

Challenges of Increased Resolution for the Local Ensemble Tangent Linear Model

DOUGLAS R. ALLEN

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

SERGEY FROLOV^a AND ROLF LANGLAND

Marine Meteorology Division, Naval Research Laboratory, Monterey, California

CRAIG H. BISHOP

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

KARL W. HOPPEL AND DAVID D. KUHLE

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

MAX YAREMCHUK

Oceanography Division, Naval Research Laboratory, Stennis Space Center, Mississippi

(Manuscript received 21 January 2020, in final form 6 April 2020)

ABSTRACT

An ensemble-based linearized forecast model has been developed for data assimilation applications for numerical weather prediction. Previous studies applied this local ensemble tangent linear model (LETLM) to various models, from simple one-dimensional models to a low-resolution ($\sim 2.5^\circ$) version of the Navy Global Environmental Model (NAVGEM) atmospheric forecast model. This paper applies the LETLM to NAVGEM at higher resolution ($\sim 1^\circ$), which required overcoming challenges including 1) balancing the computational stencil size with the ensemble size, and 2) propagating fast-moving gravity modes in the upper atmosphere. The first challenge is addressed by introducing a modified local influence volume, introducing computations on a thin grid, and using smaller time steps. The second challenge is addressed by applying nonlinear normal mode initialization, which damps spurious fast-moving modes and improves the LETLM errors above ~ 100 hPa. Compared to a semi-Lagrangian tangent linear model (TLM), the LETLM has superior skill in the lower troposphere (below 700 hPa), which is attributed to better representation of moist physics in the LETLM. The LETLM skill slightly lags in the upper troposphere and stratosphere (700–2 hPa), which is attributed to nonlocal aspects of the TLM including spectral operators converting from winds to vorticity and divergence. Several ways forward are suggested, including integrating the LETLM in a hybrid 4D variational solver for a realistic atmosphere, combining a physics LETLM with a conventional TLM for the dynamics, and separating the LETLM into a sequence of local and nonlocal operators.

1. Introduction

Tangent linear models (TLMs) and their adjoints play a key role in generating optimal initial conditions in four-dimensional variational (4DVar) data assimilation

(DA) systems used for numerical weather prediction (NWP) (e.g., Rabier et al. 2000; Rosmond and Xu 2006; Gauthier and Thépaut 2001; Gauthier et al. 2007; Rawlins et al. 2007; JMA 2019; Zhang et al. 2019). Initial conditions are obtained by minimizing a cost function that quantifies the combined error-weighted differences between the forecast and observations and between the forecast and a control background forecast. The minimization requires the cost function gradient relative to the state vector, which has millions of elements.

^a Current affiliation: Cooperative Institute for Research in Environmental Sciences, Boulder, Colorado.

Corresponding author: D. R. Allen, douglas.allen@nrl.navy.mil

TLMs are used to determine this gradient and to propagate initial conditions forward over the DA time window. Ideally, TLMs incorporate all features of the nonlinear forecast model. In practice, development and maintenance of TLMs of physical parametrizations can be difficult; therefore, operationally they only approximate the ideal linear model (Janisková and Lopez 2012; Zhang et al. 2019). In some circumstances, this limitation is not serious [e.g., when neglected processes are too slow to noticeably impact the state on the DA time scale (6–12 h)]. In other cases, the limitation may be crucial and may even disallow certain observations from being assimilated (Geer et al. 2017). While certain DA approaches avoid TLMs, variational DA based on TLMs continues to exhibit superior forecast accuracy (e.g., Lorenc et al. 2015; Poterjoy and Zhang 2015; Bowler et al. 2017).

One alternative approach to conventional TLMs is the local ensemble tangent linear model (LETLM), in which the forecast is determined from nonlinear ensemble forecasts within a “local influence volume,” which we define as a specified geometric shape surrounding a grid point. This statistical approach has been tested in a hierarchy of cases, starting with simple models in Frolov and Bishop (2016) and Bishop et al. (2017). The latter proved that when the time evolution of each model variable over a single LETLM time step Δt depends only on variables within the local influence volume, then the LETLM is guaranteed to be accurate when the ensemble size exceeds the size of the “computational stencil,” which we define here as the number of grid points contained within the local influence volume multiplied by the number of variables used for the LETLM. Bishop et al. (2017) also demonstrated how the LETLM and its adjoint can be applied to 4DVar in strongly nonlinear regimes where multiple outer loops are required to achieve convergence. Allen et al. (2017) then demonstrated an accurate LETLM-based hybrid 4DVar system using a global shallow-water model. The first LETLM demonstration using an NWP forecast model (Frolov et al. 2018, hereafter F18) showed that, when low resolution was used (triangular truncation of T47 or $\sim 2.5^\circ$), the LETLM successfully forecasted realistic analysis perturbations, with skill exceeding the conventional TLM in the troposphere, but slightly worse in the stratosphere and mesosphere. Considerable attention was paid to the sensitivity of the LETLM to model parameters and tuning of the system for optimal results.

This paper furthers LETLM development for NWP using simulations at 1° (~ 100 km) resolution, which is the inner-loop resolution currently employed in the U.S. Navy’s operational global NWP system. A central assertion is that a precise LETLM approximation to the

true TLM is possible if the ensemble size exceeds the computational stencil size. For a fixed LETLM Δt , the required influence volume remains the same (due to the speed of wave propagation), but the computational stencil size increases with increased model resolution. Hence, to maintain the same approximate LETLM accuracy the ensemble size used to construct the LETLM should increase. F18 used a large ensemble size (up to 400); larger ensembles would be highly undesirable due to computational constraints. To cope with this problem, Yaremchuk et al. (2020) developed an extension of the LETLM that relaxes the locality assumption, allowing accurate (nonlocal) ETLMs for semi-implicit operators. This development may help avoid substantial increases of the ensemble size in future ETLM applications (more details in section 6).

Here we investigate several approaches to increase LETLM accuracy. First, we modify the geometry of the local influence volume to better account for physics processes that use a vertical column. Second, we switch from full Gaussian grid to a thin grid to avoid a 10-fold increase in density from equator to pole. And third, we attempt a reduced Δt , which should in principle lead to the reduction in the optimal computational stencil size.

In addition, the F18 results suggested that the LETLM was struggling to compete with the TLM in the upper stratosphere and lower mesosphere (USLM). This article investigates a potential cause of the degradation: rapidly propagating large-scale gravity waves, or so-called normal modes. The conventional TLM captures these using nonlocal operators associated with semi-implicit time stepping and global spectral transforms. The LETLM, however, is limited by the finite local influence volume. We will examine the effects of these waves on USLM performance by applying nonlinear normal mode initialization, while retaining the local nature of the LETLM.

The paper is organized as follows. Descriptions of the forecast model and LETLM are provided in section 2. Section 3 discusses normal mode initialization and its potential impact on LETLM forecast skill. Section 4 compares LETLM and TLM errors for a test case in November 2014. Sections 5 and 6 give a summary and discussion, respectively.

2. Model description

a. NAVGEM forecast model

The NAVGEM atmospheric forecast model employs a semi-Lagrangian/semi-implicit integration of the hydrostatic dynamical equations, the first law of thermodynamics, and conservation of moisture and ozone (Hogan et al. 2014). The configuration used here has a

horizontal resolution of T119 (Gaussian grid of 360 longitudes \times 180 latitudes), which is the current inner-loop operational resolution. Vertically, the model uses 60 levels (top at 0.05 hPa, ~ 65 km) with a hybrid-sigma coordinate (Eckermann 2009). The model is run with a 15 min time step and the model state is saved every time step. The predicted variables are vorticity, divergence, virtual potential temperature, specific humidity (Q), and surface pressure. In addition, the zonal (U) and meridional (V) wind, temperature (T), geopotential height (Z), $\dot{\eta}$ (vertical motion, where η represents the hybrid model levels), and $3D$ pressure (P) are derived fields. As in F18, the LETLM variables include U , V , T , P , Z , $\dot{\eta}$, and Q . NAVGEM incorporates several stochastic physics packages, including stochastic kinetic energy backscatter, 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 LETLM development, we therefore turned off these stochastic processes. One additional change is the timing of when the model fields are output. As coded, NAVGEM calculates the dynamics followed by the physics for each time step. Operationally, wind and temperature fields are saved before the physics calculation, while the water and ozone are saved after the physics. To synchronize all model variables, we modified the code to save all fields after the physics to complete the time step.

The NAVGEM DA solver is a hybrid 4DVar system (Kuhl et al. 2013) that employs a strong-constraint approach using the ensemble transform (ET) method (McLay et al. 2008) with 80 members. The 6 -h cycling window is centered at 0000 , 0600 , 1200 , and 1800 UTC. The background includes the last 6 h of a 9 -h forecast from the middle of the previous window. The solver employs the accelerated representer method described by Xu et al. (2005) and Rosmond and Xu (2006). The original NAVGEM TLM and adjoint models, described in Rosmond (1997), were based on the earlier Navy Global Atmospheric Prediction System (NOGAPS). While this NOGAPS TLM is currently operational, we use a newly developed semi-Lagrangian (SL) TLM, which is expected to become operational in 2020. The SL TLM has simplified boundary layer physics, including vertical diffusion, simplified gridscale precipitation, and simplified convection. It neglects several other physical processes including gravity wave drag, radiation, and ozone photochemistry. Note that the preoperational version of the SL TLM used for this study did not yet include moist physics (precipitation and convection). Also, for this study, several of the SL TLM settings were

adjusted to match the nonlinear forecast model, including the same horizontal and vertical resolution, time step, horizontal diffusion, and sponge layer levels.

b. LETLM description

The LETLM is described in F18. Here we summarize the formulation and highlight new aspects used in this paper. We note here that since the LETLM is still in the development state, it has not yet been optimized for timing tests, so we do not present efficiency comparisons with the TLM. However, discussions of computational requirements and sensitivities in section 8 of F18 are still relevant to this study.

The LETLM is a sparse matrix \mathbf{M} approximating the evolution of the deviations $\delta\mathbf{X}$ of the ensemble \mathbf{X} of n model states from their mean:

$$\delta\mathbf{X}_{m+1} = \mathbf{M}\delta\mathbf{X}_m + \xi_m. \quad (1)$$

Here n is ensemble size, N is the size of the model state vector, and \mathbf{X}_{m+1} and \mathbf{X}_m are $N \times n$ matrices representing ensemble states listed columnwise at time indices $m+1$ and m , and ξ_m is the ensemble of Taylor series truncation errors. For large ensembles ($n = N$) spanning the entire state space, \mathbf{M} can be retrieved by minimizing the norm of ξ_m with respect to the elements of \mathbf{M} . For realistic ensemble sizes ($n \sim 100$), the retrieval can also be performed if the nonlinear model operators driving the ensemble are local (i.e., the approximating matrix \mathbf{M} is sparse and has at most n nonzero elements in a set of rows associated with a given grid point). Under this locality assumption the number of unknowns does not exceed the number ($N \times n$) of linear constraints in Eq. (1), and the respective solution for the rows of \mathbf{M} associated with the p th grid point can be represented by

$$\mathbf{M}^p = (\hat{\mathbf{S}}^p \circ \delta\mathbf{X}_{m+1})[(\mathbf{S}^p \circ \delta\mathbf{X}_m)(\mathbf{S}^p \circ \delta\mathbf{X}_m)]^{-1}(\mathbf{S}^p \circ \delta\mathbf{X}_m)^T, \quad (2)$$

where \circ stands for the Schur (elementwise) product, \mathbf{S}^p is an $N \times n$ selection matrix whose (identical) columns contain ones in the grid points where the elements of \mathbf{M} are expected to influence the model state at the p th point and zeros elsewhere, while $\hat{\mathbf{S}}^p$ selects only the ensemble variables in the p th grid point. The action of \mathbf{M} on a state vector \mathbf{x}_m can then be computed as

$$\mathbf{x}_{m+1} = \sum_{p=1}^{N_{\text{grid}}} [\hat{\mathbf{s}}_p^T \mathbf{M}^p (\mathbf{s}_p \circ \mathbf{x}_m)], \quad (3)$$

where N_{grid} is the number of model grid points, and \mathbf{s}_p and $\hat{\mathbf{s}}_p$ are the columns of \mathbf{S}^p and $\hat{\mathbf{S}}^p$. In LETLM computations, the ensemble members are also normalized

(Eq. (3) in F18) to improve the condition numbers of the matrices in Eq. (2).

Equation (3) exposes two major challenges for LETLM implementation. The first is the necessity to know the exact structure of \mathbf{S}^p [locations of the nonzero elements within the LETLM local influence volume, referenced by the symbol ω_p , involved on the right-hand side of Eq. (2)]. F18 used the cylindrical vicinity surrounding the p th point. As we will show, further refining of ω_p based on physical considerations may substantially improve the LETLM.

The second challenge is the assumption of locality. Most NWP models, including NAVGEM, contain non-local operators, such as integrals for pressure computation or semi-implicit solvers for filtering fast gravity waves. As a consequence, the locality assumption $|\mathbf{s}_p| < n$ is violated for manageable ($n \sim 100$) ensemble sizes, and the LETLM loses accuracy in representing the TLM with increasing model resolution. To deal with this problem, one has to either increase the ensemble size to satisfy the solvability condition for Eq. (1), or augment the equation with additional constraints by replacing the inverse of the local ensemble correlation matrix in Eq. (2) by a pseudoinverse (see the appendix of F18 for details on applying the pseudoinverse). In the latter case, the major difficulty is correctly defining the pseudoinverse metric.

One possibility to mitigate the problem of nonlocality is to compute the LETLM on a thinner grid followed by reinterpolation to the grid of the parent model. Another option is to reduce the size of Δt for the LETLM in order to shrink ω_p . A third option is to alter the shape of ω_p to account for physical processes with known spatial locality, such as physics parameterizations acting in a vertical column. In what follows, we will explore the impact of these approaches using a fine-resolution (1°) NAVGEM model that was used for low-resolution LETLM tests in F18.

c. Changes from the previous formulation of the LETLM

Moving to higher resolution increases the number of grid points in the computational stencil, which grows as the inverse square of the model grid spacing. In addition to the increased computational expense, the computational stencil size should be roughly equivalent to the ensemble size. It is undesirable that increases in resolution should require more ensemble members. To ameliorate this behavior, we followed traditional numerical methods where increasing horizontal resolution results in decreasing Δt . We will compare results using $\Delta t = 15$ and 30 min to the 60-min Δt used in F18. We expect that smaller Δt will lead to smaller local influence

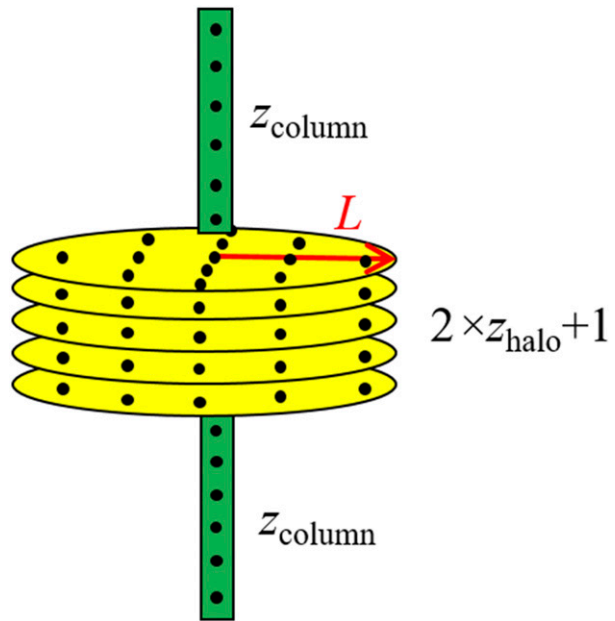


FIG. 1. The LETLM local influence volume ω_p (colored regions) is composed of a central cylindrical region of radius L (yellow), with the number of vertical levels equal to $2 \times z_{\text{halo}} + 1$. In addition, a vertical column (green) is included that extends z_{column} levels above and below the highest and lowest layers of the cylinder. This figure illustrates the configuration used in this study ($z_{\text{halo}} = 2$ and $z_{\text{column}} = 6$). The computational stencil, as defined in this paper, includes all of the grid points (dots) that are located within the local influence volume, multiplied by the number of variables. Here the computational stencil size is $(13 \times 5 + 6 \times 2) \times 7$ variables = 539.

volumes, increased skill, and reduced computational time of LETLM forecasts. We test these expectations in section 4b.

We also modified the shape of the local influence volume ω_p that is used to determine the computational stencil (the total of all model grid points that lie inside ω_p multiplied by the number of variables). Following the observation that dynamical processes often operate in a horizontal plane, while physics processes operate in a vertical 1D column, we designed a ω_p that combines a cylinder (as in F18) with an additional vertical column. The cylinder (yellow region in Fig. 1) has a radius L and encompasses all model levels within $\pm z_{\text{halo}}$ of the central level. Away from the upper and lower boundaries, the computational stencil includes $2 \times z_{\text{halo}} + 1$ levels, while near the boundaries the computational stencil is limited to the number of available levels (e.g., the lowest model layer will include only $z_{\text{halo}} + 1$ levels in the stencil). The new component of ω_p is a single vertical column that extends beyond the cylinder directly above and below the central point (green regions in Fig. 1). The integer value z_{column} is chosen so that an additional $2 \times z_{\text{column}}$ points

are added to the computational stencil away from the boundaries (this quantity is similarly limited near the boundaries). The additional expense of using z_{column} is small, but provides a noticeable benefit (about a few percent). For the LETLM tests performed in this study, we use fixed $z_{\text{halo}} = 2$ and $z_{\text{column}} = 6$, based on several tuning tests.

In addition to the new ω_p , we implemented the LETLM on a reduced grid. In F18, we used the full Gaussian grid, which has the same longitude grid at each latitude (Fig. 2a, black line), so the density of grid points varies greatly with latitude. NAVGEM can be run on a “thin grid,” where the longitude grid varies with latitude to better match the inherent resolution of the spectral model (red line on Fig. 2a). The number of grid points for a single NAVGEM T119 level is 64 800 for full grid and 42 984 for the thin grid, resulting in $\sim 33\%$ reduction in total points. The benefit to the LETLM is significant in that for fixed L , the computational stencil size as a function of latitude is more uniform. Figure 2b compares the number of horizontal points included in the computational stencil for the NAVGEM T119 full and thin grids for $L = 500$ km, showing that while the full-grid ranges from 69 to 1552, the thin grid ranges from 63 to 135. The thin grid speeds up the LETLM calculations by an order of magnitude without reducing forecast skill. All NAVGEM LETLM tests are now run with the T119 thin grid. Note, however, that the nonlinear forecasts and the TLM forecast were actually made with the full T119 grid, and the output from these was converted to the T119 thin grid in postprocessing.

Finally, we also introduced minor changes to the calibration protocol. F18 examined sensitivity to several model parameters, but the main ones were the local influence volume cylindrical radius L and the unitless pseudoinverse cutoff parameter β (see the appendix of F18 for more details on the parameter β). Here we also tune these two parameters, but change the values used; L varies from 50 to 1250 km, in 50 km increments, and for β we use values of $10^{(i-5)/5}$ where i ranges from 0 to 10 by increments of 1.0 (exact values of β are 0.100, 0.158, 0.251, 0.398, 0.631, 1.000, 1.585, 2.511, 3.981, 6.301, and 10.0). Using all combinations of L and β , we perform 220 offline calibration forecasts for each full tuning experiment. As in F18, optimal values of L and β are determined for each model level k . To select these optimal profiles $L_{\text{opt}}(k)$ and $\beta_{\text{opt}}(k)$, we compare the 3-h (i.e., at the center of the NAVGEM analysis window) normalized error $\varepsilon(k)$ (section 2e describes error calculations) for U , V , and T against the known truth for the perturbation forecast, which is computed using a pair of nonlinear model forecasts. For each level, we select the combinations of $L_{\text{opt}}(k)$ and $\beta_{\text{opt}}(k)$ that minimize $\varepsilon(k)$

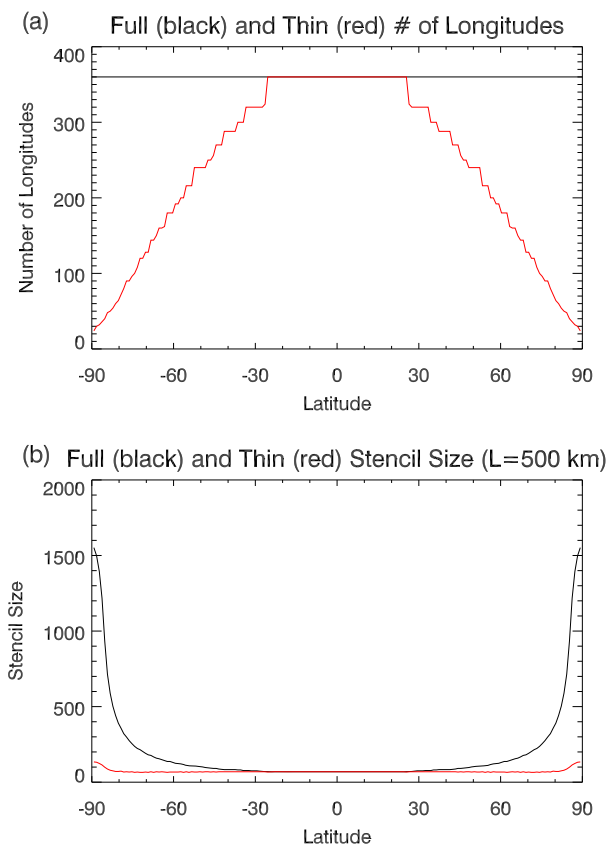


FIG. 2. (a) The number of longitudes at each latitude associated with the full (black) and thin (red) grid for triangular truncation T119. (b) The number of horizontal points in the computational stencil as a function of latitude at 0° longitude for radius of 500 km and for the full (black) and thin (red) grid.

resulting in the optimal error profile $\varepsilon_{\text{opt}}(k)$. The LETLM configuration used in this study is summarized in Table 1.

d. Experimental design

The basic experimental design is similar to F18, except for higher horizontal resolution (T119, $\sim 1.0^\circ$ rather than T47, $\sim 2.5^\circ$). In addition, while F18 assimilated conventional observations and AMSU-A radiances, here we use a more realistic observation suite that includes these observations along with radiances from CrIS, AMSU-B, Aqua, IASI, MHS, SSMIS, and ATMS, ozone from SBUV and OMPS, and GPS radio occultations. The number of observations accepted in the 6-h cycle used for this study was ~ 3.4 million. As in F18, we use the climatological covariance for cycling and for calculating the analysis perturbations. For the LETLM, rather than using the standard ET ensembles, we initialize ensembles using random samples from archived T425 (downscaled to T119 for this study) analysis perturbations from 0000 UTC 21 November 2014 to 1800 UTC

TABLE 1. Parameters used for the reference configuration of the LETLM and additional sensitivity tests.

Parameter	Reference configuration value	Additional sensitivity tests
Date time group	0000 UTC 17 Nov 2014	
Time step	60 min	15 and 30 min
Forecast time	3 h	
z_{halo}	2 (i.e., 5 vertical levels)	
z_{column}	6 (i.e., 12 additional points)	
Variables	U, V, T, P, Z, σ, Q	
β	$10^{(i-5)/5}, i = 0, 1, 2, \dots, 10$	
Ensemble size	400	50, 100, 200, and 300
L	50, 100, \dots , 1250 km	
Ensemble initialization	Archived analysis perturbations	

2 March 2015. To generate a test perturbation, we cycled at T119 from 0000 UTC 15 November to 0000 UTC 17 November 2014, and used the perturbation for 0000 UTC 17 November 2014 for the LETLM tests. Operationally, the TLM runs over the analysis time ± 3 h. To accommodate the initial condition time we shift the time sequence to run the TLM (and LETLM) from the analysis time to $+6$ h. The nonlinear forecasts used for calculating the truth are also run over this time window.

Maps of initial U , V , and T perturbations at four levels are provided in Fig. 3. The perturbation sizes increase with altitude from the surface (bottom row) up to the stratopause (top row). Finer spatial scales are seen at the surface and middle troposphere (493 hPa) than in the middle stratosphere (10.5 hPa) and at the stratopause (1 hPa), where perturbations are quite broad, reflecting larger-scale dynamics. Vertical profiles of perturbation size will be shown in section 4d. We note that in this paper, we only examine forecasts from one representative perturbation rather than performing statistical analyses over many perturbations. Our goal is to understand the LETLM mechanics and sensitivities, rather than exhaustive validation. The true test for NWP will come when the LETLM is fully integrated into a cycling hybrid 4DVar system.

e. Error metrics

We use similar error metrics to F18 to quantify the LETLM and TLM skill. Globally averaged root-mean-square errors (RMSEs) are calculated for each variable as follows:

$$U_{\text{RMSE}}(k) = \left\{ \frac{1}{n^{\text{thin}}} \sum_{i=1}^{n^{\text{thin}}} [U_{\text{TRUTH}}(i, k) - U_{\text{LETLM}}(i, k)]^2 \right\}^{1/2}, \quad (4)$$

where i and k are indices for model grid point and level, respectively, and n^{thin} is the number of points on the

thin grid. Next, a normalized error metric combines U , V , and T errors:

$$\varepsilon(k) = \sqrt{\frac{U_{\text{RMSE}}^2(k) + V_{\text{RMSE}}^2(k) + T_{\text{RMSE}}^2(k)}{U_{\text{PERT}}^2(k) + V_{\text{PERT}}^2(k) + T_{\text{PERT}}^2(k)}}. \quad (5)$$

Here PERT indicates the size of the TRUTH perturbation, and $U_{\text{PERT}}(k)$ is obtained from Eq. (4) by zeroing out $U_{\text{LETLM}}(i, k)$. Finally, $\varepsilon(k)$ is integrated vertically to give a single metric:

$$\bar{\varepsilon} = \frac{1}{n^{\text{lev}}} \sum_{k=1}^{n^{\text{lev}}} \varepsilon(k). \quad (6)$$

Here we equally weight all model levels. Since the NAVGEM levels are more closely spaced near the surface (e.g., half are below 200 hPa), this metric favors the troposphere.

3. Normal mode initialization

a. Normal modes in NAVGEM

NAVGEM can be run with normal mode initialization (NMI) applied to any nonlinear forecast. The NMI code is discussed in the NOGAPS reference manual (Hogan et al. 1992) and generally follows Machenhauer (1977). NMI uses a nonlinear iterative approach to ensure that the time tendencies of the coefficients of the selected inertio-gravity modes are approximately zero. The NOGAPS NMI uses the hybrid vertical coordinate, which differs from Žagar et al. (2015), who use the pure sigma formulation of Kasahara and Puri (1981). We initialize three vertical normal modes (NM), associated with equivalent depths of 10.147, 5.959, and 2.787 km (the leading eigenvalues of the vertical structure equation). All three NM peak in the lower mesosphere (above 1.0 hPa) and are large throughout the USLM (~ 10 hPa to the top of the model), but are small in the

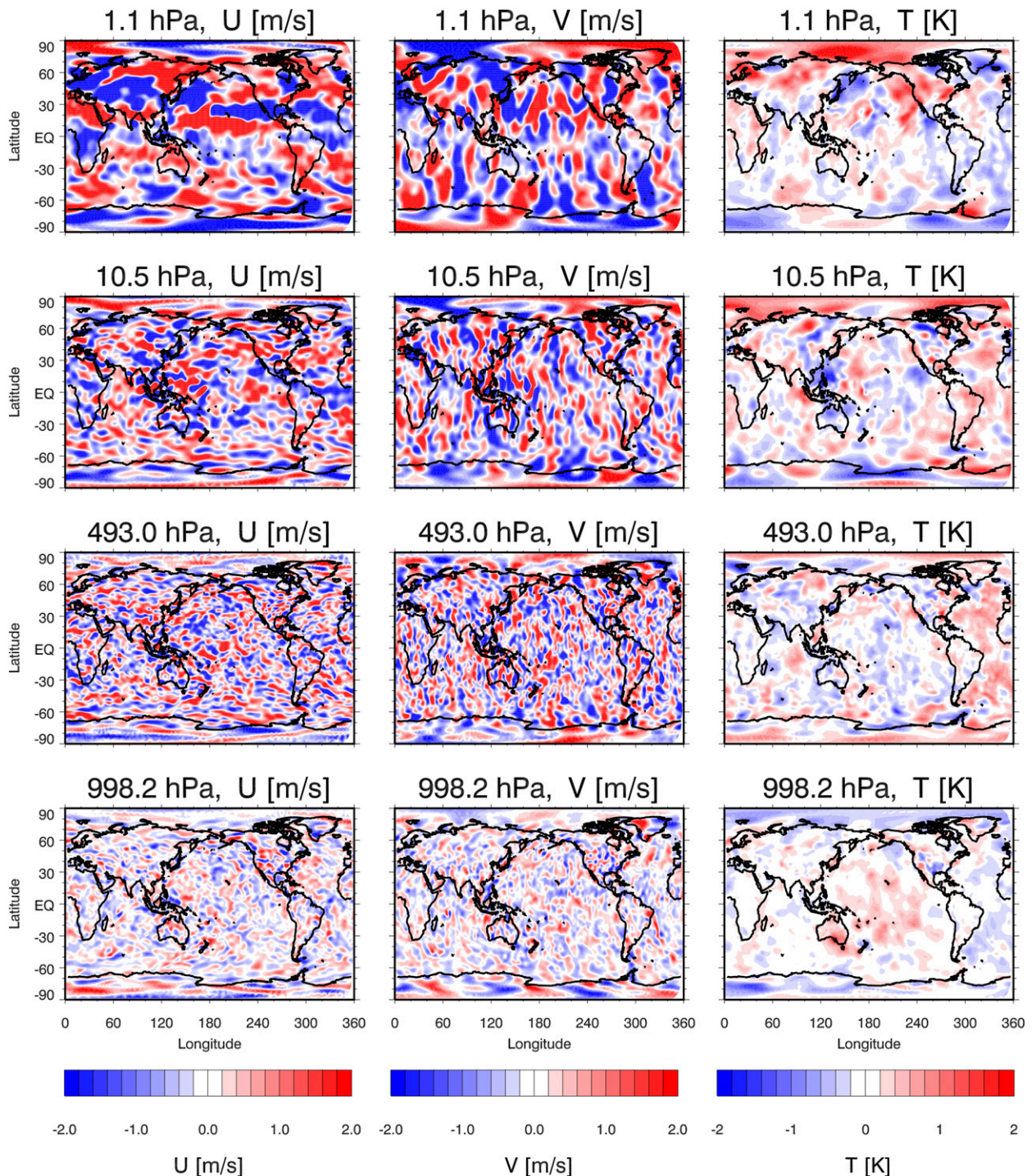


FIG. 3. The initial perturbations at 0000 UTC 17 Nov 2014 for U , V , and T at four model levels specified by a nominal pressure associated with a standard profile over the ocean. (top row) The stratopause (1.1 hPa, ~ 50 km), (second row) the middle stratosphere (10.5 hPa, ~ 30 km), (third row) the middle troposphere (493 hPa, ~ 6 km), and (bottom row) the boundary layer (998.2 hPa).

troposphere (below ~ 100 hPa). The horizontal structures of the normal modes are solutions of the shallow-water model equations with mean depth equal to the equivalent depths associated with each vertical mode. We will

examine these solutions in [section 3b](#). For this study, we set the cutoff frequency for initialized modes to 1.0 day^{-1} .

To analyze the NM structures, [Figs. 4a–d](#) provide vertical cross sections in longitude and pressure of the

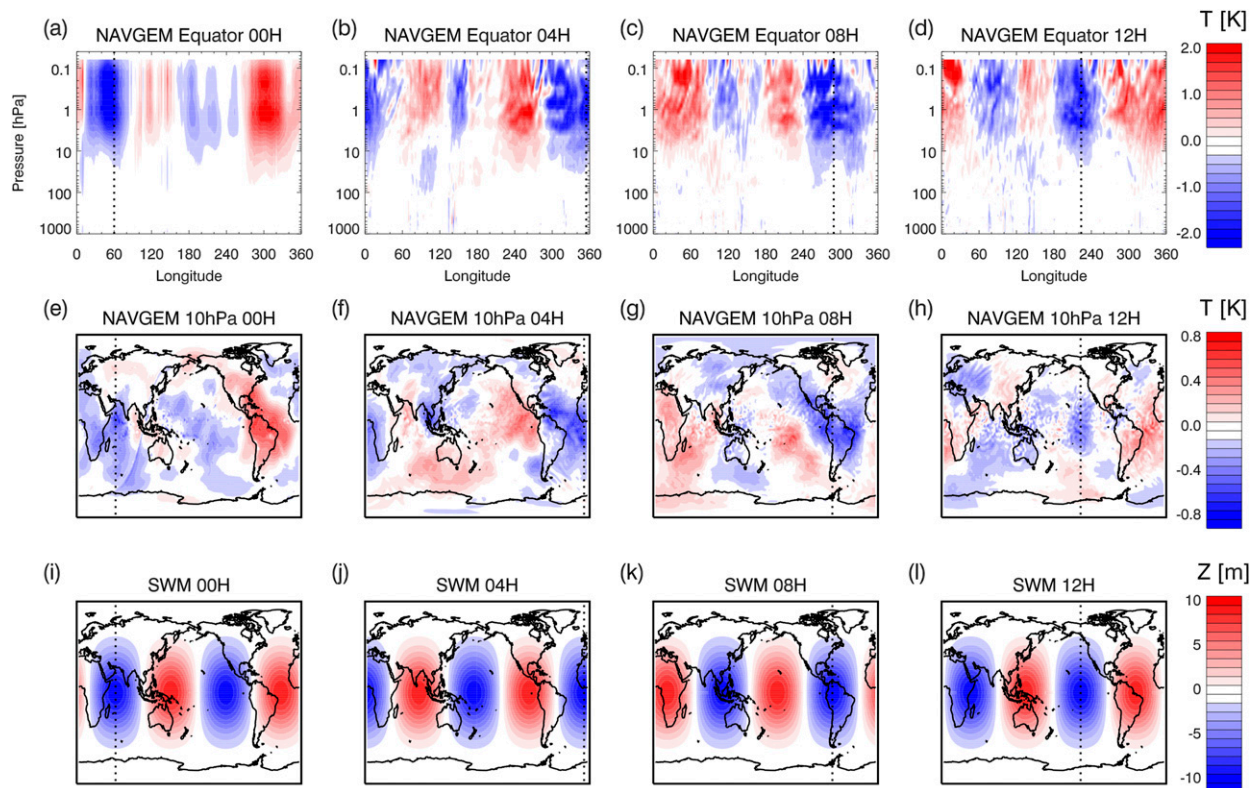


FIG. 4. (a)–(d) Difference between NAVGEM forecasts of T with and without normal mode initialization, starting from 0000 UTC 17 Nov 2014. Plots show longitude–pressure cross sections at the equator for forecast times 0, 4, 8, and 12 h. In all plots, the vertical dotted line identifies the approximate location of a trough in T , the movement of its location from one panel to the next corresponds to westward motion with a phase speed of 500 m s^{-1} . (e)–(h) Difference maps at 10.5 hPa ($\sim 32 \text{ km}$) for the same NAVGEM forecasts of T with and without normal mode initialization. (i)–(l) Shallow-water model forecast of the height for the WG(2, 2) normal mode associated with a mean depth equal to the first eigenvalue of the NAVGEM T119L60. The SWM is run at T119 ($\sim 1^\circ$ resolution) with a mean depth of $\sim 10 \text{ km}$ and basic state at rest.

equatorial T difference between 12-h deterministic forecasts from a 4DVar analysis with and without NMI. At 0 h, a zonal wave 2 structure exists at high altitude. Horizontal maps of the T difference at 10.5 hPa (Figs. 4e–h) show a wave 2 structure superposed with smaller-scale features. Forecasts with and without model physics (not shown) indicate the wave 2 structure is largely forced by radiation at upper levels combined with tropical convective processes that affect geopotential heights above. Similar mechanisms force the migrating semidiurnal tide (Hagan and Forbes 2003), but these freely propagating NM are distinct from the forced semidiurnal tide. NAVGEM forecasts without physics indicate this wave persists for at least 5 days, although at reduced amplitude. The wave 2 structure propagates $\sim 180^\circ$ westward over 11 h, for an equatorial phase speed (c) of $\sim 500 \text{ m s}^{-1}$ (dotted lines in Fig. 4 mark $c = 500 \text{ m s}^{-1}$).

b. LETLM sensitivity to c using the SWM

Calculations of the horizontal NM were computed using the shallow-water model (SWM) described in

Allen et al. (2015), with mean depth equal to the gravest equivalent depth of 10.147 km. The westward gravity (WG) mode with total wavenumber $n = 2$ and zonal wavenumber $m = 2$ [i.e., the WG(2, 2) mode], has a similar structure to that observed here, with $c = 504 \text{ m s}^{-1}$ (period of 11.1 days), consistent with the features in NAVGEM. Figures 4i–l provides global maps of the SWM height (Z) from a forecast initialized with the WG(2, 2) mode upon a basic state at rest, showing uniform westward propagation with $c \sim 500 \text{ m s}^{-1}$. There are other waves present in NAVGEM (Figs. 4e–h), but this WG(2, 2) mode plays a large role.

Fast-moving large-scale modes may cause problems for the LETLM, since it requires localization due to finite ensemble size. For each $\Delta t = 3600 \text{ s}$, the WG(2, 2) mode travels $\sim 1800 \text{ km}$, similar to the optimal LETLM localization lengths $\sim 1750 \text{ km}$ obtained in F18 (see Fig. 3a of F18). The sensitivity of LETLM errors to c is tested with the SWM experimental design of Allen et al. (2016), using T21 ($\sim 5.6^\circ$) resolution and mean depth of 10 km. First, a 100-member ensemble was created with

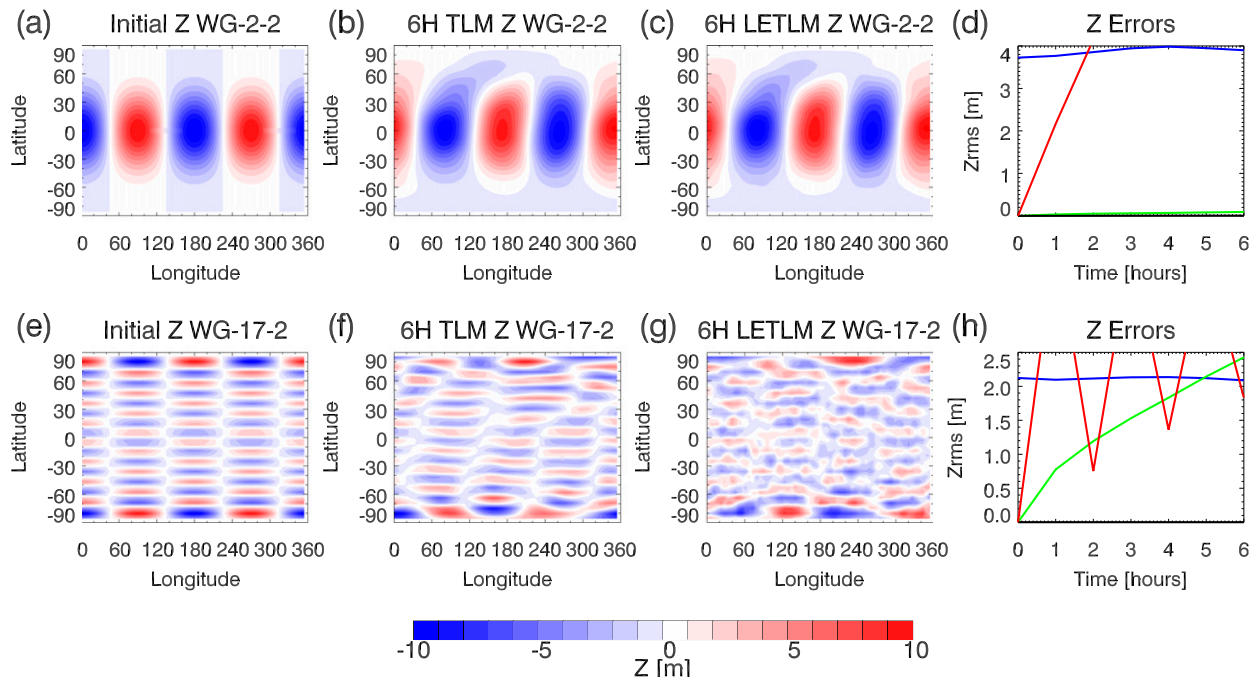


FIG. 5. Shallow-water model simulations at T21 ($\sim 5.6^\circ$) of the (top) WG(2, 2) and (bottom) WG(17, 2) modes using the system described in Allen et al. (2016). The WG(2, 2) mode moves westward with a phase speed of 501 m s^{-1} and the WG(17, 2) mode moves westward at 2750 m s^{-1} . (a),(e) Z perturbations at initialization. (b),(f) 6-h forecasts of Z from the TLM (the true perturbation at 6 h, not shown, is nearly identical to the TLM). (c),(g) 6-h forecasts of Z from the LETLM with 100 members and time step of 1 h. (d),(h) The Z globally averaged errors for the TLM (black line, near zero), LETLM (green line), persistence (red line), and size of the perturbation (blue line).

an ensemble Kalman filter that assimilated 6 days of fabricated observations from a SWM forecast with topographic forcing to generate realistic Northern Hemisphere (NH) wintertime dynamics. In Allen et al. (2016), this ensemble was used to propagate the subsequent analysis perturbation in order to tune the pseudoinverse parameter for the LETLM with a fixed local influence volume radius $L = 2000 \text{ km}$ and $\Delta t = 1 \text{ h}$. Here we use the same tuned LETLM parameters, but substitute the analysis perturbation with single NM structures having c ranging from 500 to 3362 m s^{-1} (these are the WG modes for the T21 system with fixed zonal wave-number ($m = 2$), and total wavenumber (n) from 2 to 21. The truth was calculated for each NM as the difference between the background and background plus NM nonlinear forecasts. Note that the background is not at rest in these calculations, so the wave structures become distorted, unlike the ideal structures shown in Figs. 4i–l.

Figure 5 (top row) shows Z maps for the WG(2, 2) mode. Initial perturbations are plotted in Fig. 5a, and TLM and LETLM 6-h forecasts are plotted in Figs. 5b and 5c, respectively. The truth forecasts are nearly identical to the TLM forecasts and are therefore not

shown. Both the TLM and LETLM accurately forecast this mode out to 6 h, as seen in the error time series (Fig. 5d). Results for the WG(17, 2) mode (bottom row), which has a westward phase speed of 2750 m s^{-1} , are provided in Figs. 5e–h. While the TLM accurately forecasts this wave, the LETLM has essentially no skill at 6 h, since the errors are larger than the perturbation.

A summary of 6-h forecast errors is provided in Fig. 6, using the metric of Eq. (5). While the TLM propagates all modes with high accuracy ($\varepsilon < 0.003$), LETLM errors increase sharply with c , particularly beyond $\sim 1000 \text{ m s}^{-1}$, and the LETLM has virtually no skill (i.e., $\varepsilon > 1$) for $c > \sim 2500 \text{ m s}^{-1}$. We infer that applying NMI to nonlinear forecasts used in the LETLM and the truth will lead to error reductions. This was previously tested by Allen et al. (2017), who showed that applying NMI significantly reduces the LETLM forecast errors (see Fig. 7 of Allen et al. 2017). In the next section, we will examine LETLM forecasts with NAVGEM by running the system with NMI (denoted NMIT, where T stands for true) and without NMI (denoted NMIF, where F stands for false). In the NMIT experiments, NMI is applied to both the control and the ensemble forecasts, while in NMIF, there is no application of NMI.

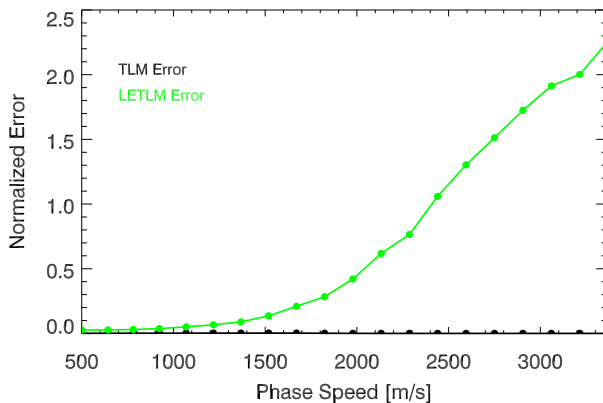


FIG. 6. LETLM and TLM errors for the propagation of a single normal mode with specified phase speed over a 6-h period using a shallow-water model system with resolution T21 and mean depth of 10 km. The modes are all westward gravity modes with zonal wavenumber 2 and total wavenumbers from 2 to 21.

4. Detailed comparison of LETLM and TLM errors for NAVGEM

a. Sensitivity to ensemble size

We now examine LETLM and TLM errors for NAVGEM. We first illustrate the sensitivity of LETLM skill to ensemble size to motivate the use of ensembles larger than the currently operational NAVGEM ensemble (80 members). Errors for 3-h LETLM forecasts were calculated using a modified tuning procedure in which we fix $\beta_{\text{opt}}(k) = 1.0$ and only tune $L(k)$. We use $\Delta t = 60$ min and ensemble sizes of 50, 100, 200, 300, and 400. The $\bar{\epsilon}$ values are plotted in Fig. 7a. In both NMIF and NMIT, there is strong sensitivity to ensemble size, with monotonic improvements out to 400 members. These tests suggest that 80 members are likely too few for an accurate LETLM, but additional improvement beyond 400 members may not be worth the additional computational cost. We therefore limited the tests to 400 members as an upper realistic limit given current computational requirements for NWP. Also, we note that NMIT errors are smaller than NMIF errors for all ensemble sizes. A detailed comparison of NMIT and NMIF will be provided below.

b. Sensitivity to Δt

With a fixed ensemble size of 400, we next examined the LETLM sensitivity to Δt . For both NMIF and NMIT, we calculated 3-h errors using $\Delta t = 15, 30$, and 60 min (900, 1800, and 3600 s), using full tuning for L and β for each case. Resulting errors are presented in Fig. 7b. While previous tests with the SWM showed LETLM errors increasing monotonically with Δt (Fig. 16f of Allen et al. 2017), here the errors are similar for NMIF

at $\Delta t = 15, 30$, and 60 min ($\bar{\epsilon} = 0.305, 0.300, 0.305$). For NMIT, errors are also similar ($\bar{\epsilon} = 0.264, 0.254, 0.255$), but have a sharper drop from 15 to 30 min. Larger errors for smaller Δt may be due to noise in the LETLM forecast that becomes amplified with the recursive application of the LETLM operator. Further tests are needed to determine the exact cause of the error behavior, but it is clear that our expectation that smaller Δt would result in smaller errors was not true.

We also expected that smaller Δt would be associated with smaller local influence volumes, which would require less computational expense, since the computational cost varies with the square of L_{opt} (as shown in Fig. 8). We tested this by comparing the mean L_{opt} values versus Δt for three vertical ranges: troposphere (1000–100 hPa), stratosphere (100–1 hPa), and mesosphere (<1 hPa). Figure 7c shows that a doubling of Δt from 900 to 1800 s results in a 15% increase in the mesospheric L_{opt} for NMIF (700 to 800 km), and in the troposphere, L_{opt} only increases slightly with Δt . To compensate for the doubling of Δt (thereby halving number of time steps), L_{opt} must increase by at least 44% to offset the increase in the overall cost. Therefore, running the optimal cases actually takes less computational time for larger Δt , since the decreased number of time steps is not entirely compensated by the increase in L_{opt} . So our expectation of reduced LETLM execution time with reduced Δt is also not true. Since smaller Δt does not result in improved skill or reduction in computation time, we decided to use $\Delta t = 60$ min results in the detailed comparisons with the TLM. This also allows comparison with results in Fig. 8, which used $\Delta t = 60$ min, but with lower resolution.

We also note that L_{opt} is highly sensitive to NMI, particularly in the stratosphere and mesosphere (Fig. 7c). The mesospheric values of L_{opt} for NMIF are ~ 200 km higher than NMIT and the stratospheric values are ~ 100 km higher. The tropospheric values are similar, as expected, since the NM have only a small contribution at lower altitudes (see Fig. 4). We also note in Fig. 7d that β_{opt} increases with Δt and β_{opt} is generally larger for NMIF than for NMIT. β_{opt} also decreases with increased altitude, with tropospheric values exceeding those in the stratosphere and mesosphere.

c. Globally and vertically averaged errors for reference configuration

We next examine the globally and vertically averaged errors $\bar{\epsilon}$ of the optimally tuned LETLM with $\Delta t = 60$ min. We note that the LETLM was tuned at 3 h, but single 6-h forecasts were subsequently made using the 3-h values of $L_{\text{opt}}(k)$ and $\beta_{\text{opt}}(k)$. The 3-h errors computed

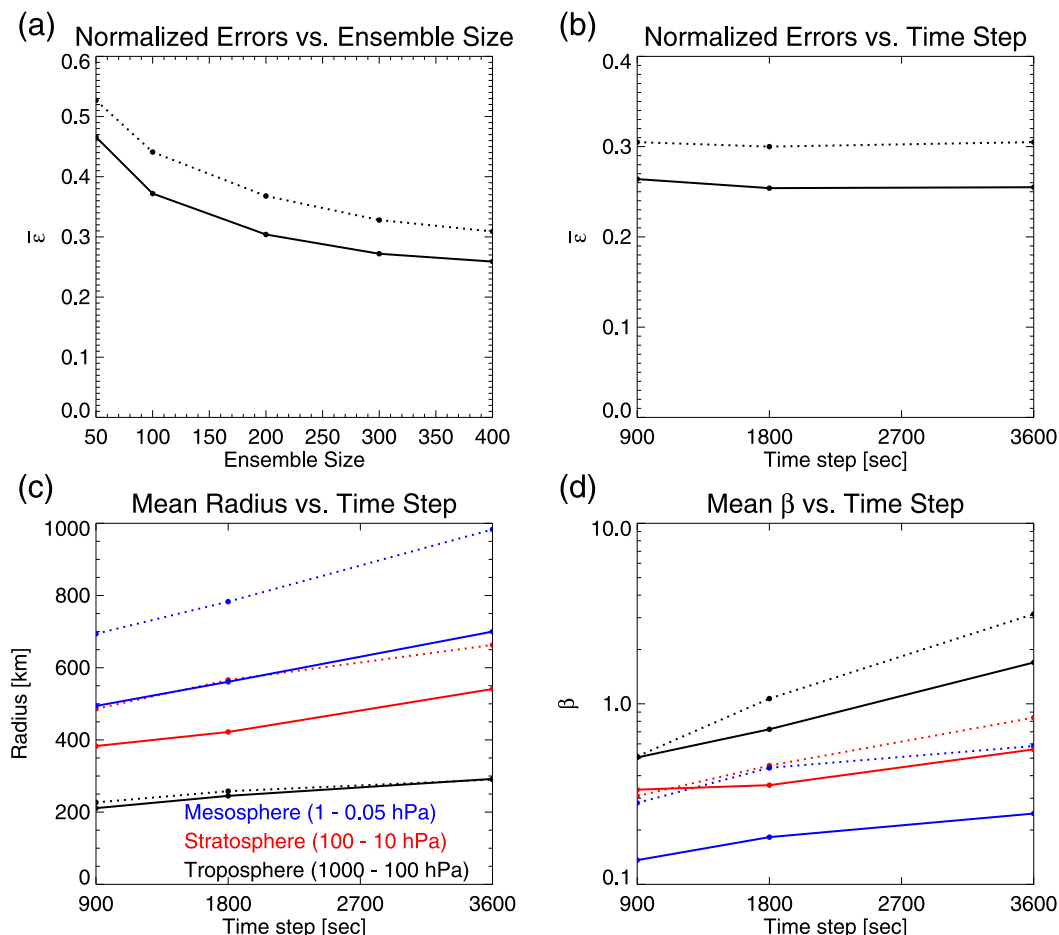


FIG. 7. (a) Vertically integrated 3-h forecast errors $\bar{\epsilon}$ as a function of ensemble size for NMIF (solid) and NMIT (dotted). (b) $\bar{\epsilon}$ as a function of time step for NMIF (solid) and NMIT (dotted). (c) Vertically integrated optimal radius as a function of time step for three vertical ranges: mesosphere (1–0.05 hPa), blue lines; stratosphere (100–10 hPa), red lines; troposphere (1000–100 hPa), black lines. Solid (dotted) lines are for NMIF (NMIT). (d) As in (c), but for β .

from these single forecasts (0.304 and 0.255 for NMIF and NMIT, respectively) are very close to the errors computed using the full-tuning method with 220 forecasts (0.305 and 0.255 for NMIF and NMIT, respectively). For comparison, the TLM gives $\bar{\epsilon} = 0.262$ for 3-h forecasts, so the NMIT version of the LETLM provides smaller overall errors.

Figure 8 shows $\bar{\epsilon}$ for LETLM (green), TLM (black), “persistence” forecast using TLM = I (red), and perturbation size in normalized units (blue). The solid (dashed) lines indicate NMIT (NMIF). The persistence errors grow to nearly the perturbation size after 6 h. The TLM and LETLM have considerable skill compared to the persistence. The LETLM skill for NMIF is worse than the TLM beyond 1 h, but NMIT shows better LETLM skill from 1 to 3 h and similar skill from 4 to 6 h. These results provide encouragement that at the operational inner loop resolution, the LETLM is on

par with the TLM in propagating a realistic analysis perturbation.

d. Globally averaged vertical error profiles for reference configuration

Figure 9 shows vertical profiles of tuning parameters and 3-h forecast errors. As discussed above, we fixed z_{halo} and z_{column} , and only tune L and β , but all four parameters are included for illustration. Additional forecast improvements could eventually be achieved by simultaneously tuning all four parameters. The second and third rows show $\epsilon(k)$ and errors for each model variable. Figure 10 is similar to Fig. 9, but emphasizes the lower troposphere (500–1000 hPa).

Figure 10a shows that $L_{\text{opt}}(k) \sim 250\text{--}300$ km near the surface for both NMIF and NMIT. Figure 10b shows $\beta_{\text{opt}}(k)$ is large near the surface, but decreases over the boundary layer, and is larger for NMIF than NMIT.

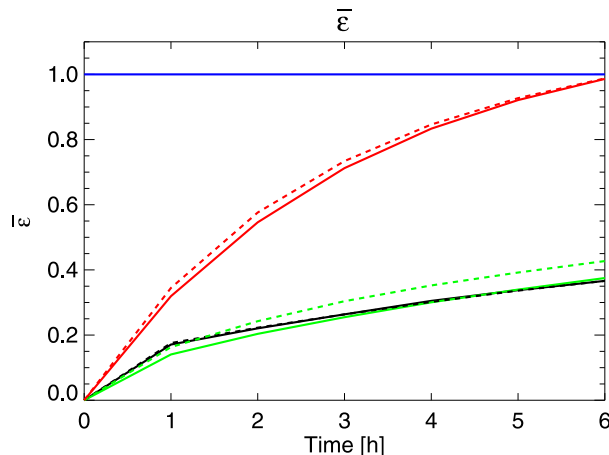


FIG. 8. Vertically integrated errors ($\bar{\epsilon}$) for the propagation of a 4DVar perturbation over 6 h. Lines indicate the magnitude of the propagated perturbation in normalized units (blue), errors under the assumption of a trivial “persistence” TLM = \mathbf{I} (identity matrix), equivalent to the 3DVar case where the increments are not propagated in time (red), errors for propagation of the perturbation using the LETLM (green), and errors for the propagation of the perturbation using the TLM (black). The solid (dashed) lines are for NMIF (NMIF). The LETLM experiments are run with 400 ensemble members, a time step of 60 min, and full tuning of L and β .

At higher altitudes (Fig. 9a), $L_{\text{opt}}(k)$ increases from ~ 100 to ~ 0.5 hPa to values of up to 1100 km (1000 km) for NMIF (NMIF). This indicates faster processes at higher levels, so the LETLM needs a wider local influence volume. The $L_{\text{opt}}(k)$ profiles determined in F18 were significantly larger than those seen here (~ 750 km in the troposphere and ~ 1750 km at higher levels). This indicates that $L_{\text{opt}}(k)$ is not solely determined by physical processes, but also depends on model resolution (the denser T119 grid may allow optimization with smaller lengths due to more available points for fixed L). Figure 9a shows that $L_{\text{opt}}(k)$ in the USLM is smaller for NMIF than for NMIF. At 2 hPa, for example, $L_{\text{opt}}(k)$ decreases from 1000 to 700 km when NMI is applied; $\beta_{\text{opt}}(k)$ is also sensitive to NMI in the USLM with lower values occurring for NMIF. We also attempted tuning L_{opt} as a function of latitude and level. Error profiles (not shown) indicate the additional latitudinal tuning does not significantly affect the overall errors ($\sim 2\%$ – 3% error reduction).

The wind and T errors show LETLM skill at all levels, exceeding TLM skill up to ~ 700 hPa (Figs. 10f–h). This suggests the LETLM physics provides additional information not captured in the TLM’s simplified physics. Tropospheric U , V , and T errors are not very sensitive to NMI. At altitudes above ~ 700 hPa, LETLM errors for NMIF are generally larger than TLM errors, but still

show considerable skill. The NMIF results show sharply decreased errors in U , V , and T relative to NMIF, with $\epsilon(k)$ on par with the TLM above ~ 2 hPa. Note that TLM errors for NMIF (dotted black lines in Figs. 9, 10) are similar for U , V , and T , suggesting the TLM accurately propagates the fast-moving modes, as discussed in section 3.

LETLM errors for Q , Z , P , and $\dot{\eta}$ also show skill, and these are generally similar to TLM errors (due to strong variation with altitude we show these errors as percentage errors relative to the perturbation size). Note that P is horizontally constant for hybrid vertical levels above 87 hPa, so P errors are not shown above this level. The NMIF errors are reduced relative to NMIF for these variables as well, with a very large reduction ($\sim 50\%$) for Z at high altitudes (Fig. 9j). The LETLM does generally better with these variables than the TLM throughout the troposphere as well (Figs. 10i–l).

e. Zonal mean error cross sections for reference configuration

The globally averaged errors show the LETLM is competitive with the TLM, particularly for NMIF. As a final comparison, we examine latitude–pressure error cross sections. Figures 11a–d and 12a–c show error standard deviations for all model variables for the LETLM with NMIF. U , V , and T show enhanced errors in the tropical tropopause and the lower mesosphere, particularly in the NH; T errors are also large in the tropical troposphere, likely associated with convection; Z errors are very large throughout the USLM region; P errors maximize in the tropics as well, while $\dot{\eta}$ errors are large over a broad range of latitudes and altitudes; and Q errors are largest in the troposphere and smaller in the stratosphere and mesosphere. Localized regions of larger Q errors occur in the polar regions of both hemispheres.

For NMIF (Figs. 11e–h), errors are moderately reduced relative to NMIF at high altitudes for U , V , and T and strongly reduced for Z , due to the NM having a large Z signal in the USLM. For NMIF, there are still elevated Z errors in the equatorial region from the troposphere to the top. This appears to be associated with the tropospheric T errors in the convective regions, which affects the entire column due to the hydrostatic relationship between T and Z . There are also error reductions in P , Q , and $\dot{\eta}$ (Figs. 12d–f) in NMIF relative to NMIF, but these are not as dramatic as for Z .

The conventional TLM errors for U , V , T , and Z are provided in Figs. 11i–l. While the TLM is better over large regions of the atmosphere, the LETLM does slightly better overall at 3 h (Fig. 8). This is because the

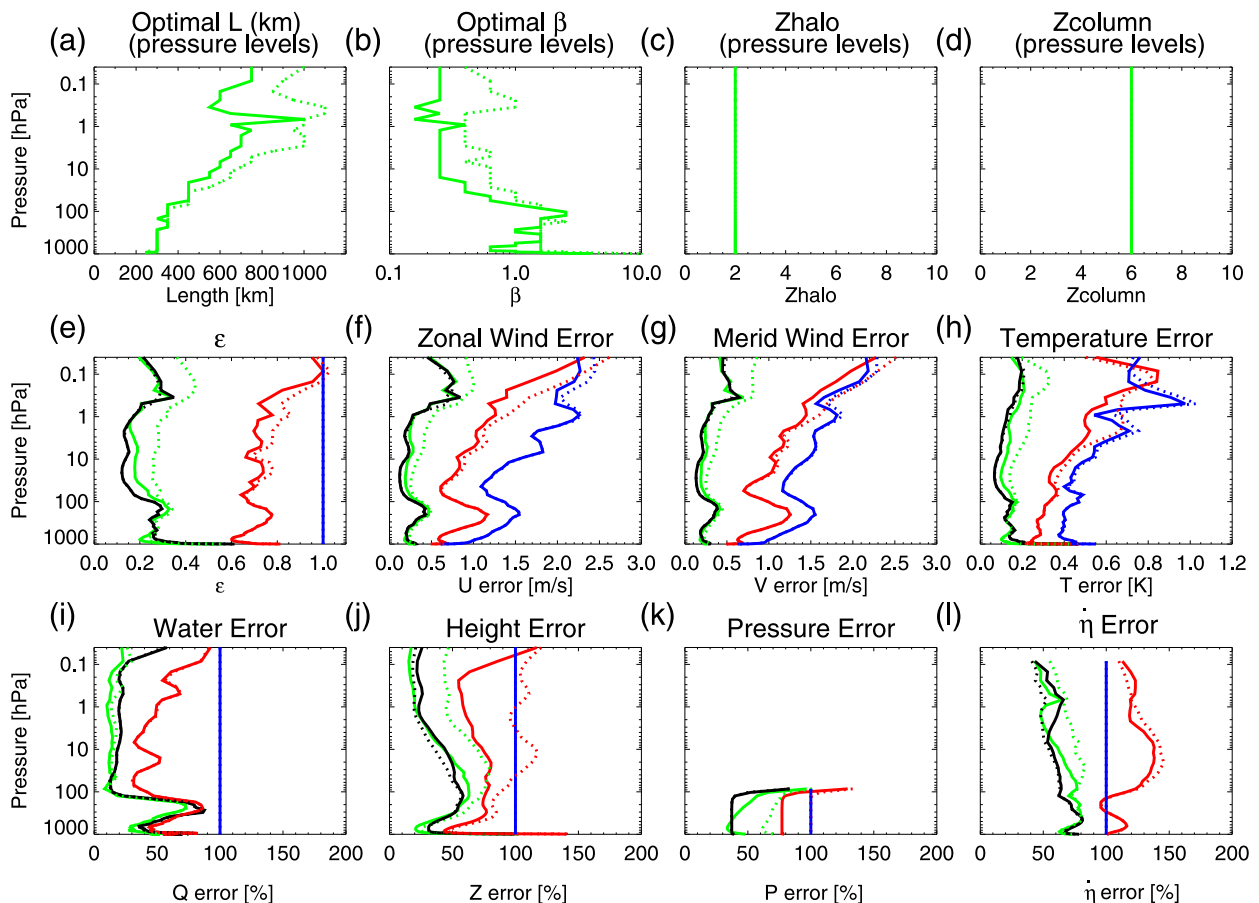


FIG. 9. Optimally tuned profiles of (a) the horizontal influence volume radius L and (b) the pseudoinverse cutoff β . (c), (d) Profiles of the prescribed vertical parameters z_{halo} and z_{column} . (e)–(l) Globally averaged profiles of 3-h forecast error ϵ and the 7 model state variables: U , V , T , Q , Z , P , and η . Color coding is green for the LETLM, black for the TLM, red for persistence, and blue for the perturbation size. Solid (dotted) lines are for NMIT (NMIF).

vertical weighting used for $\bar{\epsilon}$ (which gives each model level equal weight) favors the lower atmosphere, where model levels are more closely spaced and where the LETLM does better. The TLM error structures are similar to NMIF errors for all variables, with maximum U , V , and T errors in the tropical troposphere and the mesosphere and large Z errors in the tropics. TLM errors in P , Q , and η (Figs. 12g–i) are also similar to NMIF. One noticeable difference is smaller stratospheric Q errors in NMIF than in the TLM, which may be due to parameterized water chemistry in the stratosphere and mesosphere (details in McCormack et al. 2008), which is modeled in the LETLM, but not modeled in the TLM. Overall, these error comparisons provide a consistent picture. 1) Both the LETLM and TLM have consistent skill relative to persistence for all model variables. 2) The LETLM with NMI has generally smaller errors than the LETLM without NMI. 3) The LETLM with NMI is equal to or better than the TLM in

the lower troposphere (700–1000 hPa) and in the lower mesosphere (above ~ 2 hPa), but is slightly worse from 700 to 2 hPa.

5. Summary

In this study, we increased the LETLM resolution to that of the currently operational NAVGEM inner loop (T119). One key hypothesis was that increasing resolution would require maintaining balance between ensemble size and LETLM computational stencil size. To avoid increasing the ensemble size, we reduced the LETLM time step, suspecting that smaller optimal lengths would result and therefore smaller ensembles would be needed. We found that while optimal lengths were smaller for smaller time step, they did not offset the computational need of more iterations for fixed length forecasts. Errors also increased slightly with reduced time step, so there was not a clear benefit.

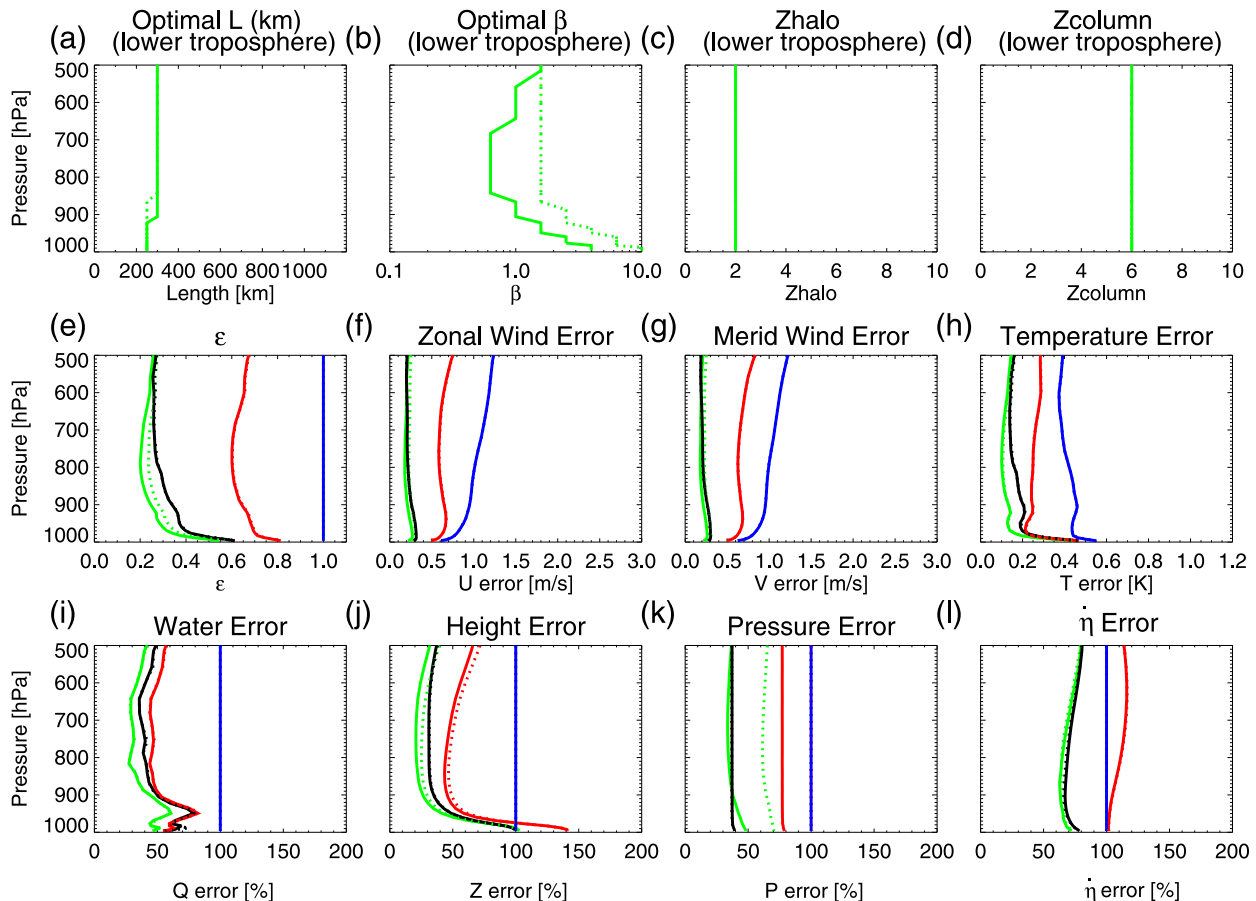


FIG. 10. As in Fig. 9, but with vertical axis that is linear in pressure from 500 to 1000 hPa in order to emphasize the lower troposphere.

Other proposed improvements included an enhanced local influence volume and using a thin grid. These changes yielded a factor of ~ 20 reduction in computational time and slightly increased the LETLM skill.

Comparisons with the traditional TLM (which includes boundary layer physics and vertical diffusion, but neglects moist physics, radiation, gravity wave drag, and ozone photochemistry) showed that the LETLM provides a viable alternative to the TLM, although currently requiring ~ 400 members to match overall TLM skill. However, if we focus on the troposphere (e.g., the lowest 30 model layers up to ~ 200 hPa), then the LETLM skill with NMI actually matches the TLM skill with ~ 200 members. We attribute the superior tropospheric skill to physical processes (e.g., moist physics) in the LETLM that are excluded from the TLM used in this study. The LETLM performed slightly worse in the upper troposphere and stratosphere. One cause was the presence of fast-moving gravity waves, which are difficult for the LETLM, but not for the TLM. We mitigated this problem by NMI application to the first three vertical normal modes. This improved LETLM

errors, such that the LETLM compared better with the TLM in the upper stratosphere, and matched the TLM skill in the lower mesosphere.

6. Discussion

LETLM performance at higher resolution exceeds the performance of the traditional TLM in the lower troposphere, while in the upper troposphere and stratosphere the LETLM performance continues to lag. While the performance lag in the stratosphere might be tolerated in practical applications (especially when mitigated using the NMI filtering), this is unsatisfactory from a theoretical perspective. Since the LETLM is better adapted to local operators, in order to understand these limitations we explore the computational stencils employed by NAVGEM. The forecast model can be decomposed into a sequence of local and nonlocal operators (see the [appendix](#)). The nonlocal operators include forward and inverse Fourier spectral transforms, implicit solvers, and some aspects of the physics such as deep convection and parameterized gravity wave drag.

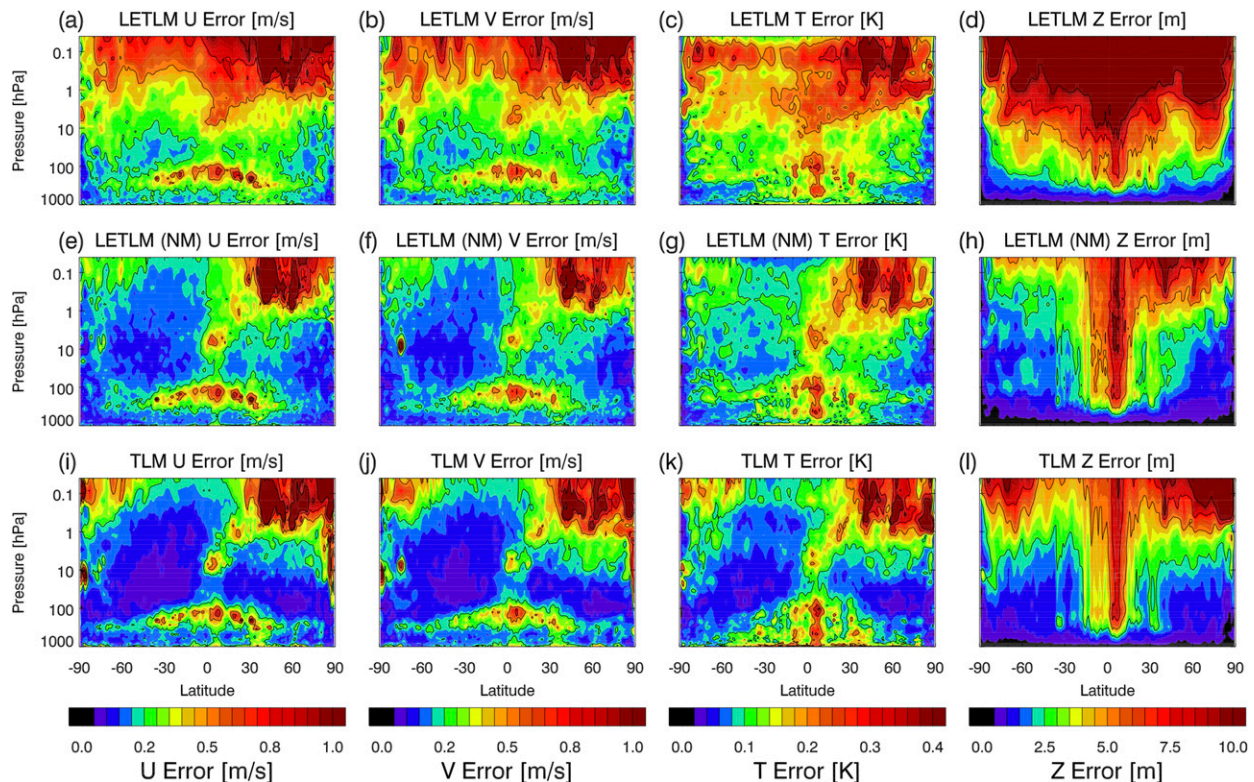


FIG. 11. Latitude–pressure cross sections of 3-h LETLM NMIF forecast error standard deviations for (a) U , (b) V , (c) T , and (d) Z . (e)–(h) As in (a)–(d), but for NMIT. (i)–(l) As in (a)–(d), but for the TLM.

The fact that the semi-implicit time step is performed in the coefficient space of global spherical harmonics means the computational stencil is global. Bishop et al. (2017) showed that the LETLM is precisely equal to the true TLM whenever (i) the ensemble size exceeds the degrees of freedom of the actual computational stencil, and (ii) ensemble perturbations are small enough to neglect the nonlinear terms affecting their evolution. Obviously, condition (i) cannot be satisfied for $O(100)$ member ensembles when the actual computational stencil is global. Thus, the global aspect of the semi-implicit time step in NAVGEM fundamentally limits the accuracy of the NAVGEM LETLM as currently configured, since the LETLM always operates in the grid space and incorrectly assumes a local computational stencil.

These insights motivated Yaremchuk et al. (2020, hereafter Y20) to transform the LETLM technique to ETLM (i.e., nonlocal) by removing the locality assumption. The approach represents an ETLM as a sequence of operations on sparse matrices constructed in parallel with the process of nonlinear ensemble propagation. The underlying idea is that most geophysical fluid dynamics (GFD) models (including NAVGEM, see the appendix) can be factored into a product of

linear (not necessarily local) and local nonlinear operations on a state vector. The linear operators (such as Fourier transforms and implicit solvers) are coded in the parent model, and therefore can be readily used in coding the ETLM application to a state vector. Furthermore, since implicit solvers in GFD models are usually applied to sparse matrices arising from discretization of the differential operators, the structure of these matrices can be obtained via the LETLM technique applied to auxiliary ensembles produced during the parent ensemble propagation. In principle, these retrievals can be performed in parallel with the parent ensemble, providing an accurate ETLM model by the end of integration. To demonstrate the feasibility of this approach, Y20 reconstructed the ETLM operator with machine accuracy for a SWM featuring a semi-implicit solver. Y20 also showed that additional computational saving can be achieved by assuming the ETLM operator evolves slowly compared to the model time step and, therefore, costly LETLM retrievals of the ETLM can be conducted less frequently, and the respective sparse matrices can be linearly interpolated and applied on every time step of the parent model.

Our diagnostics suggest that the component of the perturbation that the NAVGEM LETLM fails to capture

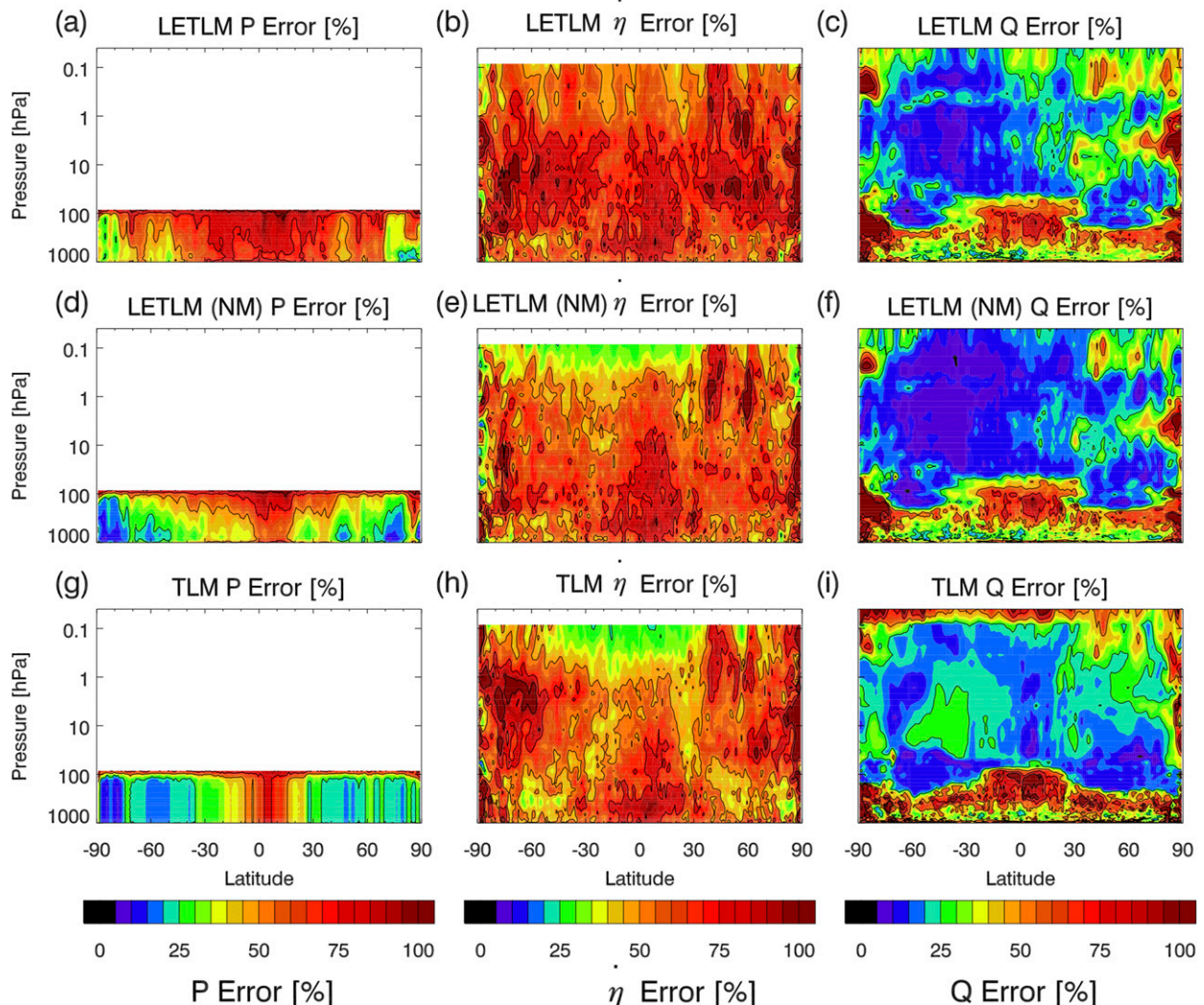


FIG. 12. Latitude–pressure cross sections of 3-h LETLM NMIF forecast error standard deviations for (a) P , (b) η , and (c) Q . (d)–(f) As in (a)–(c), but for NMIF. (g)–(i) As in (a)–(c), but for the TLM.

is very large scale in both the horizontal and the vertical. This raises the possibility of introducing a separate global ETLM for the very large-scale part of the perturbation. In theory this could be accurate provided the number of basis modes used to describe the large scale was smaller than the ensemble size. The result of this global ETLM time step for the very largest scales could then be blended with the results of an LETLM as described in this paper. We also note that the global computational stencil associated with semi-implicit numerical methods is not present in some operational environmental models such as ICON (Zängl et al. 2015) and MPAS (Klemp et al. 2018). In these models, only vertically propagating sound waves are handled implicitly; therefore, the entire computational stencil is horizontally localized. Hence, in some future study it would

be interesting to test whether the LETLM configuration presented here would yield a more accurate TLM for models like ICON and MPAS than it does for a semi-implicit model like NAVGEM.

Given the results from this paper and the insights from Y20, we suggest several routes for further development:

- 1) Testing the LETLM in hybrid-4DVar: Given the close agreement between the TLM and LETLM, LETLM integration in the hybrid 4DVar solver is warranted. Several practical considerations need to be addressed, including precomputing and storing the LETLM as a sparse matrix operator for the iterative hybrid 4DVar computation. The local nature of the LETLM might warrant implementation on massively parallel graphical processing units (GPUs).

We note the LETLM has been demonstrated with a hybrid 4DVar solver in the SWM framework (Allen et al. 2017).

- 2) Development of hybrid TLM methods: As shown in the appendix, a typical GFD numerical model can be represented as a sequence of linear operations that can be handled by the method proposed in Y20, and local nonlinear operators. The nonlinear operators may either have easily derivable TLM and adjoint (ADJ) codes (e.g., semi-Lagrangian advection) or lack precise TLM interpretation (e.g., discontinuous physical processes and vertical mixing). Given this decomposition, it might be beneficial to construct a hybrid TLM model that combines traditional numerical methods to derive the TLM and ADJ of the differentiable nonlinearities with ETLM techniques for approximation of discontinuous and linear operators. As a practical example, a hybrid NAVGEM TLM could be constructed using the semi-Lagrangian TLM for the model dynamics and an ETLM to handle the physics. Since NAVGEM physics is calculated in a vertical column, the ETLM computational stencil size could simply be the number of vertical levels (~ 60) multiplied by the number of variables, which could be balanced with realistic ensemble sizes. A major benefit of this approach is that changes to the physics would be automatically included in the LETLM, thereby avoiding the difficult tasks involved with linearization.
- 3) Apply the LETLM to other NWP models: As discussed above, the LETLM may better approximate the linearized versions of models that have inherently localized computational stencils.

Acknowledgments. This work was funded by the U.S. Office of Naval Research. NAVGEM analyses and forecasts as well as LETLM and TLM forecasts were produced under a grant of computer time from the Department of Defense High Performance Computing Modernization Program. This work also benefitted from the helpful comments and suggestions by two anonymous reviewers.

APPENDIX

Decomposition of the NAVGEM Forecast Model into a Sequence of Local and Nonlocal Operators

The forecast model execution over one time step can be decomposed in a sequence of operations:

$$\mathbf{x}_{m+1} = \mathbf{P}\mathbf{F}^{-1}\mathbf{E}^{-1}\mathbf{D}_v\mathbf{E}\mathbf{D}_H\mathbf{F}\mathbf{S}\mathbf{x}_m, \quad (\text{A1})$$

where the variables are defined as the following:

- \mathbf{x}_m and \mathbf{x}_{m+1} : model states in grid space at time index m and $m + 1$;
- S**: semi-Lagrangian operator that calculates backward trajectories and interpolates state variables and forcing terms to the departure points. This operation is nonlinear (trajectory computations and a fixed-point iteration for computing departure points), and mildly nonlocal (interpolation as far as several grid points);
- F** (\mathbf{F}^{-1}): spectral transform (and its inverse) from 2D gridded U, V fields to spectral harmonics of vorticity and divergence. This operation is strongly nonlocal because it performs both the global Fourier decomposition and computes derivatives in the spectral coefficient space. However, due to unitarity of the Fourier transform and availability of the respective code in NAVGEM, the TLM and ADJ code development is not required.
- E** (\mathbf{E}^{-1}): eigenvector transform (and its inverse) in the vertical. This is a global computation in the vertical but local horizontally;
- D_H** and **D_v**: diagonal diffusion operators in spectral harmonic space. Development of TLM and ADJ codes is not required; and
- P**: the physics tendency operator, which is mildly nonlocal because it operates in the vertical only. **P** can further be decomposed in the action of local and mildly nonlocal (implicit diffusion) operators. Since number of levels is comparable with ensemble size, the structure of **P** could be accurately retrieved by the LETLM technique.

REFERENCES

- Allen, D. R., K. W. Hoppel, and D. D. Kuhl, 2015: Wind extraction potential from ensemble Kalman filter assimilation of stratospheric ozone using a global shallow water model. *Atmos. Chem. Phys.*, **15**, 5835–5850, <https://doi.org/10.5194/acp-15-5835-2015>.
- , —, and —, 2016: Hybrid ensemble 4DVar assimilation of stratospheric ozone using a global shallow water model. *Atmos. Chem. Phys.*, **16**, 8193–8204, <https://doi.org/10.5194/acp-16-8193-2016>.
- , 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>.
- Eckermann, S., 2009: Hybrid σ - p coordinate choices for a global model. *Mon. Wea. Rev.*, **137**, 224–245, <https://doi.org/10.1175/2008MWR2537.1>.
- 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>.
- , D. R. Allen, C. H. Bishop, R. Langland, K. W. Hoppel, and D. D. Kuhl, 2018: First application of the local ensemble tangent linear model (LETLM) to a realistic model of the global atmosphere. *Mon. Wea. Rev.*, **146**, 2247–2270, <https://doi.org/10.1175/MWR-D-17-0315.1>.
- Gauthier, P., and J.-N. Thépaut, 2001: Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Météo-France. *Mon. Wea. Rev.*, **129**, 2089–2102, [https://doi.org/10.1175/1520-0493\(2001\)129<2089:IOTDFA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2089:IOTDFA>2.0.CO;2).
- , M. Tanguay, S. Laroche, S. Pellerin, and J. Morneau, 2007: Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada. *Mon. Wea. Rev.*, **135**, 2339–2354, <https://doi.org/10.1175/MWR3394.1>.
- Geer, A. J., and Coauthors, 2017: The growing impact of satellite observations sensitivity to humidity, cloud and precipitation. *Quart. J. Roy. Meteor. Soc.*, **143**, 3189–3206, <https://doi.org/10.1002/qj.3172>.
- Hagan, M. E., and J. M. Forbes, 2003: Migrating and nonmigrating semidiurnal tides in the upper atmosphere excited by tropospheric latent heat release. *J. Geophys. Res.*, **108**, 1062, <https://doi.org/10.1029/2002JA009466>.
- Hogan, T. F., and Coauthors, 2014: The Navy global environmental model. *Oceanography*, **27**, 116–125, <https://doi.org/10.5670/oceanog.2014.73>.
- , T. E. Rosmond, and R. Gelaro, 1992: The NOGAPS forecast model: A technical description. NRL ADA247216, Naval Research Laboratory, Monterey, CA, 218 pp., <http://www.dtic.mil/docs/citations/ADA247216>.
- Janisková, M., and P. Lopez, 2012: Linearized physics for data assimilation at ECMWF. ECMWF Tech. Memo. 666, 26 pp.
- JMA, 2019: Outline of the operational numerical weather prediction at the Japan Meteorological Agency. Japan Meteorological Agency, accessed 22 November 2019, <https://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline2019-nwp/index.htm>.
- Kasahara, A., and K. Puri, 1981: Spectral representation of three-dimensional global data by expansion in normal mode functions. *Mon. Wea. Rev.*, **109**, 37–51, [https://doi.org/10.1175/1520-0493\(1981\)109<0037:SROTDG>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0037:SROTDG>2.0.CO;2).
- Klemp, J. B., W. C. Skamarock, and S. Ha, 2018: Damping acoustic modes in compressible horizontally explicit vertically implicit (HEVI) and split-explicit time integration schemes. *Mon. Wea. Rev.*, **146**, 1911–1923, <https://doi.org/10.1175/MWR-D-17-0384.1>.
- 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>.
- Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow water model, with applications to normal mode initialization. *Contrib. Atmos. Phys.*, **50**, 253–271.
- McCormack, J. P., K. W. Hoppel, and D. E. Siskind, 2008: Parameterization of middle atmospheric water vapor photochemistry for high-altitude NWP and data assimilation. *Atmos. Chem. Phys.*, **8**, 7519–7532, <https://doi.org/10.5194/acp-8-7519-2008>.
- 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>.
- Poterjoy, J., and F. Zhang, 2015: Systematic comparison of four-dimensional data assimilation methods with and without the tangent linear model using hybrid background error covariance: E4DVar versus 4DVar. *Mon. Wea. Rev.*, **143**, 1601–1621, <https://doi.org/10.1175/MWR-D-14-00224.1>.
- Rabier, F., H. Jarvinen, E. Klinker, J. F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143–1170, <https://doi.org/10.1002/qj.49712656415>.
- Rawlins, F., S. P. Ballard, K. J. Bovis, A. M. Clayton, D. Li, G. W. Inverarity, A. C. Lorenc, and T. J. Payne, 2007: The Met Office global four-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **133**, 347–362, <https://doi.org/10.1002/qj.32>.
- Rosmond, T., and L. Xu, 2006: Development of NAVDAS-AR: Non-linear formulation and outer loop tests. *Tellus*, **58A**, 45–58, <https://doi.org/10.1111/j.1600-0870.2006.00148.x>.
- Rosmond, T. E., 1997: A technical description of the NRL adjoint modeling system. Tech. Rep. NRL/MR/7532/97/7230, Naval Research Laboratory, 57 pp., <http://www.dtic.mil/dtic/tr/fulltext/u2/a330960.pdf>.
- 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>.
- Yaremchuk, M., D. Nechaev, and S. Frolov, 2020: On the ensemble-based linearization of numerical models. *Quart. J. Roy. Meteor. Soc.*, **147**, 1026–1039, <https://doi.org/10.1002/qj.3723>.
- Žagar, N., A. Kasahara, K. Terasaki, J. Tribbia, and H. Tanaka, 2015: Normal-mode function representation of global 3-D data sets: Open-access software for the atmospheric research community. *Geosci. Model Dev.*, **8**, 1169–1195, <https://doi.org/10.5194/gmd-8-1169-2015>.
- Zängl, G., D. Reinert, P. Rípodas, and M. Baldauf, 2015: The ICON (ICOSahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quart. J. Roy. Meteor. Soc.*, **141**, 563–579, <https://doi.org/10.1002/qj.2378>.
- Zhang, L., and Coauthors, 2019: The operational global four-dimensional variational data assimilation system at the China Meteorological Administration. *Quart. J. Roy. Meteor. Soc.*, **145**, 1882–1896, <https://doi.org/10.1002/qj.3533>.