the analysis increment [hybrid 3D ensemble–variational (EnVar)-based algorithm; Lorenc 2013] although it is possible that one could utilize an alternate framework [e.g., an ensemble Kalman filter (EnKF)]. Many of these hybrid methods with technically different algorithms have been shown to be theoretically equivalent, whether using a combined covariance through brute force or through a variational-based control variable method (Wang et al. 2007a).

Various studies have demonstrated that the hybrid algorithm can in fact improve upon stand-alone variational or ensemble-based algorithms on their own for...
simple (Hamill and Snyder 2000; Etherton and Bishop 2004; Penny 2014), regional (Wang et al. 2007b, 2008b; Liu et al. 2008, 2009; Zhang and Zhang 2012; Liu and Xiao 2013; Zhang et al. 2013), and global operational numerical weather prediction models (Hamill et al. 2011; Buehner et al. 2013; Clayton et al. 2013; Kuhl et al. 2013). The advantages gained over the stand-alone variational systems come from the improved specification of the background error covariance with flow-dependent and multivariate definitions through the use of the ensemble. Additionally, hybrid systems have the ability to improve upon the stand-alone ensemble systems by applying localization in physical space whereas most EnKF algorithms for large NWP applications perform localization in observation space. Physical space localization is particularly helpful for the assimilation of those observations that are not point measurements, but instead integrated quantities such as satellite brightness temperatures (Campbell et al. 2010). This is in part because physical space localization avoids the need to explicitly assign a vertical location for such observations. It has also been suggested that the use of hybrid algorithms can be particularly useful for small ensemble size and large model error as the static error covariance helps to reduce sampling error (Etherton and Bishop 2004; Wang et al. 2007b, 2009).

Most closely related to this work are the previous studies that have focused on operational-like or pre-operational global numerical weather prediction models. A comparison of the variational, EnKF, and EnVar data assimilation algorithms was carried out for global deterministic prediction using Environment Canada’s operational model in Buehner et al. (2010a,b). It was found that using ensemble-based covariances in the EnVar system led to significant forecast improvements in the southern extratropics and modest improvements in the tropics. Two separate studies were carried out to investigate the impact of a hybrid four-dimensional variational data assimilation (4DVar) system relative to 4DVar for the Naval Research Laboratory (Kuhl et al. 2013) and the Met Office (Clayton et al. 2013) global prediction systems. Both studies showed that hybrid 4DVar was an improvement relative to 4DVar for deterministic prediction across a wide variety of variable, levels, and lead times. Kuhl et al. (2013) found that the hybrid algorithm led to particularly large improvements for tropical wind forecasts. It was also found that relying 100% on their ensemble to prescribe the error covariance resulted in a reduction in forecast skill relative to hybrid 4DVar with a blended initial error covariance. For the Met Office system, it was noted that while the inclusion of a hybrid error covariance was an improvement in the 4D context, it was not nearly as large as the improvements found when using the hybrid in the 3D context relative to the three-dimensional variational data assimilation (3DVar; Clayton et al. 2013). Unlike the findings of Kuhl et al. (2013), Clayton et al. (2013) found that forecast error increases for some metrics from the hybrid system when performing verification using self-analysis. However, further comparisons using observations and independent analysis revealed general improvements in the tropics from the hybrid 4DVar, just as in Kuhl et al. (2013).

Using the same software and algorithm as is employed in this study, Wang et al. (2013) carried out impact studies at single, low resolution for comparing 3DVar with a 3DEnVar algorithm for use with the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) model. In general, it was found that the EnVar system produced more skillful deterministic forecasts out to day 5 relative to those initialized using 3DVar, with the largest, consistent improvement in the tropical troposphere. It was also found that for their configuration, the inclusion of a static contribution to the solution did not improve forecasts relative to the EnVar that utilized the ensemble covariance directly. It was also shown that there were additional improvements within the EnVar system by utilizing a dynamic constraint within the minimization. A separate study was carried out by Hamill et al. (2011) to investigate the impact of EnKF and hybrid algorithms for initialization of the NCEP GFS at near-operational resolution for tropical cyclone track prediction. It was found that deterministic track error was reduced when using the hybrid algorithm relative to the then-operational 3DVar scheme.

Motivated by these successes in applying variational-based hybrid algorithms to the NWP data assimilation problem, a hybrid EnVar algorithm has been developed for the operational variational assimilation system at NCEP. To evaluate the impact of using a hybrid algorithm, an observing system simulation experiment (OSSE) is utilized. An OSSE allows one to directly evaluate the actual analysis error since the truth is known. This is in contrast to the aforementioned studies that utilized real observations and focused on forecast impact, where verification is complicated by the fact that the components from which the verification is being performed have errors (analyses, observations, etc.). Additionally, the use of an OSSE makes it possible to evaluate verification for variables such as moisture, which are typically difficult to verify in a real system.

The goal of this study is to not only further corroborate the findings in the previous studies that demonstrate hybrid algorithms improve deterministic global NWP
forecast skill, but to also evaluate the impact on the actual analysis error. Furthermore, unlike in Wang et al. (2013), a dual-resolution strategy is employed where the ensemble used within the EnVar scheme is run at roughly half the resolution of the deterministic component. Such a paradigm is attractive for operational applicability due to computational considerations. However, we aim to corroborate the findings of Wang et al. (2013) in terms of the improvements found as well as the importance of including a dynamic constraint, demonstrating that the results are robust and general. Section 2 describes the components of the NCEP data assimilation system including the implementation of the hybrid algorithm. Section 3 then provides details regarding the Joint OSSE project and simulated observations. This is followed by section 4, which provides a description of the various experimental components. Sections 5 and 6 describe experimental results from assimilating the simulated observations using 3DVar and hybrid 3DEnVar configurations. Last, a summary and motivation for the 4D extension described in Kleist and Ide (2015, hereafter Part II) then follows.

2. NCEP hybrid data assimilation

The gridpoint statistical interpolation (GSI) is a physical-space-based variational analysis scheme (Wu et al. 2002; Kleist et al. 2009b) that has been made operational for several NCEP applications including the global data assimilation system (GDAS) and initialization of the GFS model, used to produce global medium-range deterministic forecast guidance as well as boundary conditions for other applications. Although a variety of minimization algorithms exist within the code, a preconditioned double-conjugate gradient solver with a full-rank background error covariance $B$ (Derber and Rosati 1989) is the default choice within the GSI. Recursive filters are utilized (Purser et al. 2003 a,b) to model some of the off-diagonal components of the background error covariance matrix. Although the GSI does have a 4DVar capability embedded within it, it has yet to be exercised within the operational applications at NCEP. This is mostly due to insufficient development of the necessary tangent-linear and adjoint model codes as well as the significant computational cost associated with 4DVar. Several dynamic and physical constraint options are also available such as the penalty terms for unphysical tracers, incremental normal mode initialization, conservation of global mean dry mass, and a weak constraint digital filter for 4D applications. Prior to the implementation of the hybrid algorithm, the GSI that was run routinely as part of the GDAS utilized a static estimate for the background error covariance, which was supplemented with a tendency-based algorithm to apply flow-dependent variances. An example of the reweighting procedure can be found in Fig. 5 of Saha et al. (2010). The reweighting procedure is fairly simplistic and does not address the off-diagonal nor multivariate deficiencies in the static error covariance estimate. (More detailed information regarding the GSI can be obtained from the Developmental Testbed Center user guide available online at http://www.dtcenter.org/com-GSI/users/index.php.)

Following a similar procedure that is outlined in Wang (2010) as first proposed by Lorenc (2003), the EnVar method is implemented into the GSI using an extended control variable. Initially, the EnVar option was only developed for use with the GFS model, but has since been expanded to other applications such as the NCEP regional and hurricane models. The EnVar option in the GSI includes a dual-resolution capability, where the ensemble perturbations ($x_e$) being used can be at lower resolution than the single deterministic forecast and subsequent analysis. This reduces the computational cost of the ensemble control variable since the dimension of the ensemble control variable ($\alpha$) is the ensemble grid multiplied by the ensemble size. However, some of this cost is offset by the fact that some additional work is necessary to interpolate between analysis and ensemble grids. The dual-resolution strategy may be important for high-resolution operational applications, where it is not affordable to run ensembles at full resolution or with significantly large membership. It should be noted that such a paradigm could be complicated for applications where small scales are important, such as convective scale, tropical cyclone, or cloud assimilation.

For the hybrid system, there are tuning parameters—$\beta_f$ and $\beta_g$—that control how heavily to weight the static (fixed) and ensemble contributions, respectively. The inverses of the parameters are specified and currently assumed to sum to one (see Wang 2010; Kleist 2012; Wang et al. 2013). The localization of the ensemble-based covariance is handled through the application of the error covariance matrix $L$ for the ensemble control variable $\alpha$. The matrix $L$ (block diagonal) is explicitly local with unit amplitude on the diagonal and assumed to be Gaussian. In the GSI for global applications, this function is applied by utilizing spectral correlation functions for the horizontal and a recursive filter in the vertical instead of using quasi-Gaussian functions with compact support as is typically done for the EnKF (some details can be found in Wang et al. 2013). The GSI-specific implementation of the hybrid allows for a level-dependent specification of the decorrelation length scales used in $L$, one each for the horizontal and vertical. Since there is (currently) only a single control
variable $\alpha$ for the entire ensemble perturbation matrix, the effective localization is the same for all variables. For the case of the NCEP GFS, the ensemble perturbation variables are chosen to be virtual temperature, horizontal wind components, relative humidity, cloud mixing ratio, and surface pressure. For surface pressure, the first level of the control variable is utilized, although an option does exist to use a weighted vertical sum as an alternative. Further details on the hybrid algorithm in GSI can be found in Wang (2010). The dual-resolution setup is discussed in more detail in Part II. Results from a single-observation experiment using the GSI-based hybrid are shown in Fig. 2.1 of Kleist (2012).

3. Joint OSSE

An OSSE is typically designed to investigate the potential impact of a future observing system (Masutani et al. 2007, 2010). However, OSSE experiments can also be utilized to investigate various aspects of a data assimilation system such as analysis error (Errico et al. 2007; Wang et al. 2008a; Liu et al. 2009; Privé et al. 2013a). In an OSSE for atmospheric NWP applications, a reference state (nature run) is first generated by making a climate, uninterrupted free run of a NWP model using state-dependent boundary conditions such as observed sea surface temperatures. This free run is then considered to be the true state. Simulated observations are generated by extracting the appropriate information from the nature run and adding realistic errors. The simulated observations can then be used by a data assimilation system to assess their impact on analysis and forecast accuracy with respect to the nature run as the truth. For an OSSE to be useful, it is critical to ensure that the nature run is a suitable representation of the real atmosphere. To achieve realistic results, it is important to use a model within the data assimilation system that is different than that which was used to generate the nature run; if the same model is used for both, the so-called identical-twin experiment, the model error goes unaccounted for.

An international, collaborative effort called the Joint OSSE has formed over the past several years. The European Centre for Medium-Range Weather Forecasts (ECMWF) has generated a nature run for use within the Joint OSSE community by completing a 13-month forecast using cycle 30r1 of their Integrated Forecast System model with T511 horizontal resolution and 91 vertical levels (Andersson and Masutani 2010). The nature run was carefully evaluated to ensure it was realistic in terms of general climatology, storm tracks, as well as clouds and precipitation (Reale et al. 2007; McCarty et al. 2012). From this nature run, scientists at NCEP and the National Aeronautics and Space Administration (NASA) have simulated observations that were operationally available in 2005, including radiosonde, surface, aircraft, satellite-derived atmospheric motion vectors, wind profiler, ship and buoy, and scatterometer-based surface winds. Additionally, satellite microwave and infrared brightness temperature temperatures were simulated [e.g., High Resolution Infrared Radiation Sounder (HIRS), Advanced Microwave Sounding Unit A (AMSU-A), AMSU-B, Microwave Humidity Sounder (MHS), Atmospheric Infrared Sounder (AIRS) on the simulated polar orbiters of National Oceanic and Atmospheric Administration (NOAA)-14, -15, -16, -17 as well as Aqua, in addition to the Geostationary Operational Environmental Satellite (GOES) sounder radiances from simulated geostationary satellites GOES-10 and -12]. The simulated observations have been assimilated into a NWP model and gone through initial validation to ensure their usefulness (Errico et al. 2013). The nature run has been made available to the research community for the purpose of running OSSE experiments.

A subset of calibrated observations covering the simulated period from 1 July to 31 August was generated by NASA Global Modeling and Assimilation Office (GMAO) and made available. Observations were only simulated to correspond to the NCEP GDAS (late cutoff) cycle. Examples of the observations available for a single cycle are shown in Figs. 1 and 2. The errors that were generated and added to the simulated observations from the nature run were calibrated in an attempt to match observation impacts from the real system, evaluated by a series of data-denial experiments in the framework of observing system experiments (Privé et al. 2013a).

4. Experiment design

To test the impact of the various components and aspects of including a hybrid EnVar component to the system, it is necessary to first produce a fully cycled 3DVar run as the control experiment utilizing simulated observations from the Joint OSSE (section 3). The model used in all the assimilation experiments and longer (7 day) forecasts is a lower-resolution version of the NCEP GFS that became operational in May 2011. A description of the GFS model version 9.0.1 is available.
from the NCEP Environmental Modeling Center (EMC) website (http://www.emc.ncep.noaa.gov/GFS/doc.php). No tuning is done to the physical parameterization schemes for this work.

a. 3DVar

The 3DVar control experiment is configured to mimic as closely as possible an operational configuration, in the hope that the OSSE-based results will match the real system. The version of the GFS utilized is a T382 spectral model with 64 hybrid sigma-pressure vertical levels. The static background error estimate for this version of the model is derived using the National Meteorological Center (NMC, now known as NCEP) method (Parrish and Derber 1992) extracting statistics from 24- and 48-h lagged forecast pairs. Similar to operational applications, some tuning of the amplitudes and length scales was performed as in Kleist et al. (2009b). The Tangent Linear Normal Mode Constraint (TLNMC) with eight vertical modes and a single iteration (Kleist et al. 2009a), global mean dry mass weak constraint, and flow-dependent variance reweighting are all utilized. Radiative transfer calculations for the assimilation of satellite radiances within the GSI are performed using version 2.0.4 of the Joint Center for Satellite Data Assimilation (JCSDA) Community Radiative Transfer Model (CRTM; Han et al. 2006). This particular version of the CRTM is slightly different than the version that was used to generate the simulated brightness temperatures from the nature run, creating minor differences that contribute in part to the simulated error generation. Since the simulated observations were only generated for the GDAS (late cutoff) cycle, no early data cutoff (GFS) cycles are performed. Long forecasts are carried out for 7.5 days.
once per day at 0000 UTC from the analysis generated by assimilating the GDAS simulated observations.

The GSI/GFS is cycled through the period for which the simulated observations were made available: 1 July–31 August 2005. An initial condition for 1 July is spun up by assimilating the Joint OSSE simulated observations in a separate, offline, coarse-resolution experiment for the simulated June period. From this initial condition, an
additional two weeks of the experiment is ignored to allow for proper spinup to this experimental configuration. For analysis and forecast verification relative to the ECMWF nature run, all fields are interpolated to a common 1° regular latitude–longitude grid in the horizontal and pressure surfaces in the vertical.

b. 3D hybrid

A hybrid experiment (hereafter referred to as 3DHYB) is carried out that utilizes the variational framework to supplement flow-dependent, ensemble-based error covariance derived from a serial square root filter form of an EnKF (Whitaker and Hamill 2002; Whitaker et al. 2008) to the static error covariance of the 3DVar. The 3DHYB experiment utilizes the dual-resolution capability as in the NCEP operational hybrid GSI. The single high-resolution deterministic GFS has the same configuration as the 3DVar experiment, while the 80-member ensemble is run at a lower T190 spatial resolution in the horizontal with the same 64 hybrid sigma-pressure vertical levels. The 3DHYB experiment uses parameter settings of $\beta_f^{-1} = 0.25$ and $\beta_c^{-1} = 0.75$, so the static contribution to the analysis increment is chosen to be 25% while relying 75% on the ensemble covariances. Figure 2.1 of Kleist (2012) and Fig. 1 of Wang et al. (2013) provide a schematic flowchart of the hybrid cycling scheme. To maintain synergy between the EnKF and hybrid analyses, the ensemble mean is replaced every cycle using a recentering procedure that ensures the EnKF analysis ensemble is always centered about the hybrid analysis that is presumably better. While previous work has demonstrated that the recentering procedure produces little impact (Clayton et al. 2013; Wang et al. 2013), it was decided to utilize it within this dual-resolution context to match the operational NCEP hybrid algorithm.

The version of the EnKF in 3DHYB applies both additive and multiplicative inflation after the ensemble update step. The multiplicative inflation that is used follows the restoration to the prior spread method (Whitaker and Hamill 2012). The multiplicative inflation factor is proportional to the posterior standard deviation reduction from the prior through the assimilation of observations. The posterior (analysis) perturbations $\{x_p^a\}$ are postmultiplied by the prescribed factor. The amplitudes are controlled through the global tuning parameter $[\alpha$ in Eqs. (2) and (3) of Whitaker and Hamill (2012)]. The inflation factor is allowed to be different for each variable, vertical level, and horizontal grid point. The additive inflation extracts perturbations from a database of lagged forecast pairs (24 and 48 h) within a window about the valid analysis time to ensure the perturbations are appropriate for the given season, and the final analysis postperturbation ($x_{ap}^n$) becomes

$$x_{ap}^n = x_1^r + \kappa x_2^r. \quad (1)$$

Each ensemble member, $n$, is assigned a different quasi-random perturbation vector, $x_1^r$. The lagged forecast database contains cases spanning an entire year at the rate of 4 day$^{-1}$ associated with the data assimilation cycling frequency. There is a single amplitude parameter $\kappa$ to rescale the perturbations that are extracted from the lagged forecast pairs. The inflation was iteratively tuned in an effort to match the time mean, zonal mean background spread with the actual background error, focusing primarily on the troposphere and lower stratosphere. The inflation parameters in the 3DHYB experiment were set to $\kappa = 0.2$ for the additive inflation and a global constant of 0.7 for the multiplicative inflation (restoration of spread).

The same localization parameters as in 3DHYB are used for the EnKF but in the observation space. A level-dependent specification for the horizontal decorrelation length scales that is derived from an ensemble of forecasts similar to Pannekoucke et al. (2008) is used (Fig. 3). A single value of 0.5 scale heights is used for the vertical decorrelation in the hybrid. A conversion factor is applied in the EnKF to match the length scale of the two localization definitions [i.e., distance to zero within the Gaspari and Cohn (1999) framework for the EnKF vs the Gaussian e-folding distance for the hybrid].

For the data selection and the quality control, 3DHYB utilizes the same procedure as the 3DVar control does using the high-resolution background. However, the exact data selection and counts are allowed to differ for any given cycle based on GSI internal quality control checks such as gross error, as well as decisions made by the variational quality control procedure within the minimization. The EnKF also utilizes the GSI for all of the observation operators. To ensure that identical observations are chosen for all of the ensemble members, the data selection and quality control decisions are driven by the low-resolution, background ensemble mean. Additionally, a coarser 225-km mesh is used for the satellite thinning procedure of the EnKF rather than a 150-km mesh used for the high-resolution update. This results in fewer observations being assimilated into the lower-resolution EnKF than the deterministic hybrid analysis. Since the EnKF used in this study is a serial, square root filter, the reduced number of satellite observations assimilated helps improve computational efficiency.

---

4 A GSI-based 3D hybrid EnVar algorithm became operational for use in the NCEP GFS/GDAS system in May 2012.
The analysis increment derived from the 3DHYB update utilizes the TLNMC (Kleist et al. 2009a) on the total analysis increment. Applying the constraint in this manner maintains consistency in terms of the balance and noise characteristics of the analysis increment between the 3DVar and 3DHYB experiments. The total analysis increment ($x_0^{t}$) that is used in the observation term then becomes

$$x_0^{t} = C \left[ x_f' + T \sum_{n=1}^{N} (\alpha_n \circ x_n^{\mu}) \right],$$

(2)

where $C$ represents the forward tangent-linear normal mode constraint operator as described in Kleist et al. (2009a), $x_f'$ is the contribution that comes from the static part of the hybrid covariance, and $T$ represents the transformation from ensemble to deterministic resolution. As in the 3DVar case and described in Kleist et al. (2009a), eight vertical modes and a single iteration are used for $C$. In addition to the consistency between the increments within the 3DVar and 3DHYB paradigms, the use of such a constraint can act to help ameliorate potential imbalance and noise introduced through the localization of the control variable as well as sampling error inherent to the ensemble covariances. In all EnVar experiments, the TLNMC is applied to the deterministic analysis increment in the GSI, and the individual ensemble members are filtered with the GFS full-field digital filter. The individual members only feel the benefit of the TLNMC through the ensemble mean via the recentering procedure. Wang et al. (2013) demonstrated that the use of this constraint resulted in improved forecast skill within the context of a single resolution hybrid experiment.

The initial condition of the single, high-resolution deterministic forecast is identical to that of the 3DVar control. The initial ensemble is generated by taking the same initial condition for 1 July that is utilized in the control experiment and performing the additive inflation procedure with 100% larger $\kappa$ in Eq. (1). As with the control, the first two weeks of the experiment are ignored to allow for the system to spin up. Finally, to better understand the 3DHYB results, an additional 3D experiment is designed. The experiment (3DENVAR) investigates the direct use of the ensemble covariance by setting $\beta_x^{-1} = 1$ and $\beta_f^{-1} = 0$. All other settings including inflation and localization parameters are kept the same. This experiment is analogous to running a high-resolution analysis of 3D-EnKF using a low-resolution forecast ensemble but in a variational framework and utilizing the TLNMC.

5. 3DVar results

As an initial validation of the experimental configuration, an evaluation of the zonal mean square root of the variance in the analysis increments from the 3DVar experiment is carried out as in Errico et al. (2007). The zonal wind increment in the OSSE-based 3DVar exhibit two local maxima (Fig. 4, top right), associated with the extratropical jets and consistent with the background error variance specification. Two secondary maxima are also noted in the near-surface extratropics, with larger amplitude in the Southern Hemisphere consistent with the season for which the observations are assimilated.

The increments from the OSSE-based 3DVar experiment are then compared in a qualitative sense to a system that assimilated real observations over a similar time period. To perform this comparison, analysis increments are extracted from an experiment that utilizes a T382 version of the GFS model as well as the GSI for assimilation (Kleist et al. 2009b). One has to keep in

---

5 The experiment was part of a preimplementation run to test the impact of replacing a spectral analysis with the GSI as part of the GDAS/GFS system.
mind that the underlying atmospheric state is different in this comparison, since the OSSE experiment utilized a model error-free run as the truth (see section 3). The zonal wind increment in the real system exhibits similar structure to that in the OSSE run, with maxima associated with the extratropical jets and a secondary maximum in the near-surface Southern Hemisphere extratropics (Fig. 4, top left). Interestingly, the magnitude of the low-level maxima is similar between the OSSE and real observation experiments, with larger
discrepancies in amplitude associated with the jet level maxima. The real system also exhibits two local maxima in each hemisphere at jet level: one associated with the tropics/subtropics and a second maximum associated with the midlatitude jet. The OSSE, on the other hand, does not exhibit these two distinct features, and instead has a much broader structure across latitude.

A comparison of other variables shows that the OSSE and real observations are quite similar in terms of incremental statistics. For example, the temperature increment exhibits very similar structure and amplitude between the two experiments (Fig. 4, middle row). Both experiments exhibit distinct maxima in the temperature increments poleward of 30° in both hemispheres, with separate local maxima near the surface and near jet level. Other than an odd feature in the lower troposphere right over the South Pole in the real observation experiment, the OSSE compares remarkably well. Likewise, the surface pressure analysis increment is very similar between the two runs (Fig. 4, bottom row), with
the OSSE exhibiting smaller amplitude. The statistics described here are similar to the findings in Errico et al. (2007) and Errico et al. (2013) where the largest discrepancies between the OSSE-based and real observation systems seem to be associated with the wind increments.

It is to be expected that there will be some quantitative discrepancies between the real analysis system and the OSSE (Errico et al. 2007). However, the focus of this work is not the impact of particular observations or the evaluation of real analysis error, and instead the relative and potential improvement in the quality of analysis resulting from algorithmic changes such as the introduction of the hybrid. In this sense, the only expectation is that one can produce something that at least behaves similarly to the real system. Of course, one would hope that the OSSE and the real system are close enough so that the findings of the OSSE-based experimentation would translate directly to the real system.

6. 3D hybrid results

a. Analysis comparison

The analysis error structure is quite similar between the 3DVar and 3DHYB experiments (Fig. 5, left and middle columns), exhibiting zonal wind error local maxima in the upper troposphere associated with jets, large temperature errors in the lower troposphere poleward of 60°S, and lower-tropospheric maxima in specific humidity errors. Despite the obvious similarities, several differences immediately stand out. A plot of the difference between the 3DHYB and 3DVar experiments reveals that analysis errors are consistently reduced in the hybrid
system for wind, temperature, and moisture (Fig. 5, right column). There is a small region where temperature errors are slightly increased in the 3DHYB experiment over the southern high latitudes in the upper troposphere and stratosphere.

A comparison of the 300-hPa zonal wind analysis error standard deviation between the 3DHYB and 3DVar experiments reveals that the biggest impact from the hybrid system can be found in the tropics (Fig. 6), where the 3DVar control appears to have the biggest problems. This large improvement in the tropics is similar to the findings of Kuhl et al. (2013) and Wang et al. (2013). In particular, the large local maxima in error over the equatorial Indian and Atlantic Ocean regions, associated with deep convection, are significantly reduced in 3DHYB. The difference in analysis errors (Fig. 6, bottom) reveals that 3DHYB is almost uniformly better than the 3DVar control for the simulated August period. Interestingly, the errors in the 3DVar control are quite large over the Southern Ocean, as expected, but quite small over the central and North Pacific. This is likely an artifact of the errors that were added to the simulated observations and trying to match the impact of each observing platform relative to realistic observing system experiments (OSEs; N. C. Privé 2010, personal communication).

Some interesting behavior in both the 3DVar and 3DHYB experiments is revealed by the time series of the analysis and background error standard deviations (Fig. 7). In general, the analysis error is smaller than the background error for any given cycle, though there are rare exceptions. For midtropospheric zonal wind, the background errors for 3DHYB are significantly smaller than the analysis errors from the 3DVar control. Interestingly, the magnitude of the difference between the analysis errors and background errors for 500-hPa zonal wind is much larger in the 3DVar control. For 850-hPa temperature, the behavior is quite different with a clear signal by cycle. The temperature errors are a minimum for the 0000 UTC analysis cycle each day in both the 3DVar and 3DHYB experiments. A subtle, secondary minimum in analysis error is associated with the 1200 UTC cycle. The smaller errors at 0000 and 1200 UTC can be attributed to the addition of the radiosonde network for those times. The zonal wind errors appear to have a bit of cyclical nature to them as well, but the pattern is much more difficult to ascertain among the much larger
day-to-day variability relative to the temperature errors. Consistent with previous discussion, the temperature errors between the 3DVar and 3DHYB experiments are much more similar to each other and exhibit much less variability, attributable to the fact that the temperature analysis errors are anchored substantially through the assimilation of satellite radiances.

Although some tuning was performed for the static error covariance as well as the ensemble inflation parameters to help control the ensemble spread, it is evident from a comparison with the actual background errors that there is still significant room for improvement (Fig. 8). For the comparison of the 3DVar experimental background error with the tuned static covariance, it can be seen for zonal wind that the static error estimate is much too large in the extratropics, particularly at jet level, and deficient in the tropics. Although the ensemble spread for zonal wind in the 3DHYB experiment matches the actual background error a bit better than the static estimate, it seems to be under-representative of the errors outside of the extratropical jet regions, and overdispersive in the troposphere at the southern high latitudes. Furthermore, the spread was tuned through the global inflation parameters with a focus primarily on the troposphere. This results in an underdispersive ensemble for upper-stratospheric temperature (Fig. 9), due primarily to the lack of treatment of model error, significant differences in the nature run and assimilation model at these levels, and reduced amplitude additive perturbations to help compensate.

Within a perfect model, identical-twin OSSE configuration, Wang et al. (2008a) were able to get a good match between ensemble spread and background error. Here, several other sensitivity experiments were run in an effort to get a better match between the ensemble spread and actual background errors. However, this proved to be quite difficult given the ad hoc nature of the inflation, the fact that global parameters are utilized, and the imperfect model setting. This is further evidence that a more proper treatment of system (including model) error is needed, perhaps through the use of something like stochastic physics (Whitaker and Hamill 2012) in the model.

b. Forecast impact

It is clear that the analysis quality in 3DHYB is superior relative to that from the 3DVar control. However, for operational NWP, it is imperative that the improved analyses also translate to improved forecasts. To test the impact of the hybrid in this framework, forecasts are initialized once per day at simulated 0000 UTC, and run out to 7.5 days using the same forecast model that was used for data assimilation cycling, starting after the two-week spinup. Forecasts from the 3DVar and 3DHYB analyses are then verified by comparing them with the ECMWF nature run, instead of doing verification relative to its own analysis as is standard practice in operational NWP. It is important to note that in this context, it is the ability to recover the ECMWF forecast model result that is being demonstrated. This does not necessarily translate to the ability of forecasting the real atmosphere.

The forecasts generated by starting from the 3DVar analyses exhibit behavior similar to the real system, at least in terms of geopotential height errors. The errors generally increase significantly with time, especially at jet level and near the top of the model (Fig. 10). The use of the hybrid analyses results in an improvement for tropospheric heights, particularly in the Southern Hemisphere at longer lead times. However, it should be noted that the hybrid analysis does result in larger height errors in the upper stratosphere, consistent with the underdispersive ensemble for temperature in the same region. Just as was the case for the background and analysis errors for zonal wind, the use of 3DHYB analyses results in significantly improved forecasts for vector
wind in the tropics for all levels and all lead times (Fig. 11). The upper-stratospheric wind forecasts are improved in the hybrid (Fig. 11) despite that fact that the global heights are degraded in the same region (see Fig. 10 for extratropics, not shown for the tropics). The improvements gained by utilizing the hybrid system instead of 3DVar are statistically significant for most variables, levels, and lead times (Fig. 12). The forecast improvements are particularly impressive in the tropics and Southern Hemisphere, similar to the findings of
Buehner et al. (2010b), Kuhl et al. (2013), and Wang et al. (2013). The results shown in Fig. 12 are generally representative of other variables and levels, consistent with the analysis error reduction (Fig. 5), including moisture.

**c. 3DENVAR experiment**

To test the impact of the inclusion of a static error contribution to the solution, a 3DENVAR experiment (3DENVAR) is carried out and examined. The 3DENVAR experiment uses an identical configuration to that of the 3DHYB experiment, except to set the static contribution to zero and increase the ensemble contribution to 100%. The TLNMC is utilized in an effort to test only the impact of the static error contribution. For practical reasons and a clean comparison, the dual-resolution configuration is maintained despite the fact that the ensemble has no information in part of the spectrum of the incremental solution. The analysis errors in the 3DENVAR experiment relative to the 3DVar configuration are found to be qualitatively and quantitatively similar to those produced in the 3DHYB experiment (Fig. 13). The static error covariance does not appear to be adding much for this particular configuration, though it is not making things worse either (not shown), similar to the findings of Wang et al. (2013). This result is not guaranteed to hold true for other dual-resolution configurations with real observations or smaller ensemble sizes. As was found in Kleist et al. (2009a) and Wang et al. (2013), additional experiments demonstrate that the TLNMC has a positive impact in the 3DENVar and hybrid paradigm (not shown). It is possible that the impact of a static $B$ in a configuration without the TLNMC will be different, and is investigated in more detail within the 4D hybrid in Part II.

7. **Summary and conclusions**

Hybrid data assimilation algorithms that attempt to combine the strengths of variational and ensemble algorithms while attempting to minimize the effects of their weaknesses have become popular within the data assimilation community. This is true at NCEP where efforts have been put forth to develop and test a hybrid system, which in fact became operational in May 2012 for the global data assimilation system. For an operational center, a hybrid EnVar algorithm allows for an easier transition toward using ensemble-based, flow-dependent error covariances without a complete paradigm shift by incorporating ensemble perturbations into well-established variational algorithms. There are advantages to maintaining the variational framework beyond the practical aspects, such as supplementing the ensemble estimate with a static error covariance, the application of already established dynamic constraints, and localization in physical space.
Although impact tests with real observations have corroborated that hybrid algorithms can prove superior to their stand-alone variational and ensemble counterparts for global numerical weather prediction (Buehner et al. 2010b; Hamill et al. 2011; Buehner et al. 2013; Clayton et al. 2013; Kuhl et al. 2013; Wang et al. 2013), OSSEs provide a platform for which to evaluate the characteristics of the analysis error since the truth is known. With this in mind, a series of experiments are carried out to evaluate the impact of a hybrid data assimilation algorithm relative to 3DVar control through the assimilation of observations generated from an ECMWF-produced nature run. The errors that are added to the simulated observations were done in such a manner that their impact on the analysis and forecast skill within the OSSE framework matched their counterparts from real observation OSEs. Using the NCEP GFS model, 3DVar and 3DHYB experiments are carried out over a simulated two-month period in a framework that closely resembles a real-world NWP application.

Consistent with previous studies, the quality of analyses and forecasts from the hybrid experiment are generally superior to those derived from the 3DVar control. In particular, the analysis errors for wind and moisture are substantially reduced. The largest, consistent positive impact appears to be in the tropics, in agreement with previous studies. The 3DHYB experiment yields a reduction in forecast error for most metrics, levels, and lead times. However, the upper-stratospheric heights are degraded likely due to the ensemble being under-dispersive for temperature at these same levels. An experiment utilizing 3DEnVar is compared to the original

**Fig. 12.** Percent change in root-mean-square error from 3DHYB – 3DVar for the period covering 27 Jul–1 Sep 2005 in the Northern Hemisphere (green), tropics (red), and Southern Hemisphere (blue) for selected variables as a function of forecast lead time. The forecast variables include (top left and top right) 500-hPa geopotential height, (top middle) 850-hPa vector wind, (middle) 200-hPa vector wind, and (bottom) 700-hPa specific humidity. All verification is performed against the ECMWF nature run. The error bars represent the 95% confidence threshold for a significance test.
3DHYB run and shows that within this configuration and set of parameters, the static error contribution does not improve the quality of analysis, consistent with Wang et al. (2013). However, this is likely not a general result and is not guaranteed to hold true for dual-resolution configurations with real observations or when utilized without the TLNMC filtering.

It is encouraging that the forecast impact seen from the hybrid within this framework is somewhat consistent with results that have been achieved for more realistic systems that assimilate real observations (chapter 5 of Kleist 2012; Wang et al. 2013) and similar to the findings of Privé et al. (2013b). This system then provides a framework for which to experiment with ways to improve the hybrid system, with the expectation that gains achieved in this framework will translate to the real NWP and data assimilation systems. Two anonymous reviewers helped to significantly improve the manuscript.

Acknowledgments. This work was completed as part of the first author’s Ph.D. thesis at the University of Maryland, College Park, made possible through the support of EMC management, especially Steve Lord, John Derber, and William Lapenta. The ECMWF nature run was provided by Erik Andersson through arrangements made by Michiko Masutani. We are grateful to the Joint Center for Satellite Data Assimilation for providing access to their supercomputers to perform the experiments. The support of ONR Grants N000140910418 and N0001410557 is gratefully acknowledged. The authors wish to thank Nikke Privé and Ron Errico for kindly providing access to the calibrated, simulated observations. Finally, the authors wish to thank Dave Parrish and Jeff Whitaker for contributing to various aspects of the hybrid, EnVar, and EnKF developments. Two anonymous reviewers helped to significantly improve the manuscript.

REFERENCES

Andersson, E., and M. Masutani, 2010: Collaboration on observing system simulation experiments (Joint OSSE). ECMWF Newsletter, No. 123, ECMWF, Reading, United Kingdom, 14–16.


