First Application of the Local Ensemble Tangent Linear Model (LETLM) to a Realistic Model of the Global Atmosphere[Ⓔ]

SERGEY FROLOV

Marine Meteorology Division, Naval Research Laboratory, Monterey, California

DOUGLAS R. ALLEN

Remote Sensing Division, Naval Research Laboratory, Washington, D.C.

CRAIG H. BISHOP AND ROLF LANGLAND

Marine Meteorology Division, Naval Research Laboratory, Monterey, California

KARL W. HOPPEL AND DAVID D. KUHL

Remote Sensing Division, Naval Research Laboratory, Washington, D.C.

(Manuscript received 1 November 2017, in final form 25 May 2018)

ABSTRACT

The local ensemble tangent linear model (LETLM) provides an alternative method for creating the tangent linear model (TLM) and adjoint of a nonlinear model that promises to be easier to maintain and more computationally scalable than earlier methods. In this paper, we compare the ability of the LETLM to predict the difference between two nonlinear trajectories of the Navy's global weather prediction model at low resolution (2.5° at the equator) with that of the TLM currently used in the Navy's four-dimensional variational (4DVar) data assimilation scheme. When compared to the pair of nonlinear trajectories, the traditional TLM and the LETLM have improved skill relative to persistence everywhere in the atmosphere, except for temperature in the planetary boundary layer. In addition, the LETLM was, on average, more accurate than the traditional TLM (error reductions of about 20% in the troposphere and 10% overall). Sensitivity studies showed that the LETLM was most sensitive to the number of ensemble members, with the performance gradually improving with increased ensemble size up to the maximum size attempted (400). Inclusion of physics in the LETLM ensemble leads to a significantly improved representation of the boundary layer winds (error reductions of up to 50%), in addition to improved winds and temperature in the free troposphere and in the upper stratosphere/lower mesosphere. The computational cost of the LETLM was dominated by the cost of ensemble propagation. However, the LETLM can be precomputed before the 4DVar data assimilation algorithm is executed, leading to a significant computational advantage.

1. Introduction

Variational data assimilation (DA) methods based on tangent linear models (TLMs) continue to show superior forecast accuracy when compared to pure ensemble-based methods or TLM-free hybrid methods (Lorenc et al. 2015;

Bowler et al. 2017). However, such TLMs are difficult to maintain and develop, especially for high-resolution models of the coupled Earth system. Local ensemble TLMs (LETLMs) (Frolov and Bishop 2016; Allen et al. 2017; Bishop et al. 2017) can potentially replace traditional TLMs by using a linear matrix operator constructed from an ensemble of nonlinear forward model integrations. A hybrid four-dimensional variational data assimilation (hybrid-4DVAR) algorithm based on an LETLM can combine the benefits of traditional 4DVAR methods, such as propagation of the static covariance through time and four-dimensional localization of the ensemble covariance, with the low development costs

[Ⓔ] Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/MWR-D-17-0315.s1>.

Corresponding author: Sergey Frolov, sergey.frolov@nrlmry.navy.mil

and high computational scalability of the TLM-free methods like hybrid-4D EnVAR (Buehner et al. 2010; Kleist and Ide 2015). In fact, our previous work (Allen et al. 2017; Bishop et al. 2017) showed that it is possible to replace a traditional TLM within the context of a hybrid-4DVAR system with no loss in assimilation skill. However, these previous papers developed and tested ensemble-based TLMs only in the context of simplified models: a coupled 1D Lorenz model (Bishop et al. 2017) and a global shallow water model (Allen et al. 2017). To date, there have been no demonstrations that the LETLM can show comparable skill to a traditional TLM in the context of a realistic 3D model of the Earth's atmosphere.

In this paper, we compare the ability of the traditional TLM and the LETLM to propagate 4DVAR analysis increments generated by an operational weather prediction system run at a degraded resolution. Specifically, we used the Navy Global Environmental Model (NAVGEN) with 2.5° horizontal resolution and 60 vertical levels (surface to about 65 km). We used the degraded resolution to expedite the development of the system. We plan to evaluate the applicability of our findings to a higher-resolution model in future publications.

We compared the ability of the LETLM to propagate an analysis increment against a pair of nonlinear forecasts and against the operational TLM that linearizes the Navy Operational Global Atmospheric Prediction System (NOGAPS)—the Eulerian predecessor to NAVGEN. Our choice of NOGAPS TLM reflects the reality that TLMs used by operational centers often lag the developments in the forecast model. Our tests evaluated the ability of the LETLM to propagate increments over a wide range of altitudes, including the atmospheric boundary layer, troposphere, stratosphere, and lower mesosphere, and also examine the sensitivity of the LETLM to inclusion of radiative forcing and physics parameterizations.

The layout of the paper is as follows. Section 2 describes the NAVGEN, the traditional TLM, and the LETLM. Section 3 details the experimental design, error metrics, and tuning methodology, while the appendix describes the efficient matrix inverse implementation. The experimental results are presented first in section 4, focusing on the tuning of a reference configuration for the LETLM and detailed evaluation of the LETLM performance for a single date. Sensitivity of the LETLM to variations of the model parameters is described in section 5. In section 6, we evaluate the performance of the LETLM for seven different dates, with the goal of establishing statistical significance of the results established for a single date in sections 4 and 5. Section 7 quantifies the computational efficiency and scalability of the LETLM.

The summary and conclusions are provided in section 8, and some thoughts on future work are given in section 9.

2. Model description

a. NAVGEN

This study uses the global atmospheric NAVGEN (Hogan et al. 2014), which uses a semi-Lagrangian–semi-implicit integration of the hydrostatic dynamical equations, the first law of thermodynamics, and conservation of moisture and ozone. This study uses a 60-level hybrid sigma-pressure coordinate (top at 0.05 hPa, or about 65 km) as described in Eckermann et al. (2009). To facilitate multiple sensitivity experiments, the model was run at relatively low resolution of T47 (144 longitudes $\times$ 72 latitudes, for a Gaussian grid spacing of $\sim 2.5^\circ$ at the equator). The model time step was 1800 s. The same forecast configuration and resolution was used for the control and the ensemble forecasts. The predicted variables of the NAVGEN are vorticity, divergence, virtual potential temperature, specific humidity Q , and surface pressure P_s . In addition, the zonal U and meridional wind V , temperature T , geopotential height Z , σ (sigma dot; vertical motion), and 3D pressure P are derived fields.

The NAVGEN code incorporates several stochastic physics packages, including stochastic kinetic energy backscatter (SKEB), nonorographic gravity wave drag, and a stochastic mass flux parameterization in the boundary layer. Stochastic processes are difficult to model from a linear perspective, since there is no deterministic process that results in the random physics change. For the sake of this initial pilot study with NAVGEN, we decided to turn off these stochastic processes for testing the propagation of analysis increments.

The NAVGEN DA solver is a hybrid-4DVAR system (Kuhl et al. 2013) that employs a strong-constraint approach with an ensemble system based on the ensemble transform (ET) method. The cycling window is 6 h, centered at 0000, 0600, 1200, and 1800 UTC. The background includes the last 6 h of a 9-h forecast initialized at the center of the previous analysis time window. The solver employs the accelerated representer (AR) method (Xu et al. 2005; Rosmond and Xu 2006).

b. Traditional tangent linear model

The NAVGEN TLM and adjoint models are described in Rosmond (1997). While the nonlinear forecast model is semi-Lagrangian, the TLM uses a spectral code based on the earlier NOGAPS. This earlier version of the NOGAPS TLM is currently used in the operational 4DVAR system, while a new semi-Lagrangian TLM was still in the development stage at the time of the

publication. The TLM is run with the same horizontal and vertical resolution as the nonlinear model, but with a smaller time step of 900 s. The TLM neglects several physical processes, such as gravity wave drag, radiation, and ozone photochemistry. We will refer to this TLM as the NOGAPS TLM in this paper.

c. Local ensemble tangent linear model

In our previous papers (Allen et al. 2017; Frolov and Bishop 2016; Bishop et al. 2017), we described an ensemble-based TLM $\mathbf{M}_{m+1,m}$ that approximates the linear propagation of a perturbation from an earlier time m to a later time $m + 1$:

$$\begin{aligned}\delta\mathbf{x}_{m+1} &= \mathbf{M}_{m+1,m}[\mathbf{x}_m + \delta\mathbf{x}_m] - \mathbf{M}_{m+1,m}[\mathbf{x}_m] \\ &= \mathbf{M}_{m+1,m}\delta\mathbf{x}_m + \xi_m,\end{aligned}\quad (1)$$

where $\mathbf{x}_m$ and $\mathbf{x}_{m+1}$ are the states of the model at times m and $m + 1$; $\mathbf{x}_{m+1} = \mathbf{M}_{m+1,m}[\mathbf{x}_m]$ is the nonlinear model of the forward dynamics; $\delta\mathbf{x}_m$ is the initial-time perturbation; and ξ_m is the linear truncation error.

In this paper, we focus on a specific form of the ensemble-based TLM—the LETLM—that predicts the future state at each grid point based on the information in a surrounding influence volume. An example of such an influence volume is shown in Fig. 1, where the points in the influence volume with radius 2000 km are shown with red dots, and the location of the predicted point is shown with a green dot located near Chicago (41°N, 90°W). The influence volume also has a vertical dimension (discussed below), which may incorporate multiple model levels above or below the level that is being predicted, resulting in a three-dimensional influence volume in the shape of a cylinder.

The ensemble regression model for each grid point p can be formulated as

$$\mathbf{M}_{m+1,m}(p) = \mathbf{S}_p \mathbf{X}_{m+1} \left[(\mathbf{S}_p^{\text{vol}} \mathbf{X}_m)^T (\mathbf{S}_p^{\text{vol}} \mathbf{X}_m) \right]^{-1} (\mathbf{S}_p^{\text{vol}} \mathbf{X}_m)^T, \quad (2)$$

where $\mathbf{M}_{m+1,m}$ is one time step of the LETLM (e.g., 1 h in our experiments). Here, $[(\mathbf{S}_p^{\text{vol}} \mathbf{X}_m)^T (\mathbf{S}_p^{\text{vol}} \mathbf{X}_m)]^{-1}$ is a pseudoinverse of the matrix $[(\mathbf{S}_p^{\text{vol}} \mathbf{X}_m)^T (\mathbf{S}_p^{\text{vol}} \mathbf{X}_m)]$. (We describe details of the efficient implementations of the pseudoinverse in the appendix.) The terms $\mathbf{X}_m$ and $\mathbf{X}_{m+1}$ are the ensemble matrices that store normalized ensemble perturbations for the past ($t = m$) and future ($t = m + 1$) states of the model, where t is a time index; $\mathbf{S}_p^{\text{vol}}$ is a matrix operator that selects all variables within a given influence volume (e.g., red dots in Fig. 1 for a single layer) for the prediction point p ; and $\mathbf{S}_p$ selects just the variables at the predicted point (central point

FIG. 1. An example of an LETLM influence volume ($L = 2000$ km) on a T47 Gaussian grid. Red dots show the influence region associated with the central point shown in green, located at 41°N, 90°W. There are 200 points in this influence volume.

located next to Chicago in Fig. 1). In the vertical, we include points within $\pm z_{\text{halo}}$ levels of the central level, where z_{halo} is an integer 0, 1, 2, In the central layers of the model, the number of levels included is $2 \times z_{\text{halo}} + 1$, while near the model top and bottom, the influence volume is truncated by the number of available levels. In our implementation, the size of the influence volume $\mathbf{S}_p^{\text{vol}}$ was based on a Cartesian distance tuned for each vertical level (see details of the tuning in section 3d).

For this study, we obtained ensemble perturbations by differencing about the ensemble mean. (An alternative would be to obtain the perturbations by differencing about the 4DVAR best-guess trajectory of the model. In our limited experience, both methods produce an equivalent result.) To improve the condition number for the matrix inverse in Eq. (2), we normalized the variables by the root-mean-square perturbation amplitude $n(k, v)$ of each variable averaged horizontally and over the entire ensemble for each model level k and variable type v as follows:

$$n(k, v) = \sqrt{\frac{1}{n^{\text{lat}} n^{\text{lon}} n^{\text{ens}}} \sum_{i_{\text{lat}}=1}^{n^{\text{lat}}} \sum_{i_{\text{lon}}=1}^{n^{\text{lon}}} \sum_{j=1}^{n^{\text{ens}}} x(i_{\text{lat}}, i_{\text{lon}}, j, k, v)^2}, \quad (3)$$

where $x(\cdot)$ are ensemble perturbations; n^{lat} and n^{lon} are the number of latitude and longitude points on the grid; and n^{ens} is the number of ensemble members.

We assemble the solutions for each individual grid point in Eq. (2) into a global solution using the following summation notation:

$$\mathbf{x}_{m+1} = \sum_{p=1}^{N_{\text{grid}}} \left\{ [\mathbf{S}_p]^T \mathbf{M}_{m+1,m}(p) [\mathbf{S}_p^{\text{vol}} \mathbf{x}_m] \right\}, \quad (4)$$

where N_{grid} is the total number of grid points. Here, the $[\mathbf{S}_p]^T$ operator embeds the local solution $\mathbf{x}_m(p)$ onto a global grid (e.g., the green dot in Fig. 1), and $\mathbf{S}_p^{\text{vol}}$ selects the variables within the influence volume (e.g., the red dots on Fig. 1). For NAVGEM, the operator $[\mathbf{S}_p]^T \mathbf{M}_{m+1,m}(p) [\mathbf{S}_p^{\text{vol}} \mathbf{x}_m]$ computes several rows of the LETLM matrix—one row for each predicted variable (e.g., U , V , and T) at the p th grid point (green dot on Fig. 1). The number of variables used as both predictors and predictands varied among our experiments, as described in detail in section 5e.

3. Experiment setup

a. Cycling data assimilation system

For the LETLM experiments described in this paper, we required example analysis increments to propagate with the LETLM and, as explained in section 3b, a database of analysis increments to use as perturbations to initialize the 6-h ensemble forecasts required by the LETLM. The analysis increments for both the test cases and the ensemble perturbations were obtained from a run of the hybrid-4DVAR system of Kuhl et al. (2013). We configured the hybrid-4DVAR system to use only the climatological error covariance model so that it was just a pure 4DVAR formulation. One reason we chose this experiment configuration was that our prior work showed that the ability of the LETLM to propagate analysis corrections depends on the weighting given to the ensemble and climatological parts of the hybrid background error covariance model. The LETLM was found to perform at its worst (best) when full (zero) weight is given to the climatological part of the hybrid matrix (Frolov and Bishop 2016; Allen et al. 2017). Hence, to make our experiments more stringent, we chose to address the harder problem of propagating a climatological perturbation. Another reason was that degrading the system to pure 4DVAR removed any aliasing between the analysis increment and the ensemble. To ensure consistency among the model first guess, the ensemble resolution, the background trajectory for the TLM, and the 4DVAR inner loop, we used the same T47L60 (about 2.5° at the equator, or 277 km) resolution for all components of the cycling system. In our experiments, we assimilated the full suite of

FIG. 2. Schematic of the experimental period. (a) Spin up and calibration, 15–20 Nov 2014; (b) archive of ensemble perturbations, 20 Nov 2014 to 2 Mar 2015; and (c) LETLM testing, 4–15 Mar 2015.

conventional observations and Advanced Microwave Sounding Unit (AMSU)-A microwave radiances.

The 4DVAR system was initialized from a fully spun up, high-resolution (T425L60) model state that was degraded to T47L60. We cycled the system for 120 days, starting at 0000 UTC 15 November 2014. The cycling run generated a set of 480 analysis increments, 400 of which were used as perturbations to initialize the ensemble (analysis increments between 21 November 2014 and 2 March 2015). Most of the experiments were performed using a reference configuration that was initialized from the same initial condition at 0600 UTC 20 November 2014. This date was chosen because it allowed for most of the initial shocks from the resolution downgrade to fully dissipate, and it allowed the cycling system to settle into a new steady state. In addition to the reference configuration that was tuned for the cycle at 0600 UTC 20 November 2014, we performed a second tuning based at 0000 UTC 4 March 2015. We then tested the LETLM performance using the average of these two tunings on seven additional experiments from a randomly selected set of times over the period 5–15 March 2015. Figure 2 summarizes our experimental configuration above and highlights the fact that the analysis increments used to generate ensemble perturbations came from a period that was distinct from the period used for testing of the LETLM performance.

b. Ensemble generation

Computation of the LETLM relies on the availability of ensemble forecasts. These ensemble forecasts can come from a cycling ensemble forecast system (e.g., an ensemble Kalman filter or an ensemble of data assimilations). In our case, we did not have access to such high-quality ensemble forecasts; instead, we initialized the ensemble forecasts by drawing initial ensemble perturbations from a database of 4DVAR increments. We used 400 ensemble

perturbations from the time period 21 November 2014 to 2 March 2015 (Fig. 2). For sensitivity tests with smaller ensemble sizes, we randomly selected a subset of increments over the same time period. The increments were added to the initial background state, and then 6-h forecasts were completed for each ensemble member.

During early stages of our experimentation, we also tried to use the flow-dependent ensemble that was cycled in NAVGEM using the ET method described in McLay et al. (2008) and Kuhl et al. (2013). However, we found that the performance of the LETLM with ET perturbations was significantly worse. The root-normalized-mean-square error [RNMSE in Eq. (7)] with the analysis increment ensemble was 0.313, compared to 0.505 with the ET ensemble—a 65% degradation. Our preliminary analysis suggested that this degradation was because the ET ensemble did not span

the space of possible climatological increments within each influence volume. The results of the experiments with the ET-based ensemble are not shown in this paper. We concede that the optimal generation of the ensemble for the LETLM is still an open question, and we defer it to future investigations.

c. Error metrics

To quantify the skill of the TLM and LETLM, we compared the 6-h forecasts of the zonal wind U , meridional wind V , and temperature T fields with the truth forecast. By the truth forecast of the perturbation, we mean the difference between two nonlinear forecasts whose initial conditions differ by the known (TRUTH) initial perturbation. To illustrate this calculation in the case of the U perturbations, we calculate area-weighted (to account for the Gaussian grid) root-mean-square error (RMSE) between the LETLM and TRUTH as follows:

$$U_{\text{RMSE}}^2(k) = \frac{\sum_{j=1}^{n^{\text{lat}}} \left\{ \sum_{i=1}^{n^{\text{lon}}} [U_{\text{TRUTH}}(i, j, k) - U_{\text{LETLM}}(i, j, k)]^2 \cos[\text{lat}(j)] \right\}}{n^{\text{lon}} \sum_{j=1}^{n^{\text{lat}}} \cos[\text{lat}(j)]}, \quad (5)$$

where i , j , and k are indices for longitude, latitude, and model level, respectively. The TRUTH perturbation is computed as the difference between nonlinear NAVGEM forecasts initialized with and without the added perturbation. The area-weighted RMSE is calculated individually for each variable and each model level: $U_{\text{RMSE}}(k)$, $V_{\text{RMSE}}(k)$, and $T_{\text{RMSE}}(k)$. Profiles of these errors will be evaluated in sections 4 and 5. We then calculated the RNMSE over all variables as follows:

$$\varepsilon(k) = \sqrt{\frac{U_{\text{RMSE}}^2(k) + V_{\text{RMSE}}^2(k) + T_{\text{RMSE}}^2(k)}{U_{\text{PERT}}^2(k) + V_{\text{PERT}}^2(k) + T_{\text{PERT}}^2(k)}}, \quad (6)$$

where the variables with the subscript PERT are the RMS variations of the perturbation that is being propagated, calculated separately at each level. The quantity ε is then averaged over all vertical levels to get a single error metric:

$$\bar{\varepsilon} = \frac{1}{n^{\text{lev}}} \sum_{k=1}^{n^{\text{lev}}} \varepsilon(k). \quad (7)$$

In addition, we calculate the correlation of the true nonlinear perturbation with the TLM and LETLM representations of that perturbation. This is done by calculating the correlation separately at each

model level $U_{\text{CORR}}(k)$, $V_{\text{CORR}}(k)$, and $T_{\text{CORR}}(k)$ as follows:

$$U_{\text{corr}}(k) = \frac{\sum_{j=1}^{n^{\text{lat}}} \text{corr}[U_{\text{LETLM}}(:, j, k), U_{\text{TRUTH}}(:, j, k)] \cos[\text{lat}(j)]}{\sum_{j=1}^{n^{\text{lat}}} \cos[\text{lat}(j)]}, \quad (8)$$

where $\text{corr}(\cdot)$ is the linear Pearson's correlation coefficient. Then, we average over all variables:

$$r_{\text{CORR}}(k) = [U_{\text{CORR}}(k) + V_{\text{CORR}}(k) + T_{\text{CORR}}(k)]/3. \quad (9)$$

A single error metric is obtained by averaging over all levels:

$$\overline{r_{\text{CORR}}} = \frac{1}{n^{\text{lev}}} \sum_{k=1}^{n^{\text{lev}}} r_{\text{CORR}}(k). \quad (10)$$

For some of the comparisons, we also calculated a relative error, where as a reference we either use the reference configuration described in section 4 or the NOGAPS TLM. The relative error between an experiment and a reference is calculated by

$$\varepsilon_{\text{relative}} = \frac{\varepsilon_{\text{experiment}} - \varepsilon_{\text{reference}}}{\varepsilon_{\text{reference}}} \times 100\%. \quad (11)$$

d. Nonlinear index

To measure the nonlinearity of the forecast, we used the nonlinearity index (NLI) similar to Allen et al. (2017). Using U as an example, the nonlinearity index is defined as

$$\text{NLI}_U = \frac{\text{std}[(U_2 + U_3)/2 - U_1]}{\text{std}(U_1)}, \quad (12)$$

where $\text{std}(\cdot)$ is a standard deviation operator, U_1 is the forecast of the U field from unperturbed forecast, U_2 is the forecast with added perturbation to the initial condition, and U_3 is the forecast with perturbation subtracted from the initial condition. For a linear system, the NLI is equal to zero. An NLI of 1 means that the nonlinear terms are equal to the standard deviation of the variable of interest. We computed the NLI for the entire system as a simple average of the NLI for the U , V , and T .

e. Tuning methodology

To illustrate the LETLM approach to parameter optimization, we consider two “tuning” parameters that are available in our LETLM: 1) the horizontal length scale L of the localization window that determines the horizontal extent of the influence volume (red dots in Fig. 1) and 2) the pseudoinverse cutoff parameter β (see appendix for details) that determines the number of eigenmodes that are actively engaged in the pseudoinverse. In principle, different values of these parameters could be chosen for each model grid point and each time in the forecast. In practice, however, we assume these parameters are constant in time and depend only on the model level: $L = L(k)$ and $\beta = \beta(k)$. Other model parameters are kept fixed for the LETLM integration, including the size of the vertical influence volume (z_{halo}), the ensemble size (n^{ens}), the model time step, and the forecast time. Sensitivities to several of these parameters are examined in section 5.

To find the optimal values of L and β , we conduct a series of 152 offline calibration forecasts. Each forecast selects one value of L from 250 to 2000 km in steps of 250 km and one value of β from 0.1 to 0.9 in 0.1 increments and from 1.0 to 10.0 in 1.0 increments, resulting in $8 \times 19 = 152$ LETLM tuning forecasts over the same 6-h tuning period. The values of the L and β for each of the forecasts do not change in the vertical and are fixed with time. To select an optimal profile of $L(k)$ and $\beta(k)$, we compare the 6-h forecast error for U , V , and T against the known truth for the perturbation

forecast computed using a pair of nonlinear models. For each level, we select the values of the $L_{\text{opt}}(k)$ and $\beta_{\text{opt}}(k)$ that result in a smallest error for this level $\varepsilon_{\text{opt}}(k)$.

Clearly, it would be highly impractical to conduct such extensive tuning every time one would want to use an LETLM. Instead, we assume tuned profiles of $L_{\text{opt}}(k)$ and $\beta_{\text{opt}}(k)$ for one time can be applied to other times as well. However, before using the $L_{\text{opt}}(k)$ and $\beta_{\text{opt}}(k)$, we verified that $\varepsilon_{\text{opt}}(k)$ obtained through offline tuning resulted in similar errors when the vertically varying profile of $L_{\text{opt}}(k)$ and $\beta_{\text{opt}}(k)$ are used in a single LETLM forecast. We found that a single LETLM forecast with the optimal parameters was as accurate as the $\varepsilon_{\text{opt}}(k)$ obtained from the offline tuning using 152 LETLM forecasts. The profiles of U_{RMSE} , V_{RMSE} , and T_{RMSE} were nearly indistinguishable, with the maximum absolute differences of 0.014 m s^{-1} for the wind component and 0.011 K for temperature, and the $\bar{\varepsilon}_{\text{LETLM}}$ is similar (0.303 for offline vs 0.300 for online).

4. LETLM reference configuration

a. Reference configuration

To illustrate the performance of the LETLM, we assembled a reference configuration of the LETLM that provides a good balance between the accuracy and the computational efficiency. For the vertical influence volume, we chose $z_{\text{halo}} = 4$; for the ensemble size, we used the maximum number of analysis increments in our archive (400); the time step of the LETLM was fixed at 1 h, twice the time step of the nonlinear model; and the forecast time was 6 h. We chose the 1-h time step and 6-h period because the NAVDAS-AR 4DVAR needs to propagate observations backward in time in hourly batches over a 6-h DA window. We also felt that the 1-h time step allowed us to use relatively small influence volume (recall that the size of the influence volume grows with the length of the time step, which increases computational costs and complicates effective localization of the LETLM). The LETLM state variables included U , V , T , P , Z , σ , and Q (option 5 in Table 2). The reference configuration used tuned profiles of horizontal halo L and the pseudoinverse truncation parameter β shown in Fig. 3 (see detailed discussion of the tuning methodology in section 3e and tuning results in section 4b). A summary of parameters used for the reference configuration is presented in Table 1. All of the results in this section are presented for the propagation of a single 4DVAR increment, initialized at 0600 UTC 20 November 2014.

b. Tuned profiles of optimal length and pseudoinverse cutoff

Figure 3 shows how the optimally tuned values of L and β varied vertically. L was, on average, around

FIG. 3. Optimally tuned profiles of (a),(c) horizontal halo region L and (b),(d) pseudoinverse cutoff β used in the reference configuration. Profiles are shown on pressure levels in (a),(b) and on the bottom 30 model levels (up to approximately 220 hPa) in (c),(d). Horizontal line in (a) indicates the location of the model level 30 (around 220 hPa). Vertical line in (b),(d) shows the default cutoff value of 1.

750 km within the troposphere (Fig. 3c); however, it increased steadily in the stratosphere to 1500–1750 km above 1.0 hPa (Fig. 3a). The Courant–Friedrichs–Lewy (CFL) analysis of such large influence volumes suggests a maximum propagation speed of $\sim 200 \text{ m s}^{-1}$ in the troposphere and $\sim 400 \text{ m s}^{-1}$ in the lower mesosphere. We were surprised to obtain such large propagation speeds, which exceed the physical wind speeds in the atmosphere. In the stratosphere, they even exceed the speed of propagating gravity waves or the speed of sound (343 m s^{-1}). In other words, in the stratosphere,

the optimal size of the horizontal influence volume exceeded the length at which information can propagate in the volume. We speculate that such large empirically tuned horizontal halos resulted from the limited size of our vertical halo. The NAVGEM updates the entire vertical column at once. Hence, it is theoretically possible for variations at the bottom of the atmosphere to cause variations at the top of the atmosphere over a single time step. Our reference configuration for the LETLM can only accommodate information propagation over nine vertical levels. We speculate that the

TABLE 1. Parameters for reference LETLM configuration and ranges of tuning values.

Parameter	Reference configuration value	Values used for sensitivity tests
Date	0600 UTC 20 Nov 2014	0600 UTC 20 Nov 2014
Time step	1 h	1 h
Forecast time	6 h	6 h
z_{halo}	4 (i.e., nine vertical levels)	0, 2, 4, 6, 8
Variables	U, V, T, P, Z, σ, Q	See Table 2
Tuning range for β	0.1, 0.2, ..., 0.9, 1, 2, ..., 10	0.1, 0.2, ..., 0.9, 1, 2, ..., 10
Ensemble size	400	40, 80, ..., 400
Tuning range for L	250, 500, 750, ..., 2000 km	250, 500, ..., 2000 km
Ensemble initialization	Archived analysis increments	Archived analysis increments

LETLM compensates for the lack of the vertical information by enlarging the horizontal halo region to include information correlated to the missing vertical information.

The term $\beta_{\text{opt}}(k)$ also varied with height and differed from the suggested value of 1 (e.g., as used by the MATLAB's *pinv* function), and it was largest in the boundary layer, reaching 2–3 near the surface, but declined with height, with typical values of 0.3–0.6 in the troposphere and 0.1–0.2 in the stratosphere. Slightly higher values of 0.4–0.5 occur near the model top. The higher values of β_{opt} in the boundary layer result in a more severe truncation of the ensemble eigenspectrum and indicate that the LETLM is unable to accurately propagate aspects of the boundary layer perturbations associated with moderately small eigenvalues. The small values of β_{opt} (e.g., in the stratosphere) indicate that the LETLM is able to accurately propagate aspects of the stratospheric perturbations associated with moderately small eigenvalues, and, hence, the tuning retains a higher number of the eigenmodes in the pseudoinverse.

c. Performance of the reference LETLM

We found that both the LETLM and the TLM show considerable skill, compared to persistence, in predicting evolution of the 4DVAR increment over the 6-h assimilation window. The growth of the vertically averaged errors (Fig. 4) shows the LETLM errors were consistently lower than the NOGAPS TLM errors over the entire forecast, except at $t = 0$ h. We note that while the initial TLM errors are exactly zero, the initial LETLM errors are small nonzero values. This is because the initial TLM field was initialized with the “true” perturbation, while the LETLM perturbation was projected into a subspace of a localized ensemble. Since the projection was not exact, there are slight differences relative to the truth at $t = 0$ h. Overall, the normalized errors favor the LETLM. At $t = 6$ h, the LETLM error was 0.303, while the NOGAPS TLM error was 0.340, an improvement of 11%.

When we examine the globally averaged profiles of errors valid at $t = 6$ h (Figs. 5, 6), we find that the LETLM errors were roughly the same as the TLM errors in the boundary layer, but from ~ 970 to 100 hPa, the LETLM errors were smaller for all variables. (Note that the reference pressure labels on Figs. 5 and 6 were calculated using the NAVGEM 60-level hybrid sigma-pressure coefficients and assuming a surface pressure of 1000 hPa; all error calculations are actually performed on the model levels, not constant pressure levels.) From 100 to ~ 0.5 hPa, the TLM errors are generally smaller. When we compare in Fig. 7 the relative performance of the LETLM and TLM to the value of the NLI, we find

FIG. 4. Vertically integrated RNMSEs in propagation of the 4DVAR correction over 6 h [Eq. (7)]. Plotted are the magnitude of the propagated perturbation at forecast time τ (blue); errors under the assumption of a trivial TLM = I, equivalent to the 3DVAR case (red); errors for propagation of the perturbation using the LETLM (green); and errors for propagation of the perturbation using the NOGAPS TLM (black).

that LETLM performed better in the areas (below 100 hPa) where the nonlinearity was high ($\text{NLI} > 1.5$), and the TLM performed better in the regions (above 100 hPa) where the model was more linear ($\text{NLI} < 1.5$). In the boundary layer where the model was dominated by nonlinearity ($\text{NLI} > 4$), both the TLM and the LETLM had similar poor performance. This behavior is consistent with the shallow water model results of Allen et al. (2017), which showed that the LETLM performed better relative to the TLM as the nonlinearity of the analysis increment increased [see Fig. 16b of Allen et al. (2017)].

The correlation profiles $r_{\text{CORR}}(k)$ for the reference configuration (Fig. 8) show that both the LETLM and TLM have large correlations (greater than 0.9), except in the lowest 10 model levels. The LETLM generally had higher correlation than the NOGAPS TLM in the troposphere, while the TLM has higher correlation in most of the stratosphere. The vertically integrated mean value $\bar{r}_{\text{CORR}}$ was 0.975 for the LETLM, compared to 0.960 for the TLM.

We speculate that the comparatively poor performance of the NOGAPS TLM could be, in part, attributed to it being somewhat overdifusive. For example, Fig. 9 shows that the globally averaged magnitudes of the perturbations in the NOGAPS TLM are lower than the magnitude of the true perturbation after 6 h. In contrast, the LETLM-propagated perturbations retain the correct magnitude almost everywhere in the profile, except in the boundary layer, where both the NOGAPS TLM and the LETLM significantly underpredict the magnitude of the perturbation. We further found support for our theory that NOGAPS TLM is too smooth,

FIG. 5. Globally averaged profile of errors for (a) Eq. (6) and (b)–(d) Eq. (5) after 6 h.

compared to the Truth or to the LETLM, when we examined the maps of the perturbations in the middle of the troposphere (~ 700 hPa) and stratosphere (~ 10 hPa). We found that the LETLM was able to capture the

finescale structure in the U and V fields better than the NOGAPS TLM, in which the features appear to be somewhat smoothed out. For brevity, we present the extensive graphical comparisons of the maps for the

FIG. 6. As in Fig. 5, but for the lowest 30 levels of the model.

FIG. 7. NLI [computed using Eq. (12)].

Truth, the NOGAPS TLM, and the LETLM in the online supplemental material (Figs. S1 and S2 for 700 hPa and Figs. S3 and S4 for 10 hPa).

One area where both TLMs had relatively poor performance was the prediction of temperature in the boundary layer. The maps of the perturbation at level 59 (close to the bottom of the boundary layer) show that the true perturbation develops small-scale features in a 6-h forecast, such as the large temperature perturbations over Africa (row two of Fig. 10). These features are absent in the original perturbation at 0 h (top row of Fig. 10), as well as in the 6-h forecast with either the NOGAPS TLM (third row from the top) or the LETLM (bottom row). In fact, the TLM- and LETLM-propagated fields look like smooth, evolved versions of the 0-h field. We speculate that these unresolved, intense features are likely due to the nonlinearity of the convection and the planetary boundary layer schemes in the NAVGEM. Recall that these high-intensity features are present in the original ensemble forecast, but our tuning experiments (Fig. 3) suggest that we need to increase the pseudoinverse cutoff parameter in the boundary layer in order to filter out some features of the ensemble forecast that cannot be predicted by the linear forecast operator.

Finally, we examined the temporal evolution of the U perturbation along a vertical cross section taken along the 267.5° longitude (Fig. 11). The cross section shows a variety of scales captured by both the TLM and the

LETLM, including more rapid evolution of the field at the upper levels. This may be related to resolved unbalanced modes (gravity waves) that are present in the system. These fast-moving features require larger length scales, as seen in tuning of the reference configuration. The LETLM and the TLM both capture the temporal variability in the cross sections well.

Overall, we conclude that the reference configuration of the LETLM with offline tuning and $z_{\text{halo}} = 4$ competes well with the TLM in terms of the global error profiles, horizontal maps, and cross sections. In the next section, we further examine the sensitivity of the LETLM to various parameters in order to provide more insight into LETLM behavior and further guidance in tuning the LETLM for particular user needs.

5. Sensitivity tests for the LETLM

a. Sensitivity to horizontal length

In the reference configuration examined in section 4, we simultaneously tuned L and β in order to obtain optimal results. Here, we examine the sensitivity to L in more detail by running through a range of L values from 250 to 2000 km in 250-km increments. For each test, we determine the optimal β as a function of level for fixed values of L . The resulting error profiles, relative to the reference configuration errors (shown in Figs. 5 and 6), are provided in Figs. 12a and 13a. (Recall that the reference

FIG. 8. Average correlation of true and propagated perturbations after 6 h. (a) Performance from the top of the atmosphere to the surface with logarithm of pressure used as a vertical coordinate. (b) Troposphere (lower 30 model levels) with model levels used as vertical coordinates.

values for L and β vary with height in the manner given by Fig. 3). The results indicate that $L = 500$ and 750 km are outliers in the upper levels, suggesting that at least 1000 km is necessary to accurately model this region. As expected from the reference configuration results, errors at higher layers improve as length increases, while in the lower layers, errors increase as the length increases. The vertically averaged errors in Fig. 14a show a minimum RNMSE at 1000 km. A comparison of the reference configuration (RNMSE = 0.303) and the results with constant $L = 1000$ km (RNMSE = 0.317) shows that using the simpler fixed-length approach results in only a 5.6% increase in the total error. Therefore, tuning of L might not be that critical, as long as it is above a given threshold value. However, we note below that the optimal L value does depend on the size of the vertical halo z_{halo} , so one needs to take care in the approach to tuning.

b. Sensitivity to vertical halo

To examine the sensitivity to vertical localization, we varied the number of adjacent levels (above and below) that are allowed to influence the central level. We ran with z_{halo} values of $0, 2, 4, 6$, and 8 , corresponding to a vertical localization of no larger than $1, 5, 9, 13$, and 17 levels. In each case, we fixed β at 1.0 , but varied the length from 250 to 2000 km, picking the optimal value of length as a function of vertical level. Errors measured

relative to the reference configuration errors are provided in Figs. 12b and 13b. The value of $z_{\text{halo}} = 0$ is an extremely poor outlier, indicating that some vertical information is needed. For nonzero values, these results show smaller sensitivity to z_{halo} in the troposphere and lower stratosphere, while in the upper stratosphere and mesosphere, the sensitivity is quite strong (note that the errors for large z_{halo} are smaller than the reference configuration errors above ~ 1.0 hPa). This may be related to wider level spacing at higher levels. However, it is also likely that fast-moving vertically propagating resolved gravity waves in the stratosphere and mesosphere require larger vertical layers to capture these features. The optimal lengths (not shown) have generally decreasing values with increasing z_{halo} . This inverse relationship somewhat constrains the influence volume size, with stretching in the vertical corresponding to shrinking in the horizontal. For example, for the level of 1 hPa, the optimal L was 1750 km for z_{halo} of 2 and 1250 km for z_{halo} of 8 . This may be expected, since if one fails to capture all the sensitive variables in the vertical, then the LETLM can compensate by using additional variables that are correlated with the excluded variables (e.g., by increasing L). The vertically integrated errors for z_{halo} (Fig. 12b) show a minimum at $z_{\text{halo}} = 6$. However, the cost increases with increased z_{halo} , so there are tradeoffs to be made. For example, although the cost of the LETLM with $z_{\text{halo}} = 2$ was less than that

FIG. 9. Average magnitude of the true perturbation (dotted line), the NOGAPS TLM (black line), and the LETLM (green line). (a)–(c) Performance from the top of the atmosphere to the surface with logarithm of pressure used as a vertical coordinate. (d)–(f) Troposphere (lower 30 model levels) with model levels used as vertical coordinates.

with $z_{\text{halo}} = 4$, $z_{\text{halo}} = 4$ enabled the LETLM to achieve significant overall superiority to the TLM. Presumably, further superiority could have been obtained by letting $z_{\text{halo}} = 6$, but that would have increased the cost of the LETLM. In the future, we could vary the size of the z_{halo} vertically to achieve a better compromise between the cost and the accuracy of the LETLM.

c. Sensitivity to pseudoinverse cutoff

We next examine the sensitivity to the pseudoinverse cutoff β . We ran through a series of LETLM tests in which β was fixed at values of 0.1, 0.2, 0.5, 1.0, 2.0, and

5.0. For each β value, the length was tuned (range 250–2000 km) at each level to give the lowest errors. (The results are provided in Figs. 12c and 13c). Running with fixed β provides reasonable first-order errors throughout the vertical profile. As expected from the reference configuration, lower (higher) values of β provide better results at upper (lower) levels. The vertically averaged errors (Fig. 14c) show an overall minimum at $\beta = 0.5$, with $\bar{\epsilon}_{\text{LETLM}} = 0.311$, a 2.6% increase above the reference configuration value. We conclude that while simultaneous tuning of L and β is optimal, relatively good results can be obtained with more

Level 59 (~surface)

FIG. 10. Horizontal maps of the U , V , and T perturbations plotted at level 59 (almost at the surface) at (top) the initial time and (bottom three rows) 6 h (Truth is in row 2, TLM is in row 3, and LETLM is in row 4). Perturbations are shown for the reference configuration of the system.

FIG. 11. Hourly cross sections of the (left) Truth, (middle) TLM-propagated, and (right) LETLM-propagated perturbations. Cross sections of perturbations are shown for the reference configuration of the system and are taken along the 267.5°E longitude.

FIG. 12. Sensitivity to LETLM parameters with respect to the reference configuration [Eq. (11)]; positive values of error show degradation, and negative values show improvement over the reference configuration. Plotted are globally averaged vertical RNMSE profiles for (a) localization scale L , (b) vertical localization scale z_{halo} , (c) pseudocutoff parameter β , and (d) the ensemble size n^{ens} .

coarse-grained tuning. In the following sensitivity tests, in which we are not primarily concerned with optimal accuracy, but rather sensitivity to various parameters, we opted to fix β at 1.0 and tune L as a function of level over the range 250–2000 km.

d. Sensitivity to ensemble size

The LETLM was run with ensemble sizes of 40–400 in steps of 40 ensemble members. Figures 12d, 13d, and 14d show that the skill of the LETLM steadily increases as more ensemble members are added. It also appears that larger ensembles will yield even higher skill, as the RMS errors have not saturated yet at 400 ensemble members—the maximum number of ensembles that we tried (see Fig. 14d). We also found that addition of ensemble members was more effective at reducing errors above the troposphere (cf. the slopes of green and blue lines that show the RMSE below and above 100 hPa). The results in Fig. 14d suggest that even higher LETLM performance can be achieved either by using a larger ensemble (which is expensive) or by possibly using a higher-quality ensemble that can capture more information using fewer ensemble members.

In our prior work, Frolov and Bishop (2016) showed that the fidelity of the LETLM is improving steadily as

the ensemble size is increased. Bishop et al. (2017) provided a mathematical proof that the LETLM is guaranteed to be precisely accurate when the rank of the ensemble within the influence volume is equal to the number of variables within the influence volume and when the ensemble perturbations are small enough to be governed by linear dynamics. As a crude measure of the degrees of freedom in the stencil, we compared the number of grid points in the influence volume with the number of ensemble members (Fig. 15). We find that the two numbers were comparable in the troposphere (with the exception of the polar regions, where the influence volume was as large as 2000 grid points). However, in the upper stratosphere and mesosphere, the number of points in the influence volume was significantly larger than the number of ensemble members (~ 800 –2000 outside the polar regions).

e. Sensitivity to the choice of state variables

The LETLM can be used to propagate any combination of state variables available in the nonlinear NAVGEM. We tested the LETLM with $z_{\text{halo}} = 4$ and $n^{\text{ens}} = 400$ for a number of different configurations of state variables (U , V , T , P , Z , σ , and q). Table 2 lists the configurations and

FIG. 13. As in Fig. 12, but for the bottom 30 model layers.

$\bar{\epsilon}_{\text{LETLM}}$ values, while Fig. 16 shows the error profiles relative to the reference configuration errors. Using U , V , and T (option 1) as state variables resulted in poor performance in the upper layers (errors more than 50% larger than the reference configuration near the stratopause), and $\bar{\epsilon}_{\text{LETLM}} = 0.393$. Including the 3D pressure field (U , V , T , and P in option 2) improves the $\bar{\epsilon}_{\text{LETLM}}$ to 0.363 and benefits the profile from the surface up to around level 37 (~ 200 hPa). Since the vertical coordinate is constant pressure above about 85 hPa, we do not expect the inclusion of the pressure field to have much influence in the higher levels. A large jump in skill occurs in option 3 (U , V , T , P , and Z), when geopotential height is added to the state. Parameter $\bar{\epsilon}_{\text{LETLM}}$ drops to 0.317, and the errors decrease, particularly above 200 hPa. Adding σ (option 4) provides an additional improvement in the troposphere, but slightly increases the errors in the stratosphere, with an overall $\bar{\epsilon}_{\text{LETLM}}$ increasing slightly to 0.318. Including moisture (option 5) has a slight positive overall impact, reducing $\bar{\epsilon}_{\text{LETLM}}$ to 0.313. We therefore decided to use option 5 in the reference configuration and for all subsequent tests.

f. Sensitivity to physics in the LETLM

One of the main difficulties of developing traditional TLMs is that the physics package of the atmospheric model (responsible for processes like mixing, radiative transfer, and parameterization of moist processes) is

usually rapidly evolving, highly nonlinear, and often lacks a formal derivative because of conditional statements in the code that cannot be differentiated. In contrast, the LETLM has direct access to the entire physics package of the nonlinear model. Therefore, any changes to the nonlinear model will automatically be included in the LETLM. As with conventional TLMs, however, it is necessary to test the impact of each change to the nonlinear model on the performance of the TLM (Janisková and Lopez 2012). While a complete examination of the sensitivity of the LETLM to various physics packages is beyond the scope of the current study, we briefly show the impact of turning the physics and radiation packages off (i.e., running a dry adiabatic core).

To test the impact of the overall physics package on the LETLM performance, we formed a new ensemble in which the same analysis increments were used to initialize the ensemble, but the physics terms were turned off in the model, resulting in a dry adiabatic core. This no-physics LETLM was then run with the reference configuration settings and compared to the reference LETLM with physics. Tuning for both the no-physics and physics runs was made using fixed $\beta = 1.0$ and lengths from 250 to 2000 km. Figure 17 shows the relative error reduction when turning on the physics [similar format to Fig. 5 of Janisková and Lopez (2012)]. For our version of the relative improvement [Eqs. (7) and (9) of Janisková and Lopez 2012],

FIG. 14. Sensitivity of the globally averaged RNMSE to tuning parameters (a) localization scale L , (b) vertical localization scale z_{halo} , (c) pseudocutoff parameter β , and (d) the ensemble size n^{ens} . Horizontal dotted red line shows the RNMSE score for the NOGAPS TLM, and horizontal dotted black line shows the RNSME score of the reference configuration. The optimal value is indicated by a vertical tick mark.

we first calculated the mean absolute difference (MAD) between the LETLM and the TRUTH for the zonal wind U (similar calculations were made for meridional wind, temperature, and specific humidity):

$$U_{\text{MAD}}(k) = \frac{\sum_{j=1}^{n_{\text{lat}}} \left\{ \frac{\cos[\text{lat}(j)]}{n_{\text{lon}}} \sum_{i=1}^{n_{\text{lon}}} |U_{\text{LETLM}}(i, j, k) - U_{\text{TRUTH}}(i, j, k)| \right\}}{\sum_{j=1}^{n_{\text{lat}}} \cos[\text{lat}(j)]} \quad (13)$$

The relative error was calculated as

$$\frac{U_{\text{MAD,physics}} - U_{\text{MAD,nophysics}}}{U_{\text{MAD,nophysics}}} \times 100\%. \quad (14)$$

Figure 17 shows that surface winds were improved by as much as 50% by the inclusion of the physics, while the tropospheric improvement in winds was around 10%–20%. A minimum in error reduction was in the lower stratosphere, while larger improvements occur near the stratopause (1.0 hPa). The $\bar{\epsilon}_{\text{LETLM}}$ value without physics

was 0.385, $\sim 23\%$ higher than the value obtained with physics and with the same tuning (0.313). The largest impact of physics occurs in the upper stratosphere/lower mesosphere and throughout the free troposphere. These results are broadly consistent with Janisková and Lopez (2012), who examined error reductions in the European Centre for Medium-Range Weather Forecasts (ECMWF) TLM when including various physics packages. Their Fig. 5 shows a $\sim 10\%$ decrease [calculated using Eq. (11)] in temperature and wind errors in the free atmosphere when including six physics parameterizations in the TLM. However, unlike our result, their Fig. 5 also shows a much larger benefit in the boundary layer from inclusion of the regularized physics (they had a 90% improvement, while we only have a 20%–50% improvement). Further diagnosis of the effects of each physics parameterization on LETLM performance is planned for future work.

6. Performance of the tuned LETLM for other increments

While the LETLM with full tuning has been shown to provide skill throughout the atmosphere, in an operational setting, it would be helpful to have pretuned profiles of L and β . The exact method of tuning and

FIG. 15. Size of the influence volume (number of horizontal and vertical grid points) computed for the average tuning profile used in the verification tests in section 6.

frequency of updating these profiles will need to be tested in detail. Here, we test a simple method in which full tuning is performed on two different dates at the beginning (0600 UTC 20 November 2014) and end (0000 UTC 4 March 2015) of the date range used for the ensemble initialization (Fig. 2). The optimal L and β profiles are slightly different for the two cases, but show broadly similar results. We then averaged the L and β profiles from these two offline tuning tests and used these to run online tests for seven other times picked randomly from 5 to 15 March 2015.

The resulting error profiles [relative to the corresponding TLM forecast errors, using Eq. (11)] are provided in Fig. 18, with colors corresponding to the times (in YYYYMMDD format) indicated in the top panel. The LETLM errors are consistently lower than the TLM errors over approximately the same vertical range as the reference profile (Figs. 5, 6). We computed the statistical significance of the mean profiles using the Student's t test and found that the LETLM was significantly better (95% confidence level) at the surface, over a broad layer from model level 55 (~ 970 hPa) to level 30 (~ 200 hPa). The TLM errors were smaller than the LETLM errors from ~ 70 hPa to the model top. While the LETLM does not beat the TLM at these levels, the skill of the LETLM relative to persistence for these tests is significant and comparable to that of the reference configuration in Fig. 5a. Further tuning tests are required to understand how the LETLM responds in different seasons, but these preliminary results show promise that the LETLM will perform well, even with coarse tuning.

7. Computational efficiency

a. Profiling of the LETLM computations

Our profiling of the LETLM computations (Table 3) shows that computing the pseudoinverse of the ensemble

TABLE 2. Variable options and globally averaged RMSE [calculated using Eq. (7)].

Option	Variables	Error
1	U, V, T	0.393
2	U, V, T, P	0.363
3	U, V, T, P, Z	0.317
4	$U, V, T, P, Z, \dot{\sigma}$	0.318
5	$U, V, T, P, Z, \dot{\sigma}, Q$	0.313

covariance matrix $[\mathbf{C}]^{-1} = [(\mathbf{S}_p^{\text{vol}} \mathbf{X}_m)^T (\mathbf{S}_p^{\text{vol}} \mathbf{X}_m)]^{-1}$ dominated the cost of the LETLM computation (about 86% of the total cost). The cost of the inverse was almost equally split between the matrix–matrix multiplication and the Cholesky factorization. We experimented with multiple implementations of the inverse computations (see appendix for details), and we found that the Cholesky decomposition provided the fastest results. In our experience, it is possible to further reduce the memory copy time, however, at the expense of the code complexity. We discuss alternative options for decreasing computational times even further in the conclusions.

b. Parallel scalability

We conducted two types of scalability tests. In the weak scalability test, we fixed the computation hardware but increased the number of computations in the LETLM by varying LETLM tuning parameters. In the strong scalability test, we fixed the size of the problem and used progressively larger numbers of computational processing units.

The results of the weak scalability test are presented in Table 4, which shows expected scaling laws for the horizontal halo (roughly proportional to the square of the halo radius L) and for the vertical halo (linear scaling proportional to doubling of the work with each unit increase in the z_{halo} value). However, we achieved an unexpectedly low scaling exponent (1.5) for the increase in the size of the ensemble. Recall that the cost of the LETLM is dominated (Table 3) by the cost of the covariance computation $\mathbf{X}^T \mathbf{X}$ and the pseudoinverse of this matrix. The matrix–matrix multiplication is expected to scale as a square of the ensemble size and the matrix inverse as a cube. We hypothesize that our better-than-expected scalability with respect to the ensemble size was achieved because our linear algebra package was better able to optimize matrix computations as the size of the matrix increased. Also, we expect that the Cholesky decomposition of a regularized covariance should scale better than the direct matrix inverse of the original, poorly conditioned covariance. Finally, we found no dependence on β , since the number of operations is

FIG. 16. Sensitivity to the choice of the model variables. Sensitivities are computed with respect to the reference configuration using Eq. (11). The dotted black line indicates the NOGAPS TLM error relative to the reference configuration.

independent of this parameter. We can therefore fit the timing of the LETLM as follows:

$$\text{time}(L, z_{\text{halo}}, n_{\text{ens}}) = a + bL^2 z_{\text{halo}} (n_{\text{ens}})^{1.5}. \quad (15)$$

The results of the strong scalability tests showed that the LETLM was capable of close-to-perfect scalability up to 700 processing cores. The scalability at high CPU counts was limited by the startup costs of the OpenMP threads (as was evident by better scalability when more MPI tasks were used instead of OpenMP threads). It is possible to further optimize these startup costs. We also suggest that the strong scalability will improve even further once we use a higher-resolution grid because there will be more work to be done, and the startup costs would be less important.

We also found that the LETLM strong scalability was significantly better than that of the original model (the NAVGEM only scaled to 10 cores). We explain such poor scalability of the NAVGEM by the very small size of our global grid. From our experience, we know that it is possible to scale NAVGEM further on high-resolution

model simulations. We explain the positive results with LETLM strong scalability by the fact the LETLM almost completely avoids global communications.

Overall, we found that the cost of the LETLM (in terms of CPU clock) was dominated by the cost of generating ensemble forecasts (two orders of magnitude larger than the cost of a single integration of the stand-alone NOGAPS TLM or a stand-alone NAVGEM). The cost of the LETLM computation was comparable to or smaller than the cost of the ensemble forecast and depended greatly on the configuration of the LETLM [see weak scalability results in Table 4 and Eq. (15)]. We believe that it should be possible to further reduce the cost of the LETLM computation by a factor of 5–10. We discuss these strategies in the following section.

8. Summary and conclusions

This paper presents a first demonstration of the ensemble-based TLM performance in a realistic, albeit low-resolution, model of the global atmosphere. Our results show that the LETLM outperformed the NOGAPS TLM on average (by 10% over a 6-h forecast). The NOGAPS TLM is the TLM currently used by the hybrid-4DVAR system run in operations by the U.S. Navy's Fleet Numerical Meteorology and Oceanography Center. The superior performance was most prominent for winds in the troposphere (about 20% improvement). Both TLMs had similarly poor skill for forecasting temperature perturbations in the planetary boundary layer. The NOGAPS TLM had superior skill in the stratosphere and mesosphere. Comparisons of these skill scores with the measure of model nonlinearity are not inconsistent with the speculation that the NOGAPS TLM was better when the model was only slightly nonlinear [e.g., in the stratosphere ($\text{NLI} < 1.5$)]. The LETLM was better in more nonlinear regimes (NLI between 2 and 4), and both linear models had poor skill in predicting strongly nonlinear growth of small perturbations in the planetary boundary layer (NLI of 4–8).

While the superior skill of the LETLM in the troposphere is encouraging, it should be evaluated in a proper context. Our skill comparisons were against a NOGAPS TLM—a TLM that was developed for a predecessor of the current NAVGEM and has not seen any significant updates in over a decade. More recently, an effort has been undertaken to develop a more up-to-date version of the TLM for the NAVGEM. This newer version shows similar skill improvement to the LETLM when compared against the NOGAPS TLM (about 15%–20% in the troposphere). Furthermore, an experience with a state-of-the-art TLM at ECMWF (P. Lopez 2017, personal communication) suggests that a continued effort at

FIG. 17. Sensitivity of (a),(e) U ; (b),(f) V ; (c),(g) T ; and (d),(h) specific humidity to the inclusion of physics and radiation in the ensemble calculations, calculated using Eq. (14).

improving a traditional TLM leads to significant skill gains by the TLM, even in areas where both of our TLMs failed, such as the planetary boundary layer. We expect that similar gains can be achieved if a continuous effort is dedicated to the improvements in the ensemble system that drives the LETLM. Taking this larger context into account, it is fair to say that the LETLM is a promising technology that can deliver skill comparable to the skill of state-of-the-art TLMs.

Our sensitivity studies showed that the LETLM was most sensitive to the construction and the size of the ensemble. The ability of the LETLM to propagate analysis corrections was higher when the initial perturbations were obtained from an archive of 4DVAR analysis corrections than when using the initial perturbations that are typically used for ensemble forecasting (the ET initial perturbations). The size of the ensemble also mattered. The performance of the system continued improving up to the maximum size of the ensemble attempted (400 members). It is very likely that larger ensembles will result in even higher skill. We also

speculate that it should be possible to construct more effective ensembles that can yield similar skill with fewer members.

Our sensitivity results also showed that allowing for vertical tuning of the horizontal halo L and pseudoinverse cutoff β resulted in moderate improvement of the LETLM performance (5.6% and 2.6%, respectively). The impact of tuning the vertical halo (z_{halo}) was most prominent in the stratosphere and the boundary layer. We found that tuning of z_{halo} interacted with the optimal length of the horizontal halo (bigger vertical halos required smaller horizontal halos). In general, the horizontal halo increased with height (from 750 km in the troposphere to 1750 km on top of the stratosphere). We speculate that the larger horizontal halo is better able to represent the large-scale rapidly propagating waves in the upper stratosphere. Recent work (to be presented in future publications) has shown that a further reduction in the LETLM errors is possible by using a nonuniform influence volume, where the current cylindrical influence volume (of height z_{halo} and radius L) is complemented

FIG. 18. Verification tests with seven different analysis increments. (a) RNMSE results relative to the error in NOGAPS TLM [Eq. (11); negative values indicate improvement compared to the NOGAPS TLM and positive values, deterioration]. Black dots indicate levels where LETLM was significantly better (with 95% confidence using t test) than the NOGAPS TLM. Red dots indicate levels where the NOGAPS TLM was significantly better (with 95% confidence). (b) As in (a), but highlights the errors in the troposphere by using model levels as vertical coordinates. In (a) dates for each increment are shown in color labels.

by a pencil-like column that extends the influence volume directly above and below the central patch point. This new influence volume shape (we call it a “spinning top”) is especially effective at reducing errors at the model top and in the boundary layer, where information can propagate rapidly in the vertical due to convection or radiation processes.

Our results also showed that including physics in the LETLM computation yielded positive results through

most of the vertical column (improvement of 23%, compared with the LETLM based on the ensemble without parameterizations of subgrid-scale processes and radiation). These sensitivities to physics are in qualitative agreement with previous findings from the physics-enabled TLMs developed at ECMWF (Janisková and Lopez 2012). However, we suspect that the impact of physics will increase with higher model resolution. The experience at ECMWF also suggests that to achieve strong skill scores in the boundary layer, it might be necessary to develop versions of the model ensemble that have simplified physics in the boundary layer that will improve linear predictability of the system at short time scales.

The profiling of the computational costs in the LETLM showed that the overall cost was dominated by the ensemble forecast, with the cost of the LETLM computation being similar or lower than the cost of the ensemble integration. However, all operational NWP centers are already committing large resources for ensemble forecast integration. For example, the Canadian Meteorological Service routinely performs forecasts with as many as 500 ensemble members at 50-km resolution. The LETLM can leverage these existing computations without adding significantly to the cost of the overall system. In terms of the LETLM computation, the majority of the cost was associated with the matrix inverse (the size of the ensemble) at each of the model grid points. There are several strategies that can allow us to further reduce the cost of the LETLM computation, which we plan to examine in the follow-up publications. These strategies include computing the matrix inverse less frequently (e.g., at every other grid point), increasing the size of the vertical or horizontal computational patch (currently, every grid point forms a unique patch), and developing methods to construct a smaller subset of ensemble members that can be as effective as our current climatological ensemble.

However, regardless of the success of the above-mentioned optimization, the LETLM computation has a unique property that will allow it to significantly reduce the cost of computations where the TLM and its adjoint are applied iteratively (such as in 4DVAR). In fact, the LETLM matrix can be precomputed before the 4DVAR

TABLE 3. Profiling of the LETLM computation. Note: The computational times were obtained for the reference configuration of the LETLM.

Operation	Mathematical expression	Percent of the total cost
Memory copy	$\mathbf{U} = \mathbf{S}_p^{\text{vol}} \mathbf{x}_m$	13%
Matrix–matrix multiply	$\mathbf{A} = \mathbf{U}^T \mathbf{U}$	40%
Cholesky decomposition	$\sqrt{\mathbf{C}} = \sqrt{\mathbf{A} + \beta \mathbf{I}}$	46%
Matrix–vector operations (including Cholesky solver)	$(\mathbf{S}_p \mathbf{x}_{m+1}) \text{spotrf}(\mathbf{C}) \times \mathbf{U}^T (\mathbf{S}_p^{\text{vol}} \mathbf{x})$	<1%

TABLE 4. Weak scalability results.

Experiment	Range			
	Horizontal halo radius	Vertical halo length (z_{halo})	Ensemble size	Power law fit
Horizontal halo	250–3750 km	2	360	$L^{1.93}$
Vertical halo	500 km	0–9	360	$2.1 \times z_{\text{halo}}$
Ensemble size	500 km	2	40–360	$(n^{\text{ens}})^{1.46}$

executes and stored as a sparse matrix that can then be invoked by the 4DVAR solver iteratively at almost zero cost. Furthermore, the sparse matrix multiplication is supported by robust libraries and is likely to scale better than a traditional TLM computation. This possibility to drastically reduce the cost of the 4DVAR computation (or enable the efficient implementation of the perturbed observation 4DVAR system) is one of the most appealing properties of the proposed system. It is also likely that the speed-up based on the precomputed LETLM can be realized even in 4DVAR systems that are configured to run using an outer-loop configuration. In such systems, it might be beneficial to recompute the LETLM between the outer loops using a new set of ensemble forecasts that are conditioned based on the updated best guess from the previous outer loop. This speed-up may be possible because it is usually more efficient to compute ensemble forecasts in parallel rather than computing multiple sequential integrations of the TLM and adjoint models.

9. Future work

The feasibility of an NWP system based on the LETLM will depend on the ability of the LETLM to perform well in high-resolution test cases. We would like to provide a few remarks to the readers on how we think the findings of this study will generalize to high-resolution cases. As the resolution of the model increases, the number of grid points within an influence volume of a given size will scale by approximately $(\Delta x^{\text{old}}/\Delta x^{\text{new}})^2$. Having more grid points in the influence volume will not only slow down the computation, but also [based on the findings in Frolov and Bishop (2016) and Bishop et al. (2017)] will necessitate an increase in the number of ensemble members to maintain a similar level of skill. Recall that the LETLM performs at its best when the number of degrees of freedom in the influence volume is similar to the number of independent ensemble members. To maintain the same fidelity of the LETLM at a higher resolution, without needing to increase the number of ensemble members, we suggest that the LETLM will need to operate with shorter time steps and smaller influence volumes at higher resolution.

For example, a doubling in the resolution might translate to halving of the time step and the halo size L . This is similar to how traditional finite difference schemes operate. Shorter time steps will limit the distance at which the information can propagate in one time step, hence, allowing for use of smaller halo L .

Another important aspect for the success of the LETLM in future operational-scale NWP is the ability to construct skillful ensembles that can support accurate LETLM computations. In this paper, we used a very crude ensemble that was initialized using an archive of historic analysis increments. We suspect that better and more effective ensembles can be constructed. Our results also shed light on some of the concerns raised by anonymous reviewers. One of the concerns was that an ensemble of forecasts needs to span the space of an arbitrary analysis correction. Clearly, it is impossible for any limited-size ensemble to exactly include an arbitrary analysis correction in a high-dimensional NWP model. However, our experience with the crude ensemble generation scheme used here and the EnKF-based scheme used in Allen et al. (2017) suggests that the LETLM localization method is effective at extracting the local dynamics of the system. We suspect that to construct an effective LETLM, it is sufficient if the ensemble perturbations are effective at exciting dominant dynamical modes for each of the localization volumes.

Finally, more work might be needed to understand the ability of LETLM to deal with unbalanced modes in either the ensemble forecast or in the increment that it propagates. In our earlier work (Allen et al. 2017), we showed that the normal mode initialization of the ensemble forecasts improves the skill of the LETLM, possibly by filtering out spurious gravity waves in the perturbation forecast that are generated due to the presence of unbalanced modes in the perturbation. We also showed that increasing the value of the regularization parameter β can act in a similar way by removing smaller-scale wavenumbers from the LETLM. In this paper, we had much less control on our ability to filter out unbalanced modes from either the forecast or the initial analysis perturbation. In fact, the only control we had was through a crude geostrophic balance in the 4DVAR error covariance. In other words, it is safe to

say that both the analysis increment and the initial conditions for the ensemble forecasts had unbalanced modes. However, our results show that the LETLM was effective even in the presence of unbalanced modes (e.g., see rapid propagation of an unbalanced mode in Fig. 11 above 10 hPa and centered on 30°N). We did notice that we had to increase the size of the influence volume considerably in the stratosphere to accommodate propagation of the rapidly evolving gravity waves.

Acknowledgments. This work was funded by the U.S. Office of Naval Research. Sergey Frolov and Craig H. Bishop acknowledge support from the Naval Research Base Program PE-0601153N. Douglas R. Allen and Karl W. Hoppel acknowledge support from Office of Naval Research base funding via Task BE-033-02-42. David D. Kuhl and Karl W. Hoppel acknowledge support from Office of Naval Research base funding via Task BE-435-050. Authors thank Philippe Lopez of ECMWF for useful feedback. All are grateful for the access to the Department of Defense high-performance computing resources that enabled us to conduct this research. We also thank three anonymous reviewers for the most thorough review of our manuscript.

APPENDIX

Efficient Implementation of the Matrix Inverse for LETLM

The exact inverse of the covariance matrix $\mathbf{C}$ in Eq. (2),

$$[\mathbf{C}]^{-1} = \left[(\mathbf{S}_p^{\text{vol}} \mathbf{x}_m)^T (\mathbf{S}_p^{\text{vol}} \mathbf{x}_m) \right]^{-1}, \quad (\text{A1})$$

can only be computed when the matrix $\mathbf{C}$ is full rank. By construction, covariance $\mathbf{C}$ in Eq. (A1) is always rank deficient because we center it on the ensemble mean. Furthermore, our prior experience using a shallow water model (Allen et al. 2017) showed that the covariance matrix $\mathbf{C}$ is not only rank deficient, but also poorly conditioned (e.g., the largest eigenvalue of $\mathbf{C}$ is several orders of magnitude larger than the smallest eigenvalue). Numerically stable inverse of poorly conditioned matrices requires some form of regularization that deweights or excludes eigenvectors with poorly known eigenvalues. Finally, our experience with the shallow water model also suggests that by tuning the regularization of the matrix inverse $[\mathbf{C}]^{-1}$, we can remove fine scales of variability that lack predictability and are often associated with small eigenvalues.

Hence, to improve the rank and to stabilize the condition of the covariance matrix $\mathbf{C}$, we employed a variant

of the Tikhonov regularization (Tikhonov and Arsenin 1977):

$$[\mathbf{C}]^{-1} \approx [\tilde{\mathbf{C}}]^{-1} = [\mathbf{C} + \beta' \mathbf{I}]^{-1}, \quad (\text{A2})$$

where β' is a tunable parameter related to the norm of the covariance matrix $\mathbf{C}$:

$$\beta' = \beta n^{\text{ens}} \text{eps} \|\mathbf{C}\|; \quad (\text{A3})$$

β is the tunable parameter discussed in section 5c, n^{ens} is the number of ensemble members, and eps is the machine round-off error for single precision arithmetic ($\sim 10^{-8}$). In our previous work (Frolov and Bishop 2016; Allen et al. 2017), we used the maximum singular value as a measure of the size of the covariance $\mathbf{C}$:

$$\|\mathbf{C}\|_{\text{max}} = \max(\{\sigma_i\}), \quad (\text{A4})$$

where σ_i are the singular values of the covariance $\mathbf{C}$ computed using the singular value decomposition

$$\{\sigma_i\} = \text{svd}(\mathbf{C}). \quad (\text{A5})$$

Recall that for covariance matrices, the singular and eigenvalues are identical.

Unfortunately, Eq. (A4) requires explicit computation of the eigenspectrum (singular spectrum) of $\mathbf{C}$, which is computationally expensive and is not always needed. To circumvent this need, we switched in this paper to a Frobenius norm,

$$\|\mathbf{C}\|_2 = \text{sum}(\{\sigma_i\}) \equiv \text{trace}(\mathbf{C}), \quad (\text{A6})$$

which can be computed by taking the trace of the covariance matrix $\mathbf{C}$. Because the spectrum of the covariance $\mathbf{C}$ is often steep, the Frobenius norm in Eq. (A6) is dominated by the largest singular value and, hence, is closely related to the max norm in Eq. (A4). In our experiments, we got very similar (if not indistinguishable) results when we switched to using Eq. (A6) instead of Eq. (A4).

Once we computed the regularized covariance $\mathbf{C}$, we then used either the singular value or the Cholesky decomposition to compute the vector $\mathbf{z}$:

$$\mathbf{z} = [\mathbf{C} + \beta' \mathbf{I}]^{-1} \mathbf{x}. \quad (\text{A7})$$

In the singular value method, we compute vector $\mathbf{z}$ as

$$\mathbf{z} = \mathbf{U} \mathbf{\Lambda}^{-1} \mathbf{U}^T \mathbf{x} = \mathbf{U} [\mathbf{\Lambda}^{-1} (\mathbf{U}^T \mathbf{x})], \quad (\text{A8})$$

where $\mathbf{U}$ and $\mathbf{\Lambda}$ are the outputs of the singular value decomposition of matrix $\mathbf{C}$:

$$\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{C} + \beta^2 \mathbf{I}$$

$$\mathbf{\Lambda}^{-1} = \text{diag}[1/\sigma]. \quad (\text{A9})$$

We found that LAPACK's routine `ssyevr` specifically designed for symmetric matrixes was 30%–50% faster than the generic singular vector routine `sgesvd`. We achieved additional computational savings by only computing the upper triangular of the symmetric matrix $\mathbf{C}$ in Eq. (A1) using LAPACK's routine `ssyrk`.

In the Cholesky decomposition method, we compute vector $\mathbf{z}$ in Eq. (A10) using `spotrs` routine in LAPACK:

$$\mathbf{z} \leftarrow \text{spotrs}(\sqrt{\tilde{\mathbf{C}}}, \mathbf{x}), \quad (\text{A10})$$

where $\sqrt{\tilde{\mathbf{C}}}$ is the square root of the regularized matrix $\tilde{\mathbf{C}}$ in Eq. (A2), computed using Cholesky factorization routine `spotrf`.

The Cholesky factorization method is computationally more efficient (by a factor of 2 in our experiments) than the singular value method because it is less expensive to compute the square root of the matrix than the singular spectrum and vectors of $\mathbf{C}$. The potential benefit of the singular value method is that it can directly diagnose the rank of the unregularized covariance $\mathbf{C}$. Using the singular value method, it is also possible to precompute and store the rows of the LETLM matrix for any future computation. However, because of the computational efficiency, we used the Cholesky method exclusively in our experiments.

The template for the LETLM computations is available in supplementary material to this article.

REFERENCES

- Allen, D. R., C. H. Bishop, S. Frolov, K. W. Hoppel, D. D. Kuhl, and G. E. Nedoluha, 2017: Hybrid 4DVAR with a local ensemble tangent linear model: Application to the shallow-water model. *Mon. Wea. Rev.*, **145**, 97–116, <https://doi.org/10.1175/MWR-D-16-0184.1>.
- Bishop, C. H., S. Frolov, D. R. Allen, D. D. Kuhl, and K. Hoppel, 2017: The local ensemble tangent linear model: An enabler for coupled model 4DVAR. *Quart. J. Roy. Meteor. Soc.*, **143**, 1009–1020, <https://doi.org/10.1002/qj.2986>.
- Bowler, N. E., and Coauthors, 2017: The effect of improved ensemble covariances on hybrid variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **143**, 785–797, <https://doi.org/10.1002/qj.2964>.
- 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**, 1567–1586, <https://doi.org/10.1175/2009MWR3158.1>.
- Eckermann, S. D., and Coauthors, 2009: High-altitude data assimilation system experiments for the northern summer mesosphere season of 2007. *J. Atmos. Sol.-Terr. Phys.*, **71**, 531–551, <https://doi.org/10.1016/j.jastp.2008.09.036>.
- Frolov, S., and C. H. Bishop, 2016: Localized ensemble-based tangent linear models and their use in propagating hybrid error covariance models. *Mon. Wea. Rev.*, **144**, 1383–1405, <https://doi.org/10.1175/MWR-D-15-0130.1>.
- Hogan, T. F., and Coauthors, 2014: The Navy Global Environmental Model. *Oceanography*, **27**, 116–125, <https://doi.org/10.5670/oceanog.2014.73>.
- Janisková, M., and P. Lopez, 2012: Linearized physics for data assimilation at ECMWF. ECMWF Tech. Memo. 666, 26 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2012/10163-linearized-physics-data-assimilation-ecmwf.pdf>.
- Kleist, D. T., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part II: 4DVar and hybrid variants. *Mon. Wea. Rev.*, **143**, 452–470, <https://doi.org/10.1175/MWR-D-13-00350.1>.
- Kuhl, D. D., T. E. Rosmond, C. H. Bishop, J. McLay, and N. L. Baker, 2013: Comparison of hybrid ensemble/4DVar and 4DVar within the NAVDAS-AR data assimilation framework. *Mon. Wea. Rev.*, **141**, 2740–2758, <https://doi.org/10.1175/MWR-D-12-00182.1>.
- Lorenc, A. C., N. E. Bowler, A. M. Clayton, S. R. Pring, and D. Fairbairn, 2015: Comparison of hybrid-4DVar and hybrid-4DVar data assimilation methods for global NWP. *Mon. Wea. Rev.*, **143**, 212–229, <https://doi.org/10.1175/MWR-D-14-00195.1>.
- McLay, J. G., C. H. Bishop, and C. A. Reynolds, 2008: Evaluation of the ensemble transform analysis perturbation scheme at NRL. *Mon. Wea. Rev.*, **136**, 1093–1108, <https://doi.org/10.1175/2007MWR2010.1>.
- Rosmond, T. E., 1997: A technical description of the NRL adjoint modeling system. NRL Publ. NRL/MR/7532-97-7230, 57 pp., <http://www.dtic.mil/dtic/tr/fulltext/u2/a330960.pdf>.
- , and L. Xu, 2006: Development of NAVDAS-AR: Non-linear formulation and outer loop tests. *Tellus*, **58A**, 45–58, <https://doi.org/10.1111/j.1600-0870.2006.00148.x>.
- Tikhonov, A., and V. Arsenin, 1977: *Solutions of Ill-Posed Problems*. Winston and Sons, 258 pp.
- Xu, L., T. Rosmond, and R. Daley, 2005: Development of NAVDAS-AR: Formulation and initial tests of the linear problem. *Tellus*, **57A**, 546–559, <https://doi.org/10.3402/tellusa.v57i4.14710>.